Universitat Oberta de Catalunya

Projecte Fi de Carrera

Enginyeria d'Informàtica

Supervisió i Gestió de logs en entorns d'explotació

Alumne: Blas Torregrosa García

Universitat Oberta de Catalunya

Projecte Fi de Carrera

Enginyeria d'Informàtica

Supervisió i Gestió de logs en entorns d'explotació

Alumne: Blas Torregrosa García

Consultor: Jordi Ceballos Villach

Vull dedicar aquest Projecte in memoriam a Dennis Ritchie (1941–2011). Pioner de la informàtica, Ritchie va ser el creador del llenguatge de programació C i co–creador, juntament amb Ken Thompson, del sistema operatiu Unix.

Agraïments

A Jordi Ceballos, consultor d'aquest projecte, pel seu interès i els seus valuosos consells.

Als creadors de ScribTeX, l'editor online de LaTeX, amb el qual he escrit aquesta Memòria i que m'ha facilitat molt la tasca.

Al Luis, per ajudar-me sempre en tot.

A Pili, José Luis, Jorge, Jesús i Miguel (segur que m'oblido d'algun), perquè el cafè sense ells no fa un gust tan bo.

A l'Óscar, per fer-me posar sempre els peus a terra.

Al Paco, per voler ser el meu amic en la distància.

Resum

El present Projecte examina la utilització d'eines de gestió de logs en entorns d'explotació de sistemes informàtics, i el seu objectiu és doble: proporcionar una visió general dels sistemes de gestió de logs, i dur a terme una avaluació de productes que realitzin aquesta gestió de logs usant un mètode.

Es descriu el que són els logs y la problemàtica associada a ells. Es presenten les fonts principals de logs, així com els formats més freqüents (i els intents d'estandardització). També es considera una infraestructura teòrica de gestió de logs (amb la seva arquitectura y les seves funcions), que serveix de base per ordenar les necessitats de la gestió per a qüestions d'operativa diària dels sistemes i per a les qüestions relatives a la seguretat.

Es revisen breument tres marcs de referència existents en l'actualitat per a l'avaluació de programari. Es determina que l'ús d'un mètode formal per avaluar productes programari assegura que –en l'avaluació– es minimitza la influència dels possibles prejudicis de l'avaluador, i que es poden examinar més productes candidats. El mètode utilitzat en aquest Projecte està basat principalment en un dels marcs de referència descrits (a saber: QSOS v1.6), i consta de quatre fases: definició de criteris d'avaluació, una llista inicial de productes candidats, obtenció de dades mesurables, i una avaluació final.

Partint de zero, s'ha efectuat una definició completa de criteris d'avaluació per a sistemes de gestió de logs. Considerant com a punt de partida uns escenaris teòrics d'ús de logs, es defineixen uns requisits que el programari hauria de satisfer en aquests escenaris. A partir d'aquests requisits s'obtenen directament els criteris d'avaluació, dividits en tres grups principals: criteris genèrics (relatius a la fiabilitat, al manteniment i a l'explotabilitat del programari), criteris funcionals (relatius a les funcionalitats pròpies de la gestió de logs), i criteris tècnics (relatius a la usabilitat, a la interoperabilitat y a la seguretat dels usuaris).

Per mitjà del mètode desenvolupat en el Projecte s'avaluen cinc productes candidats (Octopussy, Splunk, Graylog2, Loggly i Logentries), que són suficientment representatius i cobreixen la major part del procés de gestió de logs. S'expliquen els passos del mètode en la seva aplicació a aquests productes candidats, i es comenten els resultats de l'avaluación. L'avaluació finalitza amb una sèrie de consideracions sobre el propi mètode, i amb una classificació –no només numèrica– dels productes candidats.

Paraules clau logs, gestión de logs, SIEM, avaluació de programari, QSOS, 8pussy, Splunk, Graylog2, Loggly, Logentries.

Resumen

El presente Proyecto examina la utilización de herramientas de gestión de logs en entornos de explotación de sistemas informáticos, y su objetivo es doble: proporcionar una visión general de los sistemas de gestión de logs, y llevar a cabo una evaluación de productos que realicen esa gestión de logs usando un método.

Se describe qué son los logs y la problemática asociada a ellos. Se presentan las fuentes principales de logs, así como los formatos más frecuentes (y los intentos de estandarización). También se considera una infraestructura teórica de gestión de logs (con su arquitectura y sus funciones), lo cual sirve de base para ordenar las necesidades de la gestión para cuestiones de operativa diaria de los sistemas y para las cuestiones relativas a la seguridad.

Se revisan brevemente tres marcos de referencia existentes en la actualidad para la evaluación de productos software. Se determina que el uso de un método formal para evaluar productos software asegura que –en la evaluación– se minimiza la influencia de los posibles prejuicios del evaluador, y que se pueden examinar más productos candidatos. El método utilizado en este Proyecto está basado principalmente en uno de los marcos de referencia descritos (a saber: QSOS v1.6), y consta de cuatro fases: definición de criterios de evaluación, una lista inicial de productos candidatos, obtención de datos medibles, y una evaluación final.

Partiendo de cero, se ha efectuado una definición completa de criterios de evaluación para sistemas de gestión de logs. Considerando como punto de partida unos escenarios teóricos de uso de logs, se definen unos requisitos que el software debería satisfacer en estos escenarios. A partir de estos requisitos se obtienen directamente los criterios de evaluación, divididos en tres grupos principales: criterios genéricos (relativos a la fiabilidad, al mantenimiento y a la explotabilidad del software), criterios funcionales (relativos a las funcionalidades propias de la gestión de logs), y criterios técnicos (relativos a la usabilidad, a la interoperabilidad y a la seguridad de los usuarios).

Por medio del método desarrollado en el Proyecto se evalúan cinco productos candidatos (Octopussy, Splunk, Graylog2, Loggly y Logentries), que son suficientemente representativos y cubren la mayor parte del proceso de gestión de logs. Se explican los pasos del método en su aplicación a estos productos candidatos, y se comentan los resultados de la evaluación. La evaluación finaliza con una serie de consideraciones sobre el propio método, y con una clasificación –no sólo numérica– de los productos candidatos.

Palabras clave logs, gestión de logs, SIEM, evaluación de software, QSOS, 8pussy, Splunk, Graylog2, Loggly, Logentries.

Abstract

This work examines the use of log management tools in operational environments of computer systems, and its goal is twofold: it aims to provide an overview of the log management systems, but also to carry out an evaluation of products that perform log management; such an evaluation is accomplished by using a method.

One describes what the logs are, and the questions posed by their treatment are considered. The main log sources and the most common formats (as well as the standardization attempts) are also presented. In addition to this, a theoretical infrastructure for log management (with its corresponding description of architecture and functions) is considered, which provides the basis for a classification of management needs in dealing with daily operational issues and with security issues.

Three currently existing reference frameworks for the evaluation of software products are briefly examined. It is determined that resorting to a formal method to evaluate software products ensures that –in the evaluation process– the influence of potential evaluator prejudices is minimized, and that more candidate products can be incorporated into the evaluation process. The method used in this Work is mainly based on one of the abovementioned reference frameworks (namely, QSOS v1.6), and consists of four phases: definition of evaluation criteria, a preliminary list of candidate products, obtaining of measurable data, and a final evaluation.

A complete definition of evaluation criteria for log management systems has been accomplished from scratch. Starting from some theoretical scenarios in the use of logs, the desired requirements that the software should fulfill in these scenarios are defined. From these requirements, the sought evaluation criteria are directly obtained; such criteria are divided into three main groups: generic criteria (relating to software reliability, maintenance and exploitability), functional criteria (related to specific functionality within the field of log management), and technical criteria (regarding usability, interoperability and security of users).

By means of the method developed in this Work, five candidate products (Octopussy, Splunk, Graylog2, Loggly and Logentries) are evaluated; these products are sufficiently representative and cover most of the log management process. The steps taken by the method, when applied to these candidate products, are explained, and the results of the evaluation are discussed. The evaluation ends with several considerations concerning the method itself, and the candidate products are rated according to a (not merely numerical) classification.

Keywords logs, log management, SIEM, Software Evaluation, QSOS, 8pussy, Splunk, Graylog2, Loggly, Logentries.

Índex

Índex de figures

Índex de taules

Capítol 1
Introducció

"La vida no és solament somiar i començar projectes, també cal ser perseverant i insistir fins a fer realitat el que cadascun es proposa."

Héctor García –Kirai– en Kirainet.com

1.1 Justificació del PFC i context en el qual es desenvolupa: punt de partida i aportació del PFC

Dia a dia els sistemes d'informació generen una informació molt valuosa que, pel seu format i el seu volum, amb prou feines s'utilitza. Es tracta dels *registres d'activitat o logs* que recullen cadascuna de les aplicacions i sistemes que configuren la infraestructura TIC.

En general, la relació dels tècnics TIC amb els logs es redueix a consultar-los després d'una fallada en els sistemes o aplicacions, generalment a petició d'un usuari, utilitzant el típic grep, el awk i molta paciència.

No obstant això, els logs són la principal font d'informació sobre l'activitat de la xarxa, els sistemes i les aplicacions, i resulten imprescindibles en:

- Detecció de problemes de maquinari/programari
- Detecció d'atacs i intruses
- Anàlisis forense de sistemes

Per tant, els logs haurien d'analitzar-se i monitoritzar-se, sobretot en els sistemes de producció crítics per evitar la degradació del servei o interrupcions del mateix. En ocasions, això no és una possibilitat, més aviat és una obligació en compliment de la normativa vigent.

El tractament dels logs no resulta senzill per la seva pròpia naturalesa:

- El volum de dades que es pot generar en pot ser molt gran
- Els log són generats per màquines i la informació es troba emmagatzemada segons un format, normalment textual, determinat per cada tipus de servidor o aplicació.

Els logs realment són només dades que, processats i analitzats, poden convertir-se en informació valuosa. Els logs es componen d'entrades de log, i cadascuna d'elles conté dades relatives a un esdeveniment concret que ha succeït en el sistema o a la xarxa.

El procés per a la completa gestió de logs inclou les etapes següents:

- **Recol·lecció de logs**. En un entorn de producció, molts dispositius, sistemes operatius i aplicacions generen i emmagatzemen logs. Els problemes amb els quals cal tractar en aquesta fase són la diversitat de fonts dels logs, la inconsistència en el contingut, la inconsistència de les marques temporals (*timestampt*) o la inconsistència de formats. En moltes organitzacions, per facilitar les fases posteriors, es procedeix a la conversió de logs a un format únic que usi una representació consistent dels camps.

  Un altre problema associat és el volum de logs que, depenent del tipus d'organització i nivell de generació, pot arribar a suposar un gran consum de recursos.

- **Reducció del soroll**. Com s'ha comentat, el volum de logs generat pot ser molt gran i, a més, el percentatge d'informació rellevant dins dels logs és, en ocasions, molt reduït. Aquesta reducció de soroll es realitza mitjançant el filtratge d'esdeveniments (eliminar aquells esdeveniments que no aporten informació rellevant) i l'agregació d'esdeveniments (en la qual entrades similars s'agrupen en una sola).

- **Anàlisi**. L'anàlisi de logs consisteix a estudiar les entrades de logs tractant d'identificar esdeveniments d'interès. Aquesta tasca, la més important en tot el procés, era realitzada pels tècnics TIC i considerada com de baixa prioritat. La forma de realitzar aquestes tasques és utilitzant tècniques de visualització (convertir les entrades de logs en un format adaptat als humans), informes (resums d'activitat significativa durant períodes de temps) o la correlació d'esdeveniments (cercar relacions entre dues o més entrades de log).

- **Gestió d'alertes**. L'anàlisi de logs genera, en la majoria de les ocasions, falses alarmes, el que redunda en una falta de confiança en el procés. L'objectiu d'aquesta fase és l'ajust de totes les fases per reduir les falses alarmes.

1.2 Objectius del PFC

L'objectiu principal és aprofundir en la gestió de registres d'activitat (logs) en entorns d'explotació amb les eines disponibles actualment en internet.

- L'objectiu del Projecte és, vista la problemàtica existent en els entorns de producció, realitzar un estudi de les eines, en relació a les seves capacitats i limitacions per gestionar logs.

- Realitzar una anàlisi comparativa que permeti discriminar el programari més adequat segons l'entorn.

1.3 Enfocament i mètode seguit

Per limitar l'abast del Projecte, la primera qüestió que em plantejo és decidir les eines de tractament de logs que siguin completes, representatives, actualitzades, permetin integració amb altres eines i, o es tracti de programari lliure, o tinguin versions comercials utilitzables gratuïtament.

A la vista dels objectius, el PFC està orientat primer a la recerca d'unes eines que permetin realitzar la majoria de les fases que s'han definit a l'introducció. Com els entorns de producció poden ser molt heterogenis, a continuació s'estudia quines són les fonts de logs que normalment es poden trobar, per poder definir un conjunt de criteris amb els quals avaluar les diferents eines. Aquests criteris estaran, òbviament, en relació amb les capacitats de les eines per realitzar les diferents fases del tractament de logs.

A continuació, caldria determinar els criteris d'avaluació per a les eines. Es necessitaria un conjunt complet de criteris d'avaluació capaç de mesurar les diferents eines i fixar la funcionalitat esperada de les mateixes.

Una vegada establits els criteris, el següent pas seria utilitzar-los per avaluar les eines i disposar d'una llista de productes que serveixin per a la totalitat (o la major part) de la gestió de logs.

S'ha suposat un horari de 4 hores diàries de dilluns a divendres, utilitzant els caps de setmana per absorbir possibles retards tenint en compte que les dates dels lliuraments estan prefixades.

1.4 Planificació del Projecte

La planificació original del Projecte pot observar-se en el cronograma de la Figura 1.1. Es tracta d'un pla de treball guiat pels lliuraments de l'avaluació contínua i marca les fites del Projecte.

1.4.1 Desviacions de la planificació

Encara que s'ha tractat de seguir la planificació inicialment proposta, existeixen diverses qüestions que sorgeixen durant el desenvolupament de qualsevol projecte i que fan que es desviï del pla originalment marcat. En el cas concret que ens ocupa, a continuació descric els punts que han motivat diferències entre la planificació inicial i el desenvolupament final:

- L'elecció del programari a avaluar ha portat més temps de l'inicialment previst. La presència de noves formes d'afrontar el problema amb solucions basades en SaaS ha fet que s'intenti incloure aquestes eines encara que la seva maduresa no sigui molt gran. A més, inicialment no estava tancat el nombre d'eines a analitzar, que finalment ha quedat reduït a cinc eines, la qual cosa ha fet que el procés d'avaluació hagi estat més laboriós.
- En la planificació inicial no estava contemplada la utilització d'una metodologia per guiar el procés d'avaluació. Arribats a aquest punt, es va considerar convenient fixar una metodologia concreta; per a això, primerament es va fer una revisió d'algunes metodologies utilitzades últimament. El resultat d'aquesta revisió va donar com resultat una modificació d'una d'elles (QSOS) per adaptar-la a les necessitats específiques del nostre estudi. A més, va caldre definir completament els criteris de valoració ja que no existia gens per valorar programari de gestió de log. Tot això ha ocasionat un retard a tenir llests els criteris de valoració, i que el procés de valoració hagi començat més tard del previst.
- L'obtenció de les dades de l'avaluació ha estat més costosa del previst.

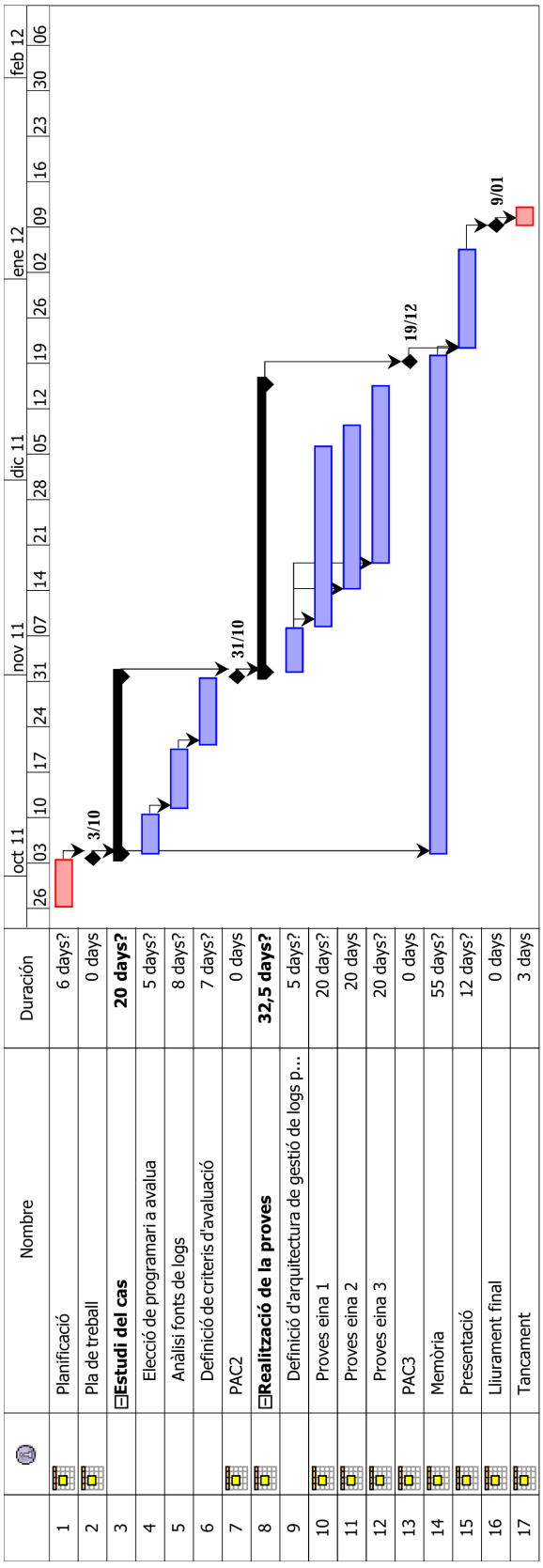

Figura 1.1: Planificació del Projecte

1.5 Productes obtinguts

Els productes generats pel present Projecte són els següents:

- **Memòria del Projecte**. És el principal producte del Projecte i en el qual es descriu què s'ha fet i com s'ha fet.
- **Llistat de criteris de valoració per a programari de gestió de logs**. Llista dels criteris de valoració utilitzats en l'avaluació dels productes, així com els possibles puntuacions en cada criteri. Pot localitzar-se en l'Apèndix C.
- **Valoració del programari seleccionat**. Llista amb la puntuació obtinguda en cadascun dels criteris pels productes candidats. Pot localitzar-se en l'Apèndix D.

1.6 Breu descripció dels altres capítols de la memòria

Els capítols de la memòria es troben distribuïts en la manera que es descriu a continuació. En primer lloc, en el Capítol 2, hi ha una descripció teòrica dels principals elements que componen la gestió de logs. A continuació, ja que un dels objectius és avaluar eines, es descriuen algunes metodologies d'avaluació de programari existent i s'introdueix la metodologia seguida per avaluar les diferents eines. El següent capítol està dedicat íntegrament a tot el procés d'avaluació, on es descriuen tots els passos seguits, així com els resultats de l'avaluació. El capítol final està dedicat a l'apartat de conclusions i possibles futurs treballs en aquest àmbit.

Els elements que resulten d'interès, però que no entren en la lògica d'aquesta distribució en capítols, s'inclouen com a apèndixs.

Capítol 2
Gestió de logs

"La fàbrica del futur tindrà només dos empleats: un home i un gos. La comesa de l'home serà donar de menjar al gos. El del gos serà cuidar que l'home no toqui l'equip."

Warren G. Bennis, Universitat de Sud de Califòrnia, assessor de Ronald Reagan i John F. Kennedy

2.1 Introducció

Un **log** és un registre dels esdeveniments que succeeixen dins dels sistemes i xarxes d'una organització. Els logs es componen d'entrades i cada entrada conté informació relativa a un esdeveniment que ha ocorregut dins dels sistemes o de la xarxa. Al principi, els logs es van utilitzar per a la depuració de programes però actualment els logs ens poden servir per a múltiples funcions, com l'optimització del rendiment en sistemes i xarxes, registre de les activitats dels usuaris, o la investigació d'activitat maliciosa.

2.1.1 Fonts

Els logs contenen una gran varietat de dades dels esdeveniments que succeeixen en sistemes i xarxes. Hi ha dues categories de logs que són d'especial interès:

- Logs de sistemes operatius i aplicacions, que contenen una gran varietat de dades.
- Logs del programari de seguretat, que contenen dades relatives a la seguretat informàtica.

Moltes de les fonts de log s'estan executant contínuament, per tant es generen entrades de forma contínua. Mentre que altres fonts s'executen periòdicament, sovint a intervals regulars.

2.1.1.1 Sistemes operatius

Els sistemes operatius dels servidors, estacions de treball i dispositius de xarxa (enrutadors i commutadors) normalment generen gran varietat de log. Els tipus més comuns de log són :

- **Esdeveniments del sistema**. Els esdeveniments del sistema són accions realitzades per algun component del sistema operatiu; per exemple, arrencada i parada del sistema o l'inici d'un servei. Per defecte, els sistemes operatius registren els esdeveniments fallits o els esdeveniments reeixits

més significatius, però és possible per als administradors especificar quins tipus d'esdeveniments es registren. Els detalls de cada tipus de log poden variar àmpliament, però cada esdeveniment sol portar una marca temporal així com dades sobre l'estat, codis d'error, nom del servei o usuari, o detalls associats a l'esdeveniment.

- **Auditoria**. Els sistemes operatius permeten als administradors especificar quins tipus d'esdeveniments s'han d'auditar, així com indicar quines accions han de registrar log. La auditoria pot contenir informació sobre intents d'autenticació, accés a fitxers, canvis en les regles de seguretat, canvis als comptes (creació, esborrament, assignació de privilegis, etc.), o utilització de privilegis.

Molts sistemes operatius utilitzen el format syslog (veure Apèndix A.1.1), mentre que els sistemes Windows (veure Apèndix A.1.2) tenen un format propi pels logs.

2.1.1.2 Aplicacions

Les aplicacions s'utilitzen per emmagatzemar, accedir i manipular les dades que l'organització utilitza en tot el seu procés de negoci. Les organitzacions utilitzen una gran varietat d'aplicacions comercials, com a servidors i clients de correu, servidors web, navegadors, servidors de fitxers, bases de dades, etc. A més d'aquest tipus d'aplicacions, en moltes organitzacions també utilitzen aplicacions desenvolupades a mida, adaptades a les seves necessitats específiques.

Algunes aplicacions generen els seus propis fitxers de log, mentre que unes altres utilitzen les capacitats de log dels sistemes operatius en els quals s'instal·len. Les aplicacions varien enormement en els tipus de dades que aboquen al log. En la següent llista estan alguns dels tipus de log més comuns i els potencials beneficis de cadascun:

- **Sol·licituds a servidors**, que resulten útils per reconstruir seqüències d'esdeveniments i el seu resultat. Per exemple, els servidors web registren cada URL sol·licitada i el tipus de resposta donada pel servidor. Aquesta informació es pot utilitzar per monitoritzar l'ús d'una aplicació o investigar incidents.

- **Informació relativa al compte** (com a intents d'autenticació o canvis en els privilegis del compte) permet identificar l'ús d'una aplicació pels usuaris o intents d'autenticació per força bruta.

- **Informació d'utilització** (com el nombre de transaccions per unitat de temps o la grandària de les mateixes) permet identificar usos anormals de les aplicacions.

- **Accions significatives per a l'operativa**, com a arrencada i parada d'aplicacions, fallades o canvis en la configuració. Això pot utilitzar-se per identificar problemes de seguretat o fallades operacionals.

Els logs de les aplicacions són especialment útils per a les finalitats que hem comentat, però presenten el problema que es troben, generalment, en un format propi, la qual circumstància fa que sigui més difícil el seu ús; a més les dades que contenen són relatius a un context i es requereix tenir un cert coneixement del mateix.

2.1.1.3 Programari de seguretat

Les organitzacions utilitzen diversos tipus de programari per gestionar la seguretat (tant instal·lat en els servidors com en els dispositius de xarxa) que serveixen per detectar activitat maliciosa i protegir els sistemes i les dades:

- **Programari anti-malware**. Els més comuns són els antivirus i, típicament, registren el malware detectat, desinfecció de fitxers i sistemes o quarentenes de fitxers. També poden registrar actualitzacions del propi antivirus i les exploracions en els sistemes.
- **Sistemes de detecció d'intrusió**. Els sistemes de detecció i de prevenció d'intrusions registren informació bastant detallada d'activitats sospitoses o atacs detectats, així com les accions realitzades per detenir activitats malicioses.
- **Programari d'accés remot**. L'accés remot es gestiona a través de les xarxes privades virtuals (VPN). Les VPN registren intents d'inici de sessió (*login*), tant reeixits com a fallits, així com data i hora de connexió i desconnexió dels usuaris o la quantitat de dades enviades o rebudes en cada sessió.
- **Proxies web**. Els proxies són equips intermediaris a través dels quals s'accedeix als servidors web. Per tant, els proxies poden usar-se per restringir l'accés a la web afegint una capa de protecció entre els clients i els servidors web. Els proxies registren totes les URL accedides a través d'elles.
- **Servidors d'autenticació**. Els servidors d'autenticació, com els servidors de directori o els servidors single sign-on (SSO), registren cada intent d'autenticació incloent l'origen, nom d'usuari, data i hora així com si va tenir èxit o va fallar.
- **Enrutadors (routers)**. Els routers poden configurar-se per permetre o bloquejar certs tipus de tràfic basant-se en polítiques. El log generat sol contenir les característiques més bàsiques de l'activitat bloquejada.
- **Tallafocs**. Igual que els routers, els tallafocs permeten o bloquegen el tràfic de xarxa en funció de polítiques, encara que els tallafocs utilitzen mètodes més sofisticats per examinar el tràfic de xarxa. Els tallafocs poden també inspeccionar el contingut del tràfic de xarxa i realitzar un seguiment del seu estat. Per tot això, els tallafocs generen una activitat de log més detallada que els routers.

2.2 La gestió de logs

L'enfocament convencional a la gestió dels logs seria alguna cosa així: quan s'ha produït un incident, procedir a revisar els logs en els punts d'origen. Aquesta forma de procedir és ineficient, complexa i cara a causa de la quantitat de logs que generen els diferents dispositius, fins i tot en organitzacions petites.

La gestió de logs comprèn **el procés de generar, transmetre, emmagatzemar, analitzar i eliminar els registres d'activitat o logs**.

Els beneficis que la gestió de logs pot aportar a una organització serien:

- Poder emmagatzemar els logs durant un període de temps adequat amb el suficient nivell de detall.
- La revisió periòdica i l'anàlisi poden obtenir una informació útil per resoldre més ràpidament problemes operacionals, incidents de seguretat o activitat fraudulenta.
- Identificar les tendències en l'operativa i en els problemes a llarg termini.
- Realitzar auditories i anàlisis forense.

En resum, les organitzacions necessiten disposar d'una imatge de la seva infraestructura informàtica i, per a això, necessiten una mica més que generar i recol·lectar logs. Els tècnics TIC, per la seva banda, han de configurar els seus sistemes per generar i guardar logs, així com realitzar anàlisis en temps real com a recerques i informes en profunditat.

2.2.1 Problemàtica de la gestió de logs

El problema principal en la gestió de logs consisteix a mantenir un equilibri entre una quantitat limitada de recursos amb una font de dades cada vegada major.

2.2.1.1 Generació i emmagatzematge

Com hem vist, les fonts de log són abundants i variades, la qual cosa complica la gestió en els següents aspectes:

- **Quantitat de fonts**. Els logs es troben en els servidors per tota l'organització; per tant, la gestió de logs hauria de realitzar-se en tota l'organització. A més, una font de log pot generar múltiples logs (per exemple, una aplicació pot registrar els errors en un log i l'activitat de xarxa en un altre).
- **Contingut inconsistent**. Cada font registra en cada entrada les peces d'informació que considera que són més importants. Això dificulta el procés d'enllaçat d'esdeveniments registrats per diferents fonts, posat que podrien no tenir en comú cap valor registrat (per exemple, un log registra la font amb una adreça IP i un altre registra el nom de servidor). Cada font de log representa també els valors amb diferents formats (per exemple, la data pot estar com DD-MM-AAAA o com AAAAMMDD).
- **Marques de temps inconsistents**. Cada font genera els logs utilitzant com a referència el seu rellotge intern com a marca temporal en cada entrada. Això fa que la marca de temps del log sigui inexacta, fent que el procés d'anàlisi sigui més complicat, especialment quan cal analitzar logs provinents de diferents servidors.
- **Formats de log inconsistents**. Molts logs utilitzen diferents formats com a fitxers de text separats per comes o tabuladors, bases de dades, syslog, SNMP, XML o fitxers binaris. Alguns logs estan orientats a lectors humans; uns altres, no. Alguns logs utilitzen formats estàndard, mentre que uns altres usen formats propis.

Per facilitar l'anàlisi de logs es necessita implementar alguns mètodes de conversió automàtica a una format estàndard amb una representació de camps consistent.

A causa del nombre de servidors i dels múltiples generadors de log per servidor, el nombre total de logs en una organització, fins i tot les petites, pot ser bastant alt. Per tant, el volum diari que es pot arribar a generar és considerable.

2.2.1.2 Protecció

Les organitzacions necessiten protegir la disponibilitat de les seves logs. Molts logs tenen una grandària màxima, que quan s'aconsegueix fa que es sobreescrigui l'antic log amb noves dades. Si existeixen requisits de retenció de logs es necessita establir un procés d'arxivament de logs que mantingui còpies per un període de temps major que l'originalment designat i, a causa del volum, realitzi un filtrat d'entrades que no necessitin ser arxivades.

2.2.1.3 Anàlisi

L'anàlisi de logs consisteix en l'estudi de les entrades per identificar esdeveniments d'interès. Per realitzar aquesta anàlisi, és molt convenient l'ús d'eines que automatitzin el procés d'anàlisi, especialment en la recerca de patrons ja que als humans els costa descobrir relacions com la correlació d'entrades en diferents logs relacionades amb un mateix esdeveniment.

Aquesta anàlisi és considerada pels administradors de sistemes de forma reactiva[1] en comptes de proactiva[2].

2.3 Infraestructura de gestió de logs

Una infraestructura de gestió de logs consisteix en el maquinari, programari, xarxes i mitjans utilitzats per generar, transmetre, emmagatzemar, analitzar i eliminar les dades de log.

En aquesta secció es descriu una arquitectura prototip i com els diferents components interactuen.

2.3.1 Arquitectura

Una infraestructura de gestió de log típica consta de tres capes:

- **Generació de log**. La primera capa conté els servidors que generen les dades de logs. Alguns servidors executen aplicacions client o serveis que fan que els seus logs estiguin disponibles a través de la xarxa per als servidors de la segona capa. Uns servidors permeten a altres servidors autenticar-se en ells i obtenir una còpia dels fitxers de log.

- **Anàlisi i emmagatzematge de log**. La segona capa està composta pels servidors de log que reben les dades de log o les còpies. Les dades es transfereixen a aquests servidors en temps real (o el més semblant) o s'envien per lots en moments planificats. Aquests servidors que reben els logs a partir dels generadors es denominen recol·lectors o agregadors. Les dades de log es poden emmagatzemar en els mateixos servidors o en servidors de bases de dades. Aquesta capa pot variar considerablement en complexitat i estructura.

- **Supervisió de log**. La tercera capa conté les consoles que es podrien usar per supervisar i revisar els logs, així com els resultats de les anàlisis automatitzades, i generar informes. En algunes infraestructures, les consoles realitzen també funcions de gestió dels clients i servidors de logs.

Les comunicacions entre els components de la infraestructura es realitzen normalment a través de la infraestructura de xarxa de l'organització. No obstant això, en algunes circumstàncies és possible considerar l'ús d'una xarxa separada (física o lògicament), especialment per obtenir els logs de dispositius clau.

2.3.2 Funcions

La infraestructura de gestió de logs realitza diverses funcions que ajuden en l'emmagatzematge, anàlisi i esborrament de logs. A continuació es descriuen les funcions de la infraestructura més comunes:

- **General**
  - **Reconeixement (*parsing*) de logs**. Extreure els valors de dades d'un log que poden ser utilitzats com a entrada a altres processos de gestió de logs.
  - **Filtratge d'esdeveniments**. Supressió d'entrades de log posat que les seves característiques indiquen que no contenen dades d'interès.

[1] Realitzada després que s'hagi identificat un problema per altres mitjans.

[2] Identificant sobre la marxa l'activitat i buscant símptomes de problemes.

  - **Agregació d'esdeveniments**. Entrades similars es converteixen en una única (juntament amb un comptador).

- **Emmagatzematge**

  - **Rotació**. Consisteix en el tancament d'un fitxer de log i l'obertura d'un nou quan es considera que el primer està complet. La rotació es realitza normalment d'acord amb una planificació temporal (cada hora, dia o setmana), o quan el fitxer de log aconsegueix una certa grandària.
  - **Arxivament**. Retenció dels logs durant un període de temps major de l'habitual, freqüentment per qüestions regulatòries o legals.
  - **Compressió**. Emmagatzematge dels logs de manera que es redueixi la quantitat d'espai necessari sense alterar el seu contingut.
  - **Reducció**. Eliminació d'entrades innecessàries per crear un altre log de menor grandària.
  - **Conversió**. Reconèixer un log en un format i emmagatzemar les seves entrades en un segon. La conversió, en ocasions, inclou filtrat, agregació o normalització.
  - **Normalització**. Cada camp es converteix a una representació particular. Una de les normalitzacions més freqüents és emmagatzemar dates i hores en un únic format.

- **Anàlisi**

  - **Cerca**. Trobar els logs que compleixen uns criteris determinats és la principal activitat d'anàlisi i la forma més directa d'obtenir informació d'ells.
  - **Correlació d'esdeveniments**. Trobar relacions entre dues o més entrades. La forma més comuna de correlació d'esdeveniments és la correlació basada en regles, en la qual es busca relacionar entrades procedents de diverses fonts de logs en funció de valors com la marca temporal, l'adreça IP o el tipus d'esdeveniment. La correlació d'esdeveniments també es pot realitzar amb altres mètodes com a eines de visualització o mètodes estadístics.
  - **Visualització**. Consisteix a mostrar les entrades de log en un format comprensible per humans.
  - **Informes**. Consisteix en mostrar els resultats de l'anàlisi de logs. Els informes es realitzen resumint l'activitat rellevant durant un període de temps o el registre detallat d'un esdeveniment particular o sèries d'esdeveniments.

- **Eliminació**

  - **Neteja**. Consisteix en eliminar les entrades de logs que precedeixen certa data i hora. Es realitza quan les dades velles ja no són necessàries en el sistema perquè ja no són importants o s'han arxivat.

2.3.3 Operativa i seguretat

La gestió de log és necessària tant per a qüestions de seguretat com per a l'operativa de les aplicacions i de la infraestructura informàtica. No obstant això, entre ambdues existeixen diferències pel que fa a responsabilitat, estructura organitzativa i analítica.

La infraestructura de gestió de logs hauria de ser prou flexible com per poder compartir-se entre seguretat i operativa.

2.3.3.1 Operativa

L'àrea d'operació se centra en la disponibilitat, rendiment, manteniment i aprovisionament dels sistemes informàtics. En les organitzacions, l'àrea d'operació s'organitza segons les fronteres que marquen les diferents plataformes. El principal interès de l'àrea en la gestió de log es troba en la disponibilitat d'un monitoratge, la detecció de canvis i una anàlisi que busqui l'origen dels problemes. Les qüestions més rellevants de la gestió de logs relatives a l'operativa serien:

- **Fonts de logs – Abast**. Se centra bàsicament en el monitoratge i en l'enteniment del funcionament dels components informàtics que són crítics per a l'organització.
- **Fonts de logs – Profunditat**. No hi ha una necessitat d'un registre complet d'activitat.
- **Retenció de logs**. Des d'un punt de vista operacional, els logs decreixen en valor segons van passant les hores fins que perden tota rellevància. En ocasions, és acceptable descartar els missatges de log i mantenir només el subconjunt de dades necessari per fer mètriques.
- **Analítica**. L'atenció de l'àrea està fonamentalment en la disponibilitat i en l'anàlisi de l'origen dels problemes. També es necessita monitoritzar els canvis de configuració. Com a requisit addicional estarien les necessitats de quadres de comandament i informes.

2.3.3.2 Seguretat

La seguretat en una organització se centra a mantenir la integritat i la protecció dels recursos informàtics. Això requereix una visió unificada de l'activitat dels usuaris i de l'accés als recursos. Les qüestions més rellevants de la gestió de logs relatives a seguretat serien:

- **Fonts de logs – Abast**. A causa de la incertesa sobre la forma en la qual s'ha realitzat un atac, és necessari tenir monitoritzada una àmplia varietat de recursos (més enllà dels considerats crítics).
- **Fonts de logs – Profunditat**. La investigació forense necessita un registre complet de tota l'activitat de totes les fonts ja que no és possible predir quina informació es necessitarà en cas d'un incident.
- **Retenció de logs**. La investigació i el compliment de la normativa vigent exigeixen que s'emmagatzemin les dades de log durant llargs períodes de temps (d'1 a 3 anys, normalment). Existeix, per tant, també una necessitat de preservar la integritat d'aquestes dades durant aquest període.
- **Analítica**. L'objectiu de la seguretat és descobrir patrons de potencials atacs (o ús indegut) en l'activitat anormal de la xarxa, de l'autenticació, de l'accés a les dades o del comportament dels usuaris. Per tant, existeix la necessitat de detectar canvis en la configuració i administració d'usuaris.

2.4 Eines per a gestió de logs

2.4.1 Gestió centralitzada de log

Una infraestructura de gestió de logs centralitzada, generalment basada en el protocol syslog (veure Apèndix B.1), proporciona un marc de treball senzill per a la generació, emmagatzematge i transferència de logs.

L'objectiu d'una gestió de logs centralitzada és recol·lectar els missatges en un sol servidor central. El mètode més directe per realitzar això és utilitzar el protocol syslog, que està disponible virtualment en tots els dispositius i aplicacions.

2.4.2 Software Security Information and Event Management (SIEM)

El *Software Security Information and Event Management* (SIEM) és un tipus de programari centralitzat per a gestió de logs. Els productes SIEM tenen un o més servidors que realitzen tasques d'anàlisis de logs i una o més bases de dades que emmagatzemen logs. La majoria dels productes SIEM permeten dos mètodes de recol·lectar logs des dels generadors:

- **Sense agent**. El servidor SIEM rep dades dels generadors sense necessitat d'instal·lar cap programari especial en aquests servidors. El principal avantatge d'aquest mètode és que no cal instal·lar, configurar i mantenir agents en cada servidor.

- **Basats en agent**. S'instal·la un programa agent en cada servidor que realitza tasques de filtratge d'esdeveniments, agregació i normalització, transmetent els logs normalitzats al servidor SIEM.

La funcionalitat present en les eines SIEM (veure Figura 2.1) es pot definir segons les cinc C [13]:

- **Col·lecció**. Les eines SIEM recol·lecten els logs de múltiples fonts (normalment dotzenes o centenars). El transport de log des de les fonts podria necessitar es confidencial, autenticat (per evitar falsos logs) i fiable.

- **Consolidació**. Fa referència a la normalització i agregació de logs. En alguns entorns resulta important emmagatzemar els logs en el seu format original mentre que en uns altres ho és normalitzar els logs. El propòsit de l'agregació és posar junts diferents esdeveniments que són del mateix tipus.

- **Correlació**. En el procés de correlació consisteix a reunir esdeveniments de diferents tipus que forma part d'un incident. Aquest procés és computacionalment intensiu i utilitza informació contextual, tenint com a objectiu final entendre com s'ha produït l'incident.

- **Comunicació**. Existeixen tres formes de realitzar la comunicació: enviar una alerta quan es produeixen unes determinades circumstàncies, generar un informe en moments predeterminats o revisió de l'eina en temps real.

- **Control**. Fa referència a l'emmagatzematge, ja que mentre que són necessaris per a l'anàlisi els logs s'emmagatzemen, però quan ja no són necessaris per a aquest propòsit, normalment s'arxiven. Els logs poden emmagatzemar-se normalitzats (i agregats) per accelerar el processament o emmagatzemar-se tal com es van generar per mantenir-los com a evidència.

No hi ha cap estàndard específic per SIEM, de manera que cada producte realitza els seus procediments segons el seu propi criteri.

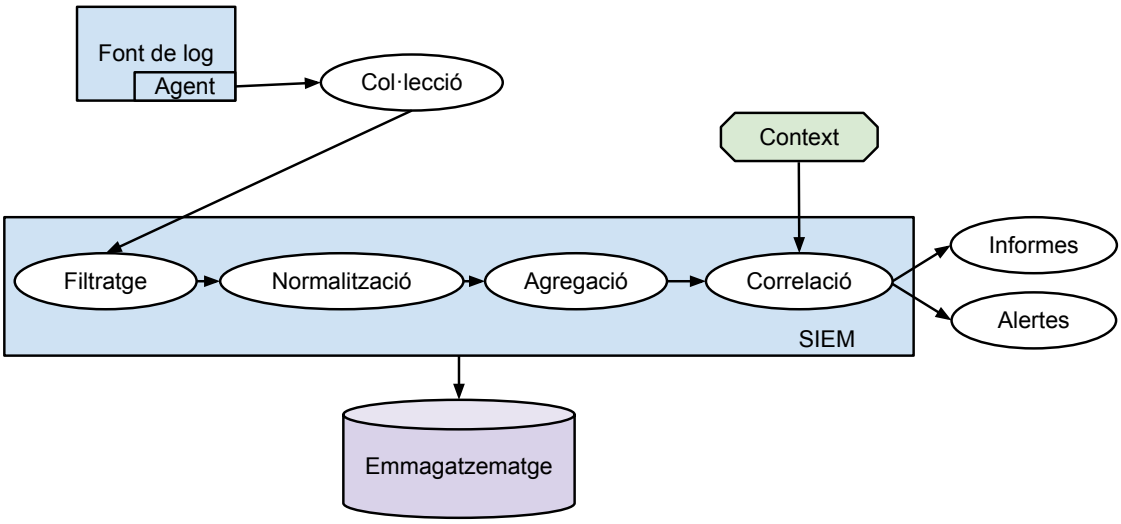

Figura 2.1: Visió general de les eines SIEM.

2.4.3 Gestió de logs en el núvol

Últimament, existeixen eines per realitzar una gestió de log en el núvol[3] (també denominat SaaS o en ocasions LaaS) que prometen simplificar la gestió, ja que l'organització pot dedicar els seus recursos i personal a les tasques pròpies, deixant la gestió dels logs en mans de tercers.

LaaS fa referència al procés de recollir, indexar i emmagatzemar els logs des de les diferents fonts de dades, sistemes operatius, aplicacions, servidors web i servidors de bases de dades de l'organització, i facilitar l'accés a ells a través d'una interfície, podent-se realitzar recerques, anàlisis i informes sobre ells.

El principal avantatge que presenta enfront de les solucions tradicionals per al tractament de logs és que es tracta d'una eina SaaS: no és necessari instal·lar ni mantenir programari en els servidors, és escalable i permet l'accés des d'un únic punt.

[3]Es defineix **computació en el núvol** com un paradigma de computació distribuïda, basat en economies d'escala, en el qual un conjunt de serveis, plataformes, emmagatzematge i capacitats de còmput virtuals, abstractes i escalables es posen a la disposició dels clients a través d'internet i sota demanda.

Capítol 3
Metodologia

"Les majors limitacions de la raça humana es deuen a la nostra incapacitat d'entendre la funció exponencial."

Albert A. Bartlett, físic.

El desenvolupament del projecte està orientat a definir una metodologia per a la gestió centralitzada de logs utilitzable en organitzacions i fent ús de productes accessibles.

Una vegada es disposa d'un marc teòric (reflectit en el Capítol 2), el següent pas és realitzar una anàlisi de la metodologia per a l'avaluació de programari (en aquest cas, d'eines de gestió de logs). Per assegurar-se que tots els aspectes rellevants es prenen en consideració, resulta convenient utilitzar un mètode o un marc de referència.

3.1 Necessitat d'una metodologia

L'avaluació de productes programari és tractada, per part de les organitzacions, d'una manera generalment informal. Aquesta avaluació sol ser duta a terme per una sola persona, sense utilitzar un mètode formal o un marc de referència. Per tant, en l'elecció del producte sol predominar l'experiència prèvia.

En comptades ocasions s'utilitzen marcs de referència o mètodes més formals que el descrit en el paràgraf anterior. Existeixen diverses raons per a això; una possible explicació és que els marcs de referència són percebuts com a rígids o no suficientment flexibles.

Un mètode d'avaluació formal assegura que es considera un rang d'alternatives més ampli, i redueix el risc de d'influència de prejudicis per part de l'avaluador.

Aquestes són les motivacions per a una revisió dels marcs de referència i la construcció d'un mètode d'avaluació.

3.2 Metodologies d'avaluació de programari

Per a una organització, l'elecció del programari per als seus sistemes d'informació, sigui de codi obert o comercial, es basa en l'anàlisi de les necessitats, en les limitacions (tècniques, funcionals i estratègiques) i en l'adequació del programari.

Existeixen diverses metodologies que ajuden a seleccionar el programari més adequat. Algunes d'elles són:

- *Open Source Maturity Model* (OSMM)
- *Open Business Readiness Rating* (OpenBRR).

- *Qualification and Selection Open Source* (QSOS) de Atos Origin.

En les següents seccions es descriuran tres marcs de referència per a avaluació de programari (millor documentats i més actius) que es troben disponibles de forma lliure i gratuïta. Les principals característiques dels tres marcs de referència considerats es troben en la Taula 3.1.

Criteri	**OpenBRR**	**QSOS**	**OSMM**
Antiguitat	2005	2006	2005
Autors	Carnegie Mellon, Intel	Atos Origin	Navicasoft
Llicència	Creative Commons	GNU Free Documentation License	Academic Free License
Model d'avaluació	Científic	Pràctic	Pràctic
Nivells de detall	2	3 nivells o més	3 nivells
Criteris predefinits	Si	Si	Si
Criteris tècnics/funcionals	Si	Si	No
Model de puntuació	Estricte	Flexible	Flexible
Escala de puntuació	1 a 5	0 a 2	1 a 10
Processo iteratiu	Si	Si	No
Ponderació de criteris	Si	Si	Si
Comparació	No	Si	No

Taula 3.1: Comparativa de marcs de referència (Font Viquipèdia).

3.2.1 Business Readiness Rating (OpenBRR)

Business Readiness Rating (OpenBRR), de 2005, és un marc de treball desenvolupat per la Universitat Carnegie Mellon i Intel. La idea principal del mètode és ponderar els factors que s'han demostrat més importants per portar a un desplegament de programari amb èxit.

El marc es basa en estandarditzar i agrupar les dades d'avaluació en categories. El procés d'avaluació s'organitza en quatre fases, al final de les quals s'obté una classificació, que és el reflex de les principals categories. Aquestes quatre fases són:

- **Avaluació ràpida**. La primera fase s'ocupa de trobar els candidats, definint els requisits dels usuaris i eliminant els candidats que no compleixen amb ells.
- **Avaluació d'objectiu d'ús**. La segona fase classifica les categories segons la seva importància, assigna a cada component un factor d'importància i defineix les mètriques per a cada categoria.
- **Recol·lecció i processament de dades**. Es recopilen les dades i s'assignen valors a cadascuna de les diferents mètriques.
- **Traducció de dades**. En l'última fase s'obté la classificació (BRR) que hauria de donar resposta a l'elecció del programari.

El marc proporciona una llista de 12 categories i un conjunt de mètriques (2–3 per categoria).

3.2.2 Qualification and Selection of Open Source software (QSOS)

El mètode *Qualification and Selection of Open Source Software*, publicat per Atos Origin en 2006 sota llicència oberta com una eina per examinar les restriccions i riscos del programari, i poder discriminar entre diferents productes candidats.

QSOS ve amb un arbre jeràrquic de criteris d'avaluació. QSOS divideix els criteris en dues seccions principals: un secció genèrica i una secció específica. La secció genèrica inclou els criteris aplicables a tots els productes programari, mentre que els criteris específics inclouen la funcionalitat esperada que, òbviament, varia segons la família dels productes.

QSOS és un marc que consta de quatre passos independents, que es realitzen iterativament fins a refinar el procés:

- **Definició**. La primera fase consisteix a definir el marc de referència. Els marcs de referència són les famílies de programari, els tipus de llicència i els tipus de comunitat.

- **Avaluació**. En la següent fase el producte s'avalua i s'obté, per a cada producte, una Targeta d'Identitat[1] consistent en dades concretes com a nom, tipus, llicència, documentació i comentaris. Aquestes dades són la base per realitzar la Fulla d'Avaluació[2], on es puntua cada criteri amb un valor que és un nombre enter entre 0 i 2. Aquesta puntuació s'atorga sense tenir en compte el context de l'organització.

- **Qualificació**. La tercera fase tracta d'indicar les necessitats de l'organització; per a això, es defineixen filtres avaluant les necessitats i limitacions relacionades amb la selecció d'un producte per al context donat.

- **Selecció**. L'objectiu d'aquesta fase és identificar el programari segons els requisits de l'organització o, més en general, comparar productes de la mateixa família.

La web de QSOS té criteris per a algunes famílies de programari; desgraciadament no existeix una família per al programari de gestió de logs.

3.2.3 Open Source Maturity Model (OSMM)

El *Open Source Maturity Model* és un marc de referència de 2005 realitzat per Navica, consultora de codi obert. L'objectiu del model és obtenir el producte que, donades les necessitats úniques de cada organització, satisfaci aquestes necessitats. Com indica el nom, el model intenta determinar la maduresa del producte.

Els dos requisits fonamentals de les organitzacions pel que fa al programari són: productes madurs (equivalent a productes d'alta qualitat) i que el producte sigui completament funcional.

El marc OSMM consta de tres fases que proporcionen un conjunt formal de criteris d'avaluació. En la primera fase es defineixen i identifiquen els elements clau del producte i s'avalua la maduresa de cada element assignant una puntuació (entre 0 i 10). En la segona fase s'assignen pesos als elements (per defecte o personalitzats). Mitjançant la puntuació de maduresa i els pesos, s'obté com a resultat final una puntuació de maduresa general (en una escala del 0 al 100).

[1] Targetes d'Identitat (Card ID): tipus de document utilitzat per QSOS que registra informació general dels diferents sistemes.

[2] Fulles d'Avaluació: tipus de document utilitzat per la metodologia QSOS que registra informació més detallada que les Targetes d'Identitat.

3.2.4 Mètode utilitzat

En revisar aquests marcs i enfocaments resulta que hi ha diverses maneres d'avaluar i que hi ha diverses formes de determinar el producte més adequat per a una organització determinada. Els tres marcs descrits utilitzan un enfocament pas a pas, que es pot reduir a un enfocament genèric que consta de quatre etapes:

- Seleccionar productes candidats.
- Definir un conjunt de criteris d'avaluació.
- Obtenir dades dels productes candidats.
- Avaluar el producte segons aquests criteris (i obtenir una puntuació).

La metodologia seguida s'inspira en el mètode QSOS v1.6, que està dissenyat per qualificar, seleccionar i comparar programari d'una manera objectiva, traçable i argumentada. No obstant això, QSOS presenta restriccions importants: està basat en XML i, per tant, el tractament de les dades resulta complex i tediós, dificultant la integració de les dades. A més, encara que és possible afegir comentaris a les avaluacions, el sistema de puntuació és tancat.

El mètode descrit comença amb la definició dels criteris d'avaluació, com en OpenBRR. Els escenaris d'ús serveixen per identificar requisits i criteris d'avaluació relacionats amb ells. La resta dels requisits s'obté de la funcionalitat descrita en el Capítol 2, així com en els Apèndixs A i B.

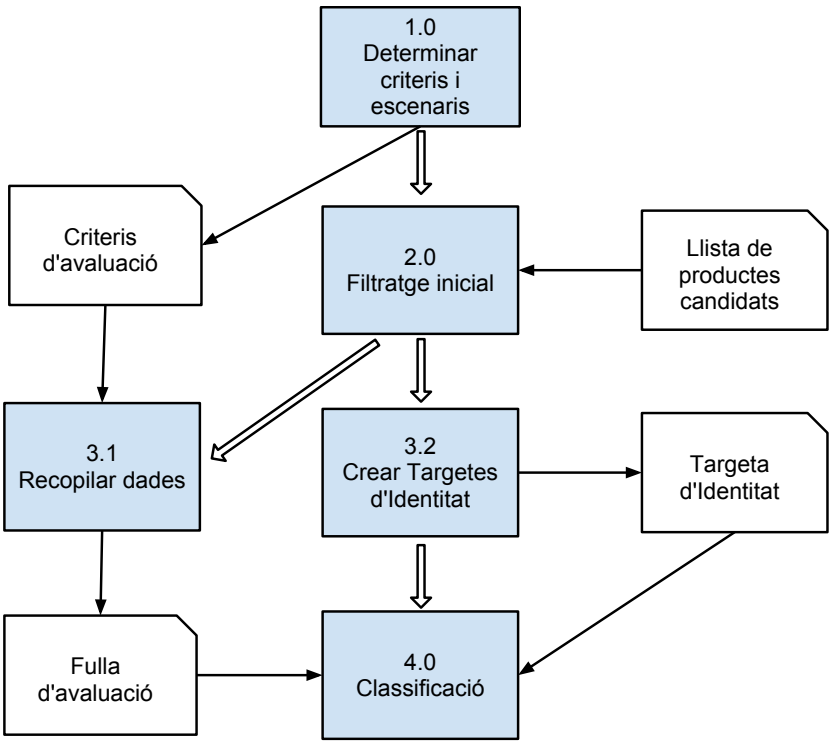

Figura 3.1: Mètode d'avaluació per a gestió de logs

El següent pas del mètode és obtenir els productes candidats. Com no existeix una llista de candidats, és necessària una cerca de productes esmentats en articles i en revistes. Partint d'una llista inicial de candidats, es realitza un filtrat eliminant productes que no són fàcilment avaluables o que no tenen activitat recent. La llista de productes candidats ha de ser adequada, la qual cosa vol dir que hi hagi un nombre de propostes suficient i que sigui abastable.

Els conceptes de Targetes d'Identitat (Card ID) i Fulles d'Avaluació, procedents de QSOS, es van a utilitzar per descriure com els productes candidats s'adeqüen (o no) als requisits segons els criteris d'avaluació.

El mètode consisteix en les següents quatre fases, tal com s'il·lustra en la Figura 3.1:

1. **Determinar criteris i escenaris**.
2. **Obtenir una llista de productes candidats i realitzar un filtratge inicial**.
3. **Crear una Targeta d'Identitat i una Fulla d'Avaluació dels productes candidats**.
4. **Ordenar els productes per puntuació i realitzar una classificació manual**.

El mètode descrit és el principi bàsic de valoració dels productes candidats. Encara que el mètode es basa en un conjunt de criteris classificats i organitzats que són directrius en la classificació dels productes, la valoració humana és un factor molt important i, per aquest motiu, el mètode realitza una classificació manual basada en la puntuació obtinguda, en comptes d'una puntuació exclusivament matemàtica.

Capítol 4
Procés d'avaluació

"Esperar que la vida et tracti ben perquè ets bona persona és com esperar que un toro no t'ataqui perquè ets vegetarià."

Héctor del Mar, comentarista esportiu

Tal com es descriu en la Figura 3.1, en aquest Capítol es va a desenvolupar el mètode esmentat en l'Apartat 3.2.4, de manera que tots els passos rellevants del procés es van a descriure detalladament. L'objectiu no és seleccionar un producte concret per a una organització, sinó valorar els productes independentment del context en el qual es vagi a usar. No obstant això, per completar la metodologia, i igual que en QSOS, una organització concreta té la possibilitat de ponderar els criteris (per exemple: prescindible, important o essencial) per reflectir el context particular d'utilització i, per tant, modular els resultats d'acord amb les necessitats concretes de cada organització.

4.1 Escenaris i requisits

El primer pas consisteix a determinar els escenaris en els quals s'aplicarà l'avaluació i, a partir d'ells, intentar extreure els requisits rellevants. Els escenaris que anem a considerar són:

- **Escenari 1**. Una organització interessada en la gestió centralitzada de logs per a l'operativa diària de la seva explotació.
- **Escenari 2**. Una organització interessada en la gestió centralitzada de logs per a qüestions de seguretat.

Aquests escenaris constitueixen el punt de partida de l'avaluació. Cada escenari està associat a diversos requisits, i cada requisit pot estar associat a més d'un escenari.

Els criteris d'avaluació obtinguts, així com les puntuacions aplicables a cadascun d'ells, es troben recollits en l'Annex C. Conforme a aquests criteris es realitzarà l'avaluació.

4.1.1 Requisits generals

Els requisits generals fan referència a la fiabilitat, el manteniment o la explotabilitat.

La **fiabilitat** és un factor molt important. Els elements que posen de manifest el grau de fiabilitat d'un producte es troben entre els criteris de "Maduresa" (dins de "Durabilitat") proposats per QSOS.

També és un punt important l'existència d'una comunitat d'usuaris que utilitza i millora els productes. L'existència d'una comunitat activa i nombrosa és important perquè, probablement, sense ella el

producte no tindrà gaire futur, la qual cosa implica que no hi haurà noves versions ni pegats que corregeixin fallades. QSOS presenta una sèrie de criteris relatius a l'avaluació de la comunitat; aquests criteris es troben dins de l'epígraf "Adopció".

La facilitat d'ús i d'administració són dos criteris que ens mesuren com de fàcil és d'utilitzar i administrar el producte, és a dir, la seva **explotabilitat**.

La **mantenibilitat** es defineix com la facilitat que presenta un producte per ser modificat, corregit, millorat o adaptat a un entorn canviant. QSOS té una sèrie de criteris per mesurar la mantenibilitat d'un producte. Aquests criteris (denominats "Adaptabilitat Tècnica") es refereixen al codi font, la documentació tècnica o la documentació per als desenvolupadors. Per comprovar aquest requisit es fa necessari revisar aquesta documentació i inspeccionar el codi font (quan està disponible), o usar webs (com ohloh.net) en les quals és possible obtenir estadístiques del codi.

4.1.2 Requisits funcionals

Els requisits funcionals estan estretament relacionats amb la funcionalitat dels productes i, per tant, són els criteris més importants. Per investigar com els diferents productes candidats satisfan els requisits (expressats en forma de criteris d'avaluació), en cada producte s'examina la documentació, els manuals de desenvolupament i també es procedeix a realitzar una instal·lació de prova. Com s'ha comentat en el Capítol anterior, QSOS no té cap criteri establert per a la família de programari de gestió de logs; per tant, la forma de definir-los és utilitzar el coneixement sobre la qüestió per determinar una sèrie de requisits que hauria de satisfer un programari de gestió de logs.

Per al cas que ens ocupa, els requisits funcionals s'han dividit en tres grups.

Els requisits d'**entrada de dades** tenen relació amb la centralització dels logs, és a dir, quins formats accepta l'eina i com es transporten des dels diferents dispositius fins al concentrador de logs[1]. Evidentment quants més formats pugui suportar l'eina (major funcionalitat), resulta més convenient i completa; però com és virtualment impossible suportar tots els possibles formats existents (i futurs), en els criteris apareixen els més coneguts, així com la possibilitat d'afegir nous formats[2]. En el transport de logs existeix un estàndard de facto suportat per la majoria dels dispositius (directament o mitjançant eines de tercers), que és syslog (i variants). A més, algunes eines disposen d'agents que poden realitzar la mateixa funció i afegir alguna funcionalitat (filtratge, normalització o agregació).

L'**emmagatzematge** és un altre requisit important, ja que registra com es van a emmagatzemar els logs una vegada rebuts en els concentradors de logs. En qüestions de seguretat, el poder determinar l'origen de cada log i mantenir-ho inalterat són factors importants per al seu ús com a evidència en cas d'anàlisi forense. També s'intenta determinar el suport sobre el qual s'emmagatzemen tots aquests logs. Tradicionalment es feia en fitxers, però actualment es fa en bases de dades[3](del tipus que siguin) per agilitar la recuperació.

Els requisits de **cerca i anàlisi** són el punt central de qualsevol eina de gestió de logs. Les capacitats de cerca es troben relacionades amb l'emmagatzematge i la recuperació de logs, i s'expressen mitjançant

[1]Aquesta no és necessàriament l'arquitectura real, car aquest concentrador pot tractar els logs o reexpedir-los cap a altres servidors, depenent de com estiguin constituïdes les diferents capes de la infraestructura de logs.

[2]Afegir nous formats pot esdevenir complicat, ja que –a més del coneixement de la sintaxi de les entrades de log– sol necessitar-se coneixement de programació i tractar amb expressions regulars.

[3]Les bases de dades representen informació d'estat com, per exemple, inventaris, reserves aèries o dades de persones. A més, cada registre té una sèrie d'atributs (nom, adreça, salari, ...), i són els valors d'aquests atributs els que varien de registre en registre i evolucionen amb el temps. El conjunt d'atributs es defineix en dissenyar la base de dades i, per tant, aquestes dades es denominen estructurats. D'altra banda, els logs són fitxers de text no estructurat o semi-estructurat associats a una marca temporal en la qual l'esdeveniment succeeix. A causa de l'organització temporal, del volum i del caràcter estàtic de les entrades de log, les tècniques de bases de dades tradicionals són poc pràctiques per al tractament d'aquest tipus de dades.

el llenguatge de consulta[4]. Els llenguatges de consulta tenen una sintaxi i operadors per cobrir les diferents funcionalitats que es necessiten en la gestió de logs. Els operadors poden agrupar-se lògicament en dos grups: els que operen sobre un esdeveniment i els que operen sobre conjunts d'esdeveniments (agregació).

Per tenir una visió holística del que succeeix amb els logs, són de gran ajuda unes eines de visualització i generació d'informes. Resulta difícil definir requisits sobre aquesta qüestió, perquè poden abastar des de simples representacions fins a elaborats diagrames. Per tant, els requisits quant a informes i visualització tracten de mesurar l'existència o absència d'unes característiques aplicables[5] a la gestió de logs.

Per poder determinar si una sèrie de logs estan relacionats, es pot recórrer a tècniques de correlació d'esdeveniments. Existeixen múltiples estratègies per establir correlació d'esdeveniments que permeti, a partir d'esdeveniments procedents de diferents fonts, generar un grup d'esdeveniments que puguin servir d'aproximació a l'origen d'un problema o detectar intents hostils. Encara que en la literatura es poden trobar múltiples formes de realitzar la correlació, la principal és utilitzar les marques de temps[6] dels esdeveniments per poder correlacionar-los.

Els últims requisits fan referència als quadres de comandament i a les alertes. Aquest tipus de funcionalitat no pot considerar-se obligatòria, però pot ser d'interès a l'hora de gestionar organitzacions de grandària mitjana o gran. Els quadres de comandament (*dashboard*) mostren de forma predefinida i integrada visualitzacions de les dades que resulten d'interès o que interessa ressaltar; per tant, són útils per tenir una visió global d'un determinat sistema o servei des d'un sol punt d'entrada. Les alertes són avisos que s'envien a certs usuaris quan concorren certes condicions o de forma planificada.

4.1.3 Requisits tècnics

Els requisits tècnics inclouen qüestions com la facilitat d'instal·lació, configuració o actualització, la interoperabilitat, la seguretat o la interfície gràfica d'usuari (GUI).

Evidentment, les **facilitats d'instal·lació, configuració o actualització** que presenta un producte, així com l'absència de **dependències respecte de llibreries externes** (que també cal instal·lar, configurar i actualitzar), són factors que indiquen la qualitat del mateix.

La **interoperabilitat** és l'habilitat de dos o més components per intercanviar informació. Per tant, mesura la capacitat que té un producte per "comunicar-se" amb l'exterior. S'han considerat només dues formes de comunicació[7]: mitjançant una API i mitjançant serveis web, que són mecanismes estàndard no excloents.

Les qüestions de **seguretat** es contemplen des de dos punts de vista: autenticació[8] i autorització [9].

Un dels requisits tècnics més habituals és la plataforma (pot ser un sistema operatiu, llibreries, o entorns virtuals com Java) sobre la qual s'executa un producte. El requisit més rellevant solia ser el sistema operatiu[10]; però, avui dia, aquest requisit té menys interès a causa dels avanços tecnològics: els entorns de virtualització (que permeten executar qualsevol sistema operatiu), i el SaaS[11]. Per aquestes

[4] El llenguatge de consulta més conegut és SQL, que exigeix que les dades estiguin completament estructurats; però les dades de logs són semi-estructurades, i no són una base de dades en el sentit normatiu.

[5] Una revisió completa de la visualització com a eina d'anàlisi es pot trobar en [21].

[6] Normalment, cada entrada de log porta una marca de temps (*timestampt*) i és l'únic element d'informació que es pot usar per correlacionar diversos tipus de logs

[7] Existeixen centenars de formes d'intercanviar informació entre dos sistemes: des de la simple transferència de fitxers per FTP fins a complicats esquemes que usen una varietat de tècniques com CORBA o sockets oberts.

[8] Autenticació, en aquest context, significa que els usuaris del producte han de demostrar que són qui diuen ser.

[9] Autorització és el procés que permet atorgar el dret per utilitzar els recursos

[10] Resulta d'interès que el producte es trobi disponible per al sistema operatiu que s'utilitza en l'organització.

[11] Amb SaaS no cal preocupar-se de qüestions com la plataforma o les actualitzacions.

raons, en aquest estudi no s'ha inclòs la plataforma com a requisit en l'avaluació.

Tampoc no s'han considerat qüestions relatives al rendiment. El motiu és que es tracta de criteris difícils de valorar i que depenen de les condicions (volum de dades, complexitat de la consulta, etc.) en les quals es mesuri aquest rendiment.

4.2 Filtratge inicial

El següent pas en el mètode de l'avaluació és un filtratge inicial de productes candidats. Com a punt de partida es va utilitzar la llista de productes que apareix al final dels articles [27] i [16] i que es reprodueix en la Taula 4.1.

HP ArchSight (Logger)	Logentries	Liquidlabs Logscape
Loggly	LogLogic	LogRhythm
Sawmill	Splunk	XpoLog XpoSearch
Apache Chainsaw	Graylog2	Logstach
Octopussy		

Taula 4.1: Llista inicial de productes.

D'aquesta llista s'eliminen aquells productes que no presenten activitat recent, no disposen de versions avaluables gratuïtes, o només cobreixen una part de tot el procés de gestió de logs. En certa manera, aquesta selecció és manual i s'ha donat preferència a les opcions més noves enfront de les més clàssiques; així, han resultat seleccionades les dues opcions basades en el paradigma de computació en el núvol. El nombre final de productes candidats (veure Taula 4.2) s'ha mantingut en un nombre limitat per poder dur a terme una avaluació prou completa.

Graylog2
Logentries
Loggly
Octopussy
Splunk

Taula 4.2: Llista de productes candidats

4.3 Recol·lecció de dades

El tercer pas del mètode d'avaluació és recopilar la informació necessària per emplenar la Targeta d'Identitat i la Fulla d'Avaluació.

La Targeta d'Identitat i la Fulla d'Avaluació estan basades en les de QSOS. La Fulla d'Avaluació reflecteix els criteris d'avaluació prèviament determinats, juntament amb les puntuacions assignades. Alguns dels criteris d'aquesta Fulla poden ser subjectius, mentre que uns altres són de naturalesa mesurable. En alguns casos, les dades no es troben disponibles, o són el resultat d'una estimació. A més, tots els productes candidats es troben en permanent evolució; per tant, les puntuacions en diversos criteris podrien canviar amb el temps. A causa d'aquestes limitacions, altres avaluacions posteriors podrien no produir un resultat idèntic, circumstància que hauria d'entendre's sota aquestes premisses.

A continuació es presenta una breu descripció dels productes candidats, així com les corresponents Targetes d'Identitat i els punts forts i febles de cadascun d'ells.

4.3.1 Octopussy

Octopussy és un concentrador que combina logs de diferents fonts en un emmagatzematge central. Octopussy és una solució de programari lliure per manejar logs que bàsicament els emmagatzema, genera informes i alertes.

Aquest producte està orientat a petites o mitjanes organitzacions que desitgin una gestió centralitzada de logs. Octopussy encaixa en una infraestructura basada en syslog (que és una forma suficientment provada de crear i transportar logs). Els esdeveniments de Windows també poden ser tractats amb Octopussy, però prèviament han de ser transformats mitjançant productes de tercers (com, per exemple, *Snare*).

Informació general	
Nom	Octopussy
Descripció	Octopussy és una eina d'anàlisi de logs, generació d'informes, alertes i gràfics (amb l'eina RRD).
Origen	Desconegut.
Pàgina web	http://www.8pussy.org
Llicència	GNU General Public License (GPL)
Sistemes Operatius	Linux
Aspectes funcionals i tècnics	
Versió	1.0rc5
Tecnologia	Perl/XML
Pre-requisits	Apache, Perl i MySQL
Funcionalitat	Anàlisi de logs, alertes, informes i visualització
Seguiment d'errors	http://sourceforge.net/projects/syslog-analyzer/
Suport i ajuda	
Documentació	http://www.8pussy.org/dokuwiki/doku.php?id=documentation
Suport gratuït	http://sourceforge.net/projects/syslog-analyzer/

Taula 4.3: Targeta d'Identitat de Octopussy

Octopussy s'autodefineix com un analitzador de logs Perl/XML, la qual cosa és una mica enganyosa i incompleta. Ofereix processament de logs, correlació d'esdeveniments sense estat i anàlisi, a més d'un mecanisme d'alerta integrat (prestació important, car suposa l'automatització i el consegüent estalvi de temps). Es poden crear alertes depenent de diverses restriccions, i l'eina d'informes pot generar informes diaris, setmanals o mensuals que mostrin la disponibilitat, avisos, errors o simplement la utilització dels diferents dispositius. Tots els components de Octopussy poden considerar-se com una peça de programari oberta que pot ser extensible.

Com a gran inconvenient, convé ressenyar que la documentació del producte és realment bastant pobre.

Punts forts
• Integració amb eines de monitoratge, com Zabbix o Nagios.
• Possibilitat de personalització.
• Gestió d'alertes.
• Programari lliure.
Punts febles
• Utilitza MySQL com a repositori que presenta limitacions per a la gestió de logs.
• Documentació.
• Procés d'instal·lació.

Taula 4.4: Punts forts i febles de Octopussy

4.3.2 Splunk

La versió de Splunk gratuïta (per a desplegaments de petita o mitjana grandària) indexa com a màxim 500MB de log al dia. A més, no estan disponibles altres característiques com alertes, control d'accés per rols, cerques distribuïdes o execució d'aplicacions premium.

Splunk presenta una complexitat interna que no s'oculta (gràcies a una extensa documentació). Per fer-ho funcionar realment, es fa necessari abandonar aviat la interfície d'usuari i introduir-se en el funcionament tècnic de l'eina, que implica editar fitxers de configuració, escriure expressions regulars i entendre com Splunk recol·lecta i indexa les dades.

No només s'indexen les dades; també s'intenta analitzar sintàcticament (*parsing*) els logs. Això es pot fer "a mà": utilitzant expressions regulars, cerques, i afegint etiquetes, la qual cosa permet que Splunk entengui qualsevol tipus de log. Però existeix una alternativa, que constitueix un dels principals avantatges del producte: la biblioteca d'accessoris (*add–on*) denominada Splunkbase[12].

Splunk còpia (intencionadament) la barra de cerca minimalista de Google; per trobar informació simplement cal començar a teclejar, seleccionar el rang temporal i prémer el botó de cerca. Una altra característica que Splunk còpia de Google és la velocitat: immediatament després de prémer el botó de cerca, els logs que contenen el text que s'ha teclejat comencen a aparèixer mentre la consulta continua executant-se en segon pla.

Les cerques simples de text són només una petita part del que les cerques de Splunk poden fer. El manual de cerques té gairebé 300 pàgines i conté informació dels 125 comandos: alguns fàcils d'entendre (com "search", "sort" o "top"); uns altres, interessants (com "rare", "dedup" i "transaction"[13]); i uns altres, complicats i difícils d'usar (com "xmlkv"[14] i "bucket").

Splunk utilitza, durant l'execució de les consultes, un mecanisme MapReduce per obtenir les dades necessàries per a cada sol·licitud. Aquest mecanisme[15] és completament transparent per als usuaris.

[12] La majoria de les aplicacions de Splunkbase, encara que no totes, són gratuïtes.

[13] Agrupa entrades de log en una sola transacció.

[14] Extreu parells atribut/valor a partir de dades en format XML.

[15] La traducció de la consulta a l'estratègia d'execució es fa automàticament, com es descriu en [4].

Informació general	
Nom	Splunk
Descripció	Splunk és un producte de cerca d'informació que indexa els logs procedents de la infraestructura TIC i permet als usuaris analitzar, generar alertes i informes, tot en un únic lloc.
Origen	Splunk Inc.
Pàgina web	http://www.splunk.com
Llicència	Free Licence. Enterprise Licence
Sistemes Operatius	Unix (Linux, FreeBSD, AIX, Solaris), Windows (32 i 64 bits) i MacOSX
Aspectes funcionals i tècnics	
Tecnologia	Python (Interfície web sobre CherryPy), Javascript
Pre-requisits	Cap
Funcionalitat	Recol·lecció centralitzada de logs, cerca, anàlisi, generació d'informes i alertes.
Seguiment d'errors	http://www.splunk.com/support
Suport i ajuda	
Documentació	http://docs.splunk.com/Documentation
Suport gratuït	http://www.splunk.com/support

Taula 4.5: Targeta d'Identitat de Splunk

Punts forts
• La potència del llenguatge de consulta. • Cerques en temps real. • Aplicacions (Splunkbase). • Documentació i comunitat.
Punts febles
• Les limitacions de la versió Free i el preu de la versió Enterprise. • Característiques d'emmagatzematge i arxivament bàsiques. • Sintaxi complexa.

Taula 4.6: Punts forts i febles de Splunk

Els informes, en forma de quadres de comandament i visualitzacions dels logs, són una part molt potent de Splunk i una valuosa eina d'anàlisi. Es poden generar diversos tipus d'informes: de gràfics simples (de pastís o de barres), així com textuals o tabulars. Comparada amb altres eines de generació

d'informes, Splunk no és la més potent, però és més que suficient per a les necessitats habituals. Cada informe es mostra dins d'un quadre de comandaments que permet a l'usuari grans possibilitats de personalització, aprofundiment (*drill–down*) en les dades o canvi de rangs temporals.

El sistema d'alertes només està disponible en la versió premium, però no és una característica especialment important. Les alertes només es poden generar basant-se en les consultes; i una vegada definides, es poden enviar via email, RSS feed o un *script*.

Per Splunk les aplicacions són una qüestió important, probablement perquè les aplicacions són claus per al futur del producte. De fet, l'eina bàsica de cerca és una aplicació pre-instal·lada denominada "Search". Les aplicacions de Splunkbase cobreixen tota la gamma: des de cerques millorades a generadors d'informes o gràfics, canvis en l'aparença (*look–and–feel*), així com les que tracten logs de dispositius concrets. Les aplicacions de Splunkbase són una de les millors característiques de Splunk, i que ho distingeixen d'altres eines. Gràcies a elles, Splunk és considerat l'estàndard per a l'anàlisi i la gestió de logs.

4.3.3 Graylog2

Graylog2 és un producte per a la recol·lecció centralitzada de missatges de log i el processament dels mateixos. Consisteix en un servidor que recol·lecta dades (graylog2–server), i una aplicació web (graylog2–web–interface) que permet cerca i visualització (veure Figura 4.1). Pot utilitzar missatges de syslog, així com els generats amb el seu propi format, denominat GELF[16] (Graylog Extended Log Format).

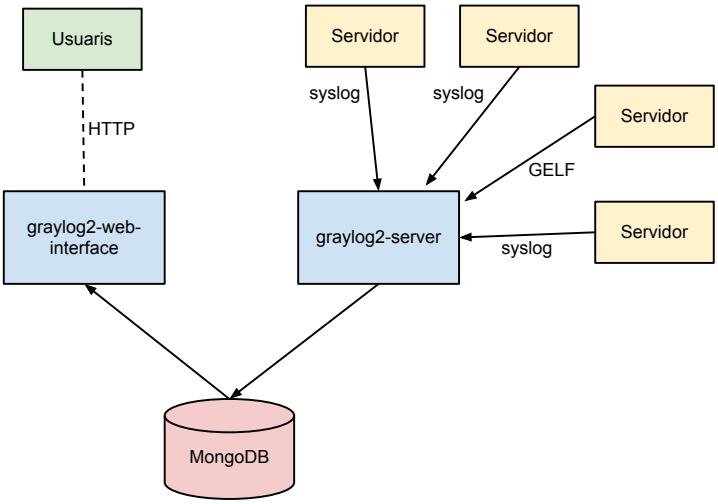

Figura 4.1: Arquitectura de Graylog2

Per a la persistència de les dades, Graylog2 utilitza un emmagatzematge amb índexs no basat en esquema (*schemaless*); és a dir, usa MongoDB[17]. Els principals avantatges d'usar MongoDB són que té

[16]Graylog Extended Log Format és una especificació per a missatges de log i altres dades de les aplicacions que està dissenyat directament per a dades estructurades: objectes JSON de parells atribut–valor comprimits i enviats al servidor usant UDP, TCP o AMQP (Advanced Message Queuing Protocol).

[17]En el moment d'escriure aquest Projecte, la versió 0.9.6 de GrayLog2 és a punt de ser llançada i deixarà d'utilitzar

Informació general	
Nom	Graylog2
Descripció	Graylog2 és una solució de gestió de logs gratuïta i oberta que permet un repositori centralitzat i l'accés als logs de la infraestructura.
Origen	El Graylog original va ser una eina interna desenvolupada per Lennart Koopmann com una interfície entre rsyslog i una base de dades MySQL (amb una interfície web en Ruby on Rails). Va començar a desenvolupar Graylog2 (2010) perquè la dependència de Graylog amb MySQL ho feia terriblement lent.
Pàgina web	http://graylog2.org/
Llicència	Apache 2.0
Sistemes Operatius	Linux (Debian, Ubuntu i CentOS)
Aspectes funcionals i tècnics	
Versió	0.9.5
Tecnologia	Java, Ruby i Javascript
Pre-requisits	MongoDB, Java JDK, RubyGems, Apache
Funcionalitat	Recol·lecció centralitzada de log, tractament, recerca i visualització.
Seguiment d'errors	http://jira.graylog2.org
Suport i ajuda	
Documentació	https://github.com/Graylog2
Suport gratuït	http://groups.google.com/group/graylog2?hl=en

Taula 4.7: Targeta d'Identitat de Graylog2

una API molt senzilla d'usar, un rendiment molt bo, i permet limitar la grandària de la col·lecció (respecta l'ordre d'inserció i elimina automàticament missatges antics). Per generar els índexs cal connectar-se directament a la interfície d'usuaris (*shell*) de MongoDB i executar allí els comandos necessaris.

Per avaluar els missatges entrants[18] segons unes regles definides per l'administrador, Graylog2 utilitza Drools Expert. Es tracta d'un motor de regles que ens permet fer el que vulguem amb els logs, però cal definir les regles i mantenir-les.

Totes les dades emmagatzemades amb Graylog2 apareixeran en la interfície web. Des d'ella, es poden buscar i filtrar dades. Una part essencial de la interfície web són els "streams", que són bàsicament cerques emmagatzemades que permeten l'accés ràpid a dades pre-filtrades. És possible monitoritzar i generar alertes des dels "streams".

Les alertes sobre "streams" es disparen quan el nombre de missatges d'un "stream" durant un temps donat aconsegueix un màxim. L'alerta consisteix en un correu enviat als usuaris subscrits o a tots. Internament, per realitzar la seva tasca les alertes requereixen d'un procés *cron* del sistema operatiu.

MongoDB per a l'emmagatzematge de missatges (encara que ho seguirà usant per al comptador de missatges), i començarà a usar ElasticSearch per agilitar les lectures i permetre cerques a text complet més ràpides.

[18] Cada missatge s'avalua abans d'inserir-se en MongoDB

Punts forts

- GELF i Drools Expert permeten configuracions flexibles.
- MongoDB, com a repositori, té un bon rendiment.
- Aplicació web senzilla per visualitzar logs.
- Programari lliure.

Punts febles

- Es fa necessari aprendre MongoDB.
- Complicat obtenir el text dels logs.
- Documentació.

Taula 4.8: Punts forts i febles de Graylog2

Llistat 4.1: Exemple de regla amb Drools Expert

```
import org.graylog2.messagehandlers.gelf.GELFMessage

rule "Llevar trafic UDP i ICMP tallafocs"
    when
        m : GELFMessage( fullMessage matches "(?i).*(ICMP|UDP) Packet(.|\n|\r)*" && host=="firewall" )
    then
        m.setFilterOut(true);
        System.out.println("[Llevar trafic UDP i ICMP tallafocs] : " + m.toString() );
end
```

4.3.4 Loggly

Loggly constitueix una altra forma d'aproximar-se al problema de la gestió de logs. En lloc de mantenir els logs dins de la infraestructura de l'organització, els logs s'envien a Loggly perquè els gestioni. Això pot suposar un estalvi de costos, car no és necessari mantenir servidors ni emmagatzematge. Una vegada que es disposa d'un compte en Loggly, simplement cal reexpedir els logs als servidors (via syslog-ng o un altre mecanisme). Quan els logs arriben a Loggly, es divideixen en elements segons la marca temporal i s'indexen. Aquests índexs són els que permeten realitzar cerques.

Per realitzar tot això, Loggly se serveix de la infraestructura AWS (Amazon Web Services). A més d'aquesta infraestructura, utilitza un sèrie de tecnologies per a cada qüestió. Així:

- syslog-ng com a entrada de syslog/TLS.
- Node.js para entrada d'HTTP/HTTPS.
- 0MQ per a cues d'esdeveniments i distribució de treballs.
- S3 bucket per a emmagatzematge local i arxivatment.
- MongoDB per a estadístiques.
- SolrCloud per a cerques escalables.

Informació general	
Nom	Loggly (Hoover Loggly)
Descripció	Plataforma de log basada en el núvol. Loggly recol·lecta i centralitza els logs i permet cerques mitjançant una interfície web senzilla.
Origen	Loggly comença l'any 2009 com una *start-up* localitzada en Sant Francisco.
Pàgina web	http://loggly.com/
Llicència	De proves (30 dies). Política de preus en funció de volum diari i retenció.
Sistemes Operatius	Cap.
Aspectes funcionals i tècnics	
Versió	Beta
Tecnologia	Python
Pre-requisits	Cap.
Funcionalitat	Aplicació web que emmagatzema, reté, realitza cerques i genera alertes i informes de logs.
Seguiment d'errors	http://loggly.com/support/
Suport i ajuda	
Documentació	http://wiki.loggly.com
Suport gratuït	http://loggly.com/support

Taula 4.9: Targeta d'Identitat de Loggly

Punts forts

- Perfecta solució per a organitzacions petites que poden obtenir gestió de logs amb poc esforç.
- Servei senzill i directe: no cal reinventar la roda.
- Eines de cerca i visualització.
- Loggly permet acceptar logs mitjançant HTTP i una API (ideal per a programadors).

Punts febles

- Poca maduresa del producte.
- Molts formats de log no estan suportats.
- Les alertes no funcionen correctament (Alert Birds en beta).

Taula 4.10: Punts forts i febles de Loggly

- Django/Python per a desenvolupament d'aplicacions.

La interfície de cerca de Loggly és la clàssica pantalla de tipus línia de comandos (CLI) mitjançant la qual es poden filtrar els logs per temps, servidor, tipus, paraules clau i altres criteris. El comando bàsic (i gairebé únic) és "search". En poder posar junts els logs de múltiples dispositius i fonts, l'usuari (desenvolupador o tècnic de sistemes) pot trobar correlació entre esdeveniments que tinguin relació amb un incident.

Com a aplicació, realment Loggly està més orientada a desenvolupadors, ja que permet buscar, monitoritzar i analitzar logs de múltiples fonts, tant procedents d'aplicacions tradicionals com a aplicacions basades en el núvol[19].

4.3.5 Logentries

Logentries és també una solució SaaS per a la gestió de logs, de concepte molt similar a Loggly.

Informació general	
Nom	Logentries
Descripció	Servei de gestió i anàlisi de logs.
Origen	En 2008, Viliam Holub i Trevor Parsons estaven treballant en un projecte amb IBM en el grup d'enginyeria de rendiment de l'University College Dublin. Com a subprojecte van crear un motor de correlació en temps real que recol·lectava, correlacionava i presentava logs per a sistemes de prova. Com els va semblar interessant, van desenvolupar en 2010 un motor de gestió i anàlisi de log basat en el núvol.
Pàgina web	http://logentries.com/
Llicència	De proves (30 dies).
Sistemes Operatius	Cap
Aspectes funcionals i tècnics	
Versió	Desconeguda (actualització online)
Tecnologia	Python
Pre-requisits	Cap.
Funcionalitat	Recol·lecció de logs en temps real, emmagatzematge, anàlisi i visualització
Seguiment d'errors	hello@logentries.com
Suport i ajuda	
Documentació	https://logentries.com/doc/search/
Suport gratuït	hello@logentries.com

Taula 4.11: Targeta d'Identitat de Logentries

[19]Els logs d'aquestes aplicacions no són fàcils de manejar, i enviant-los a un lloc centralitzat resulta més fàcil depurar aplicacions d'aquest tipus. Loggly implementa un mecanisme per poder enviar els logs mitjançant HTTP Post, que es pot afegir a les aplicacions.

Punts forts

- Barem del servei (i del preu) segons les necessitats.
- Posseeix poques funcionalitats però fa correctament la seva funció.
- L'agent per al transport de logs.

Punts febles

- Poca maduresa del producte.
- Excessivament simple i amb poques possibilitats de configuració.
- Seguretat dels logs enviats al núvol.

Taula 4.12: Punts forts i febles de Logentries

Logentries pot manejar una quantitat important de formats de log[20] (sistemes operatius, servidors web, servidors d'aplicacions o bases de dades). El llistat és bastant extens encara que sempre és possible crear nous tipus d'entrades log; però mantenint sempre la marca de temps i utilitzant expressions regulars.

Logentries emmagatzema les dades dels logs en el núvol Amazon EC2, més concretament a la regió Amazon EU (Irlanda).

La interfície de cerca és molt simple, semblada a la coneguda barra de Google. Realment la potència de les cerques no és molt gran, puix que només permet buscar per diferents servidors, arxius de logs i esdeveniments utilitzant paraules clau o expressions regulars. Tot el mecanisme de cerca i presentació està orientat a temps. Per a això cal especificar el rang temporal que interessa estudiar, i Logentries visualitzarà els gràfics i les estadístiques per al rang escollit.

La visualització és una característica bastant limitada, ja que bàsicament permet observar el volum d'esdeveniments per unitat de temps i identificar esdeveniments etiquetats com a errors, avisos o excepcions.

La impressió general que causa Logentries és que es tracta d'una aplicació en obres, però en continu canvi: cosa característica d'una *start-up*.

4.4 Classificació

La part final del procés d'avaluació és la selecció manual i la classificació. La Fulla d'Avaluació de cada producte és on es troba la puntuació de cada candidat. Els criteris poden puntuar-se amb tres possibles valoracions, que són nombres enters entre 0 i 2, sent 2 la millor puntuació, i 1 o menys indiquen que el candidat no compleix completament amb el criteri.

Les dades de l'avaluació completa poden veure's en l'Apèndix D. En la Taula 4.13 hi ha un resum de les puntuacions[21] obtingudes per cadascun dels productes, agrupades segons els principals grups de

[20] Per defecte, està limitat a entrades de log d'una sola línia. Per a entrades multilínia, com les excepcions, cal crear un nou tipus i modificar el separador de línia per defecte.

[21] Les puntuacions es calculen com la mitjana aritmètica simple de les puntuacions dels criteris que componen cada grup

requisits. En la Figura 4.2 es poden observar els mateixos resultats, però en forma de diagrama de radar.

	Splunk	Octopussy	Graylog2	Loggly	Logentries
Criteris genèrics	1.42	1.21	1.21	1.29	0.67
Criteris funcionals	1.32	0.74	0.58	0.63	0.73
Criteris tècnics	1.43	0.93	0.88	1.51	1.33

Taula 4.13: Resum de les puntuacions obtingudes pels diferents productes candidats.

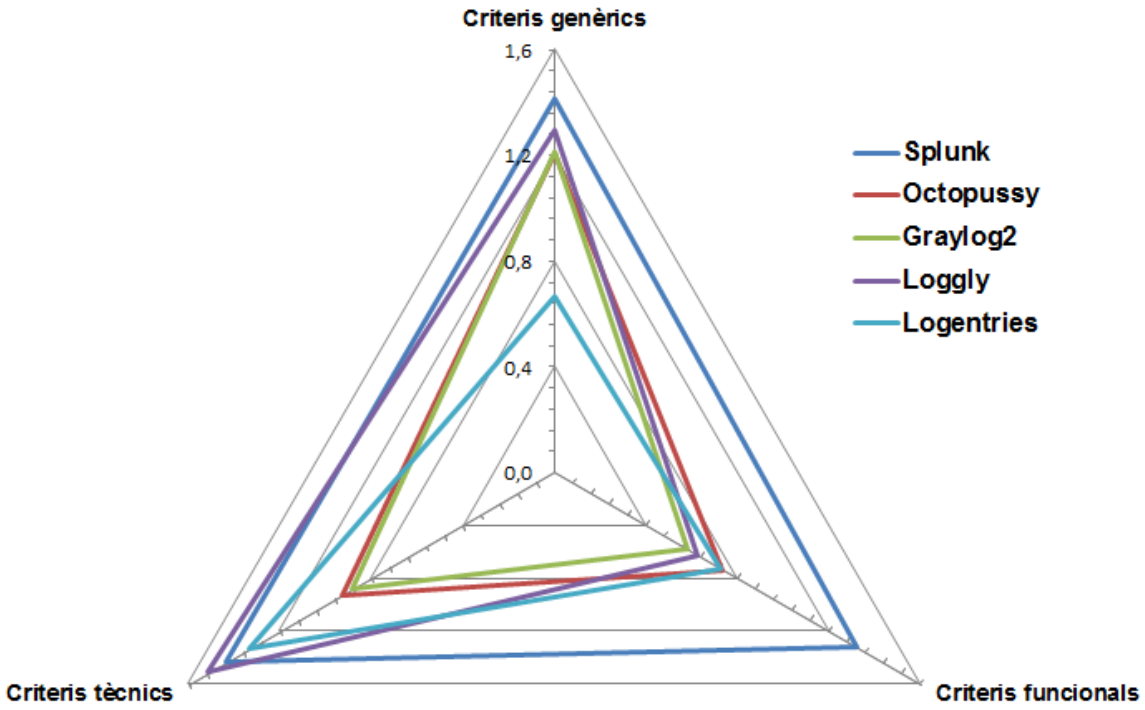

Figura 4.2: Diagrama de radar amb el resultat de l'avaluació dels productes candidats

Una ràpida inspecció de les puntuacions mostra que un dels productes candidats és clar vencedor (i a més en gairebé tots els grups de criteris): Splunk. Però caldria fer algunes consideracions addicionals que veurem a continuació.

4.4.1 Criteris generals

En els criteris generals, els candidats (amb l'excepció d'un) han puntuat de manera bastant semblant. Les puntuacions obtingudes pels productes candidats mostren que, en general, compleixen amb els criteris.

Splunk puntua una mica per sobre dels altres gràcies a les dades obtingudes en "Durabilitat" i en "Documentació". Els altres, amb l'excepció de Logentries, obtenen bones dades en "Adaptabilitat Tècnica" que –recordem– és una mesura de la mantenibilitat. Les males puntuacions de Logentries es deuen a la seva mancança de maduresa i a l'opacitat del funcionament intern de l'eina, que fa que no se sàpiga gens del seu funcionament intern.

de criteris.

	Splunk	Octopussy	Graylog2	Loggly	Logentries
1. Durabilitat	1.83	1.33	1.17	1.17	0.67
2. Documentació	2.00	1.00	1.00	2.00	1.00
3. Explotabilitat	1.50	1.50	1.00	1.00	1.00
4. Adaptabilitat tècnica	0.33	1.00	1.67	1.00	0.00

Taula 4.14: Resum de les puntuacions en els criteris generals.

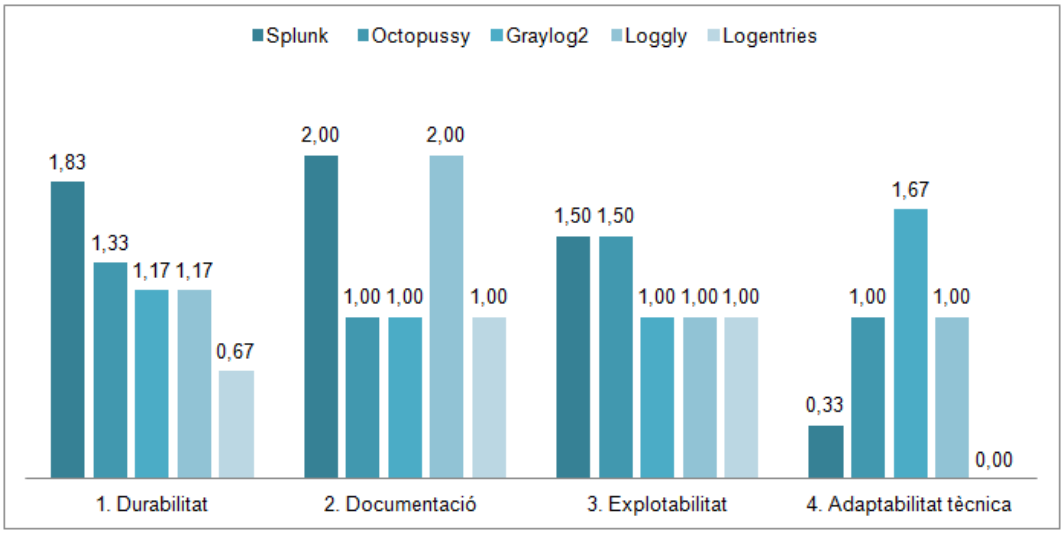

Figura 4.3: Valoració dels productes candidats per als criteris generals.

4.4.2 Criteris funcionals

Els criteris funcionals representen la funcionalitat desitjada. En aquest aspecte, Splunk es pot considerar el millor producte, ja que és el que millor puntuació obté quant a criteris funcionals; a més, aconsegueix la major puntuació en cadascun dels grups (veure Taula 4.15). A continuació, Octopussy i Logentries aconsegueixen les següents puntuacions, gràcies als resultats en "Entrada de dades" i "Cerca i anàlisi". Loggly no obté un bon resultat a causa de les males dades en la gestió d'alertes, i Graylog2 obté la puntuació més baixa en "Cerca i anàlisi".

És important ressaltar la puntuació total obtinguda per Splunk, gràcies al potent llenguatge de consulta inclòs en el producte.

	Splunk	Octopussy	Graylog2	Loggly	Logentries
5. Entrada de dades	1.33	0.92	0.83	0.75	1.00
6. Emmagatzematge	0.84	0.37	0.37	0.45	0.37
7. Cerca i anàlisi	1.80	0.92	0.54	0.68	0.81

Taula 4.15: Resum de les puntuacions en els criteris funcionals.

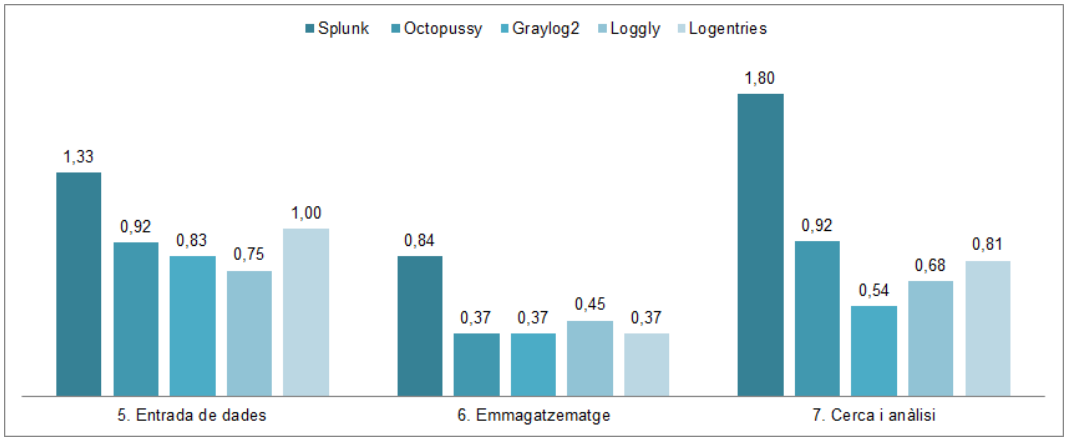

Figura 4.4: Valoració dels productes candidats per als criteris funcionals.

4.4.3 Criteris tècnics

Tal com es pot veure en la Taula 4.16, els criteris tècnics fan referència a la facilitat d'instal·lació, la interoperabilitat, la seguretat o la interfície gràfica d'usuari (GUI). Evidentment, aquells productes que no requereixin el clàssic procés d'instal·lació obtindran un bon resultat, perquè són més fàcils de gestionar, actualitzar i no presenten dependències d'altres llibreries. Són els casos de Loggly i Logentries. En el cas dels altres, han de ser instal·lats en els servidors de cada organització. Encara que aquest procés sol ser senzill i automatitzat, en el cas de Octopussy resulta bastant complex per la quantitat de dependències amb llibreries en Perl que cal instal·lar, així com el repositori de dades MongoDB.

La configuració de cada producte segueix sent una tasca laboriosa, que (encara que s'intenti simplificar) segueix necessitant de coneixement tècnic. Així i tot, la utilització d'assistents i altres facilitats simplifica algunes fases, però segueix sent necessari arribar fins als fitxers de configuració de les eines. Per tant, en les eines que és imprescindible instal·lar (Splunk, Octopussy i Graylog2) la configuració és més complexa que en aquelles (com Loggly i Logentries) que no s'instal·len, encara que el control que aconseguim amb les primeres no és comparable amb el de les segones.

Pel que fa a la "Seguretat", és Splunk l'eina que permet més i millors mecanismes per implementar autenticació i autorització, mentre que les altres només tenen mecanismes bàsics per permetre aquesta seguretat.

Sobre la interfície d'usuari, tant Splunk (pel completa) com Octopussy (pel personalitzable) obtenen les més altes puntuacions. Les altres, en aquest aspecte, són bastant més bàsiques.

	Splunk	**Octopussy**	**Graylog2**	**Loggly**	**Logentries**
8. Facilitat d'instal·lació	2.00	1.00	2.00	2.00	2.00
9. Facilitat de configuració	1.00	1.00	1.00	2.00	2.00
10. Facilitat d'actualització	2.00	2.00	1.00	2.00	2.00
11. Dependència de llibreries externes	1.00	0.00	1.00	2.00	2.00
12. Interoperabilitat	1.00	0.50	0.00	1.50	0.50
13. Seguretat	1.53	0.50	0.69	0.34	0.34
14. Interfície d'usuari	1.50	1.50	0.50	0.75	0.50

Taula 4.16: Resum de les puntuacions en els criteris tècnics.

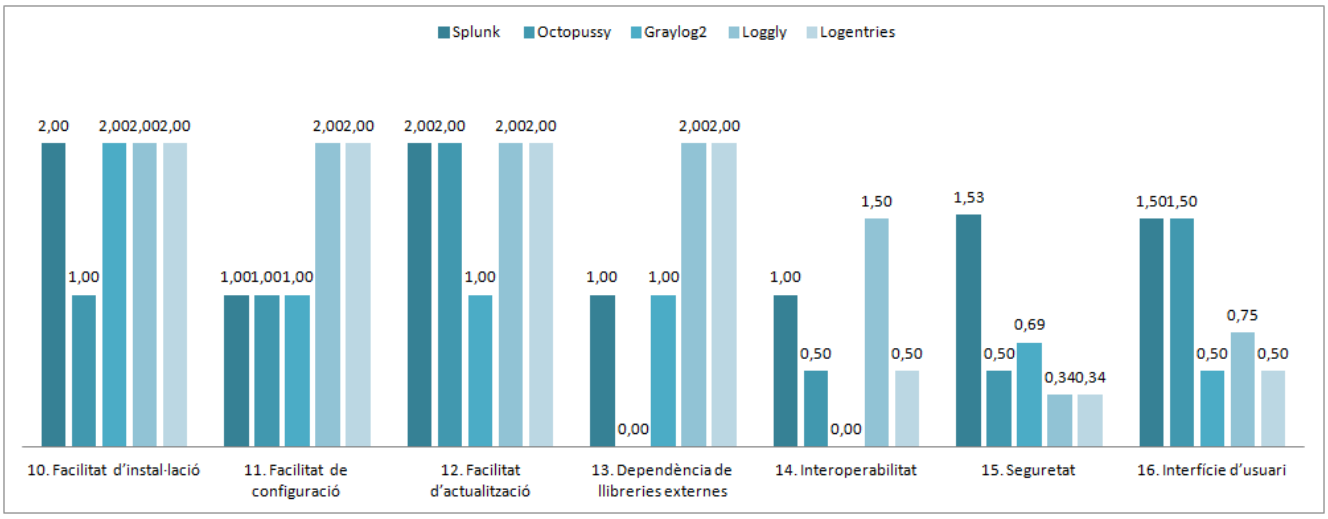

Figura 4.5: Valoració dels productes candidats per als criteris tècnics.

4.4.4 Comentaris finals

Encara que s'ha descrit un mètode d'avaluació que cobreix aspectes molt rellevants de la gestió de logs, el mètode en si mateix té algunes limitacions.

De tota manera, com es va comentar quan es va descriure QSOS (Capítol 3), **el mètode és iteratiu**, la qual cosa vol dir que en successius cicles es pot refinar, tant en relació amb els criteris d'avaluació[22] com en relació amb diversos aspectes del propi mètode.

Una qüestió rellevant, relativa als criteris de valoració, és la importància que (dins dels criteris funcionals) s'ha atribuït a la branca de "Cerca i anàlisi".

Com s'ha comentat, diversos criteris utilitzats per a l'avaluació depenen d'un judici subjectiu de l'avaluador. Una forma d'assegurar que s'assigna la puntuació correcta és encomanar el mateix procés d'avaluació a diverses persones; llavors es podrien analitzar la diferents puntuacions donades a un mateix criteri i realitzar algun tipus de ponderació.

El nombre d'alternatives (productes candidats) avaluades podria haver estat més elevat. Aquest aspecte, no obstant això, no és una qüestió crucial per al cas considerat, ja que el nombre de candidats avaluats havia de ser abastable i pràctic. Partint d'una llista inicial de 13 productes, s'ha avaluat finalment a 5; és a dir, el filtratge inicial ha eliminat el 61% dels possibles candidats.

A més, una qüestió que m'agradaria posar de manifest és que realment es tracta d'una comparació desequilibrada, perquè un dels productes és molt conegut[23] mentre que els altres són pràcticament desconeguts (o minoritaris, si es prefereix). És com si, salvant les distàncies, s'intentés fer una comparació entre Google[24] i altres cercadors en la web.

Amb tot això, les meves conclusions sobre l'avaluació serien:

- En la meva opinió, i en de vista de l'avaluació, ara per ara Splunk és la referència a seguir. És una eina molt completa que cobreix pràcticament la majoria de les necessitats de qualsevol organitza-

[22] Es podrien afegir, modificar o eliminar criteris que no resultin suficientment descriptius o importants. Per exemple, en l'avaluació cap dels productes va utilitzar SOAP com a protocol de transport; per tant podria eliminar-se en següents versions o adquirir més significat si CEE es converteix en un estàndard de log.

[23] Aquesta circumstància fa que, a més de l'extensa documentació disponible, existeixin publicacions i articles que permeten conèixer més en profunditat les característiques i el funcionament intern.

[24] De fet, a Splunk se li denomina el Google de la gestió de logs.

ció, encara que sigui complicada d'entendre i de dominar; no obstant això aquesta circumstància no hauria de ser un problema, ja que el públic destinatari és personal TIC.

- En un segon grup es troben les eines de programari lliure, Octopussy i GrayLog2, que tenen una funcionalitat que cobreix els principals punts de la gestió de logs i que tenen l'avantatge que estan disponibles perquè es puguin modificar o adaptar (això sí, a costa de temps de desenvolupament).
- Finalment, figuren Loggly i Logentries, encara limitades però amb un potencial d'actualització enorme. En un futur proper, alguna d'elles pot esdevenir una referència en l'àmbit dels sistemes de gestió de logs en el núvol.

Capítol 5
Conclusions

"No es fa més amb menys. Això és mentida. Es fa menys amb menys."

David Simon, creador de The Wire.

Aquest Projecte de Fi de Carrera ha suposat un esforç considerable per la varietat de les qüestions de naturalesa diversa tractades en ell.

5.1 Objectius assolits

En finalitzar aquest Projecte, i analitzar els resultats, podem arribar a la conclusió que la majoria dels objectius que s'havien definit al inici del Projecte han estat assolits.

D'una banda, s'ha analitzat la problemàtica que suposen els logs en els entorns de producció actuals. Aquest estudi ha inclòs la descripció dels formats de logs més habituals i els intents d'estandardització d'aquests formats.

D'altra banda, s'han estudiat diferents metodologies per a l'avaluació de programari, i s'ha adaptat una com a marc de treball per poder avaluar una sèrie de productes que suportin la gestió de logs. Com a part d'aquest marc, s'ha seleccionat un conjunt de criteris d'avaluació (que cobreixin la major part dels requisits propis de la gestió de log) amb els quals poder realitzar una avaluació el més objectiva possible d'una sèrie de productes.

Finalment, s'ha realitzat una avaluació seguint el mètode proposat, i s'han analitzat els resultats per a uns productes candidats seleccionats.

M'hagués agradat disposar de més temps per dur a terme totes les proves que m'havia plantejat inicialment.

5.2 Treball futur

Durant la realització del Projecte han sorgit idees que podrien ser desenvolupades en el futur.

Una de les primeres qüestions en aparèixer va ser la necessitat d'usar una metodologia per acomplir una avaluació. Aquesta metodologia adaptada a les necessitats de la gestió de logs presenta algunes deficiències que podrien esmenar-se. Les principals qüestions relatives a la metodologia que haurien de tractar-se en un futur serien:

- Ajust i refinatge dels criteris d'avaluació.
- Realitzar la mateixa avaluació per diferents persones (amb diferents perfils), i ponderar els resultats.

- Avaluar altres productes no inclosos en la llista de candidats.
- Desenvolupar una aplicació[1] que permeti portar control del procés d'avaluació, versions dels criteris, valoracions i perfils d'avaluadors.

D'altra banda, la recollida de dades en el procés d'avaluació s'ha dut a terme revisant la documentació tècnica dels productes, el codi (quan estava disponible), i realitzant una instal·lació de prova d'aquests productes, per comprovar funcionalitat. Aquesta instal·lació s'ha fet seguint la guia que apareix a la pàgina web correspondent, i comprovant que funcionés correctament la instal·lació. S'hauria de fer una prova més formal, descrivint i documentant amb detall el procés de prova, i utilitzant un estàndard. Aquesta prova d'instal·lació i els seus resultats haurien d'incorporar-se com a part del mètode.

5.3 Conclusions

El balanç del Projecte, contemplat en la seva totalitat, ha estat molt positiu.

D'una banda he tingut l'oportunitat d'entrar en contacte amb un camp desconegut per a mi fins a aquest moment, com era la gestió de log i les eines SIEM, àmbit en el qual he aprofundit tant a nivell teòric com a nivell pràctic. Vist en perspectiva, la principal dificultat que ha suposat aquest projecte ha estat sens dubte la gran quantitat de tecnologies que han format part del project.

D'altra banda, també m'he apropat a la problemàtica de l'avaluació de productes programari i he pogut adaptar i utilitzar una metodologia com a marc de referència. S'ha hagut de definir completament els criteris amb els quals s'ha efectuat l'avaluació, la qual cosa ha exigit un coneixement de l'estat de la qüestió per poder cobrir tots els aspectes rellevants.

L'etapa d'avaluació mereix una menció a part, car he estat en contacte amb els productes més importants per a la gestió de logs i he tingut l'oportunitat d'estudiar detalls del seu funcionament (intern i extern) i de la forma en la qual estan construïts. Considero que això és molt important, ja que m'ha permès conèixer la forma en la qual es desenvolupen actualment les aplicacions i les bases de dades NoSQL.

A més, a part de tots els coneixements pràctics sobre la gestió de projectes, aquest PFC m'ha aportat una mica més: l'interès pel món de la gestió en tecnologia i de les *start-ups*, empreses que comencen a partir d'una idea o d'una innovació tecnològica.

També caldria destacar l'etapa de documentació, i en especial de la redacció de la Memòria del Projecte.

[1] Ja hi ha algunes eines (com add-on de Firefox) proporcionades per QSOS que permeten realitzar una gestió molt bàsica del algunes fases del projecte. La idea és construir un programari (com *mashup*) que utilitzi components (com llibreries de visualització o magatzems de dades) ja existents.

Glossari

ØMQ ZeroMQ. Biblioteca de missatges asíncrons d'alt rendiment dissenyada per a aplicacions concurrents i escalables. Està dissenyada com un API similar als coneguts sockets.

Add-on Millores instal·lables per les aplicacions (similars als plug-in).

AMQP Advanced Message Queuing Protocol. AMPQ és un estàndard obert de missatgeria que permet a diferents plataformes en diferents llenguatges enviar missatges els uns als altres.

API Application Programming Interface. Interfície utilitzada per components programari per comunicar-se entre si. Una API pot incloure especificacions de funcions, estructures de dades, classes d'objectes i variables..

BEEP Blocks eXtensible eXchange Protocol. Marco de treball per crear protocols d'aplicacions de xarxa.

CLI Command–Line Interface. Mecanisme d'interactuar amb un sistema operatiu o programari consistent a escriure comandos per realitzar tasques concretes.

Cookies Les cookies són peces d'informació que el servidor HTTP envia al client juntament amb els recursos sol·licitats. El navegador pot emmagatzemar aquesta informació i, posteriorment, enviar-la de tornada al servidor HTTP quan faci altres sol·licituds. El servidor HTTP pot establir diverses cookies per cada sol·licitud HTTP. Les cookies són de la forma de parells clau/valor; quan una clau té múltiples valors, es delimiten amb punt i coma (;).

COTS Commercial-Off-The-Shelf. El programari actual es construeix integrant components de programari de diferent naturalesa i orígens, normalment desenvolupats per tercers. Aquests components es denominen OTS, de les sigles angleses "Off–The–Shelf", al·ludint a la seva disponibilitat. Els components OTS comercials, denominats COTS, es poden personalitzar i integrar amb programari desenvolupat a mida.

Creative Commons Creative Commons (CC) és una organització sense ànim de lucre que ofereix llicències *copyright* flexibles per a treballs creatius.

cron Administrador de processos en segon pla (dimoni) que executa processos o scripts a intervals regulars en sistemes Unix/Linux.

daemon Procés en segon pla que dóna servei a usuaris locals o remots, o altres processos.

FOSS/FlOSS Free/libre and Open Source Software. És el programari la llicència del qual permet als usuaris estudiar, modificar i millorar el seu disseny mitjançant la disponibilitat del seu codi font.

GUI Graphic User Interface. Interfície d'usuari que utilitza elements gràfics que permeten interactuar de forma molt més intuïtiva amb un sistema informàtic.

HTTP Hyper Text Transfer Protocol. Protocol d'aplicació per a sistemes hipertextuals.

HTTPS Hyper Text Transfer Protocol Secure. Protocol d'aplicació basat en el protocol HTTP destinat a la transferència segura de dades hipertextuals.

IDS Intrusion Detection System. Sistema dissenyat per detectar intents no desitjats d'accés, manipulació o desconnexió d'ordinadors a través d'una xarxa.

IPS Intrusion Prevention System. Sistema que monitoritza xarxes o activitat d'ordinadors a la recerca d'activitat maliciosa o comportaments no desitjats i que pot bloquejar o prevenir aquestes activitats.

JSON JavaScript Object Notation. JSON és un format lleuger en text per a l'intercanvi de dades i independent del llenguatge (encara que basat en Javascript) que s'usa com a alternativa a XML per descriure objectes.

LaaS Logging–as–a–Service (veure SaaS).

MapReduce Paradigma i marc de referència originari de Google per a computació distribuïda en *clusters* de computadors. El nom s'inspira en els noms de dos importants mètodes: Map i Reduce. L'operació Map presa un parell clau/valor d'entrada i produeix un grup intermedi de parells claus/-valors. MapReduce agrupa tots els valors intermedis associats amb la mateixa clau intermèdia i els passa a la funció Reduce. La funció Reduce accepta una clau intermèdia i un grup de valors per a aquesta clau i fusiona aquests valors per formar un grup possiblement més petit de valors. Això permet manejar llistes de valors massa grans per a la memòria física del sistema.

mashup Aplicació que usa i combina dades, presentacions i funcionalitat procedents d'una o més fonts per crear nous serveis usant normalment APIs oberts i fonts de dades per produir resultats enriquits.

Nagios Nagios és un sistema de monitoratge de xarxes de codi obert.

NoSQL Terme usat per agrupar una sèrie de magatzems de dades no relacionals que no permeten transaccions; normalment no tenen esquemes de taules ni *joins*.

NTP Network Time Protocol. Protocol que permet sincronitzar rellotges en una xarxa de computadors.

plug-in Component programari que afegeix capacitats específiques a una aplicació.

RegExp Regular Expression. Una expressió regular és una representació, segons unes regles sintàctiques d'un llenguatge formal, d'una porció de text genèric a buscar dins d'un altre text.

REST REpresentational State Transfer. Arquitectura de programari per a sistemes distribuïts consistent en l'existència de recursos (elements d'informació) que poden ser accedits utilitzant un URI i en la qual els components de la xarxa (clients i servidors) es comuniquen a través d'una interfície estàndard (HTTP) i intercanvien representacions d'aquests recursos.

RFC Request For Comment. Document publicat per la IETF (Internet Engineering Task Force) que descriu les novetats, innovacions, comportaments, investigació o mètodes relatius a Internet i als sistemes connectats a Internet.

SaaS Software–as–a–Service. Anomenat com a "programari sota demanda", és un model en el qual el programari i les seves dades associades es guarden de manera centralitzada (típicament, en el núvol) i s'accedeix mitjançant un client lleuger (un navegador).

SIM/SEM/SIEM Security Information and Event Monitoring (veure 2.4.2).

SOAP Simple Object Access Protocol. Protocol per a l'intercanvi d'informació estructurada (en format XML) en serveis web sobre xarxes d'ordinadors.

Solr Solr és una plataforma de cerca de codi obert que ve del projecte Apache Lucene.

SSL/TSL Secure Sockets Layer i el seu successor Transport Layer Security són protocols criptogràfics que proporcionen comunicacions segures per una xarxa.

SSO Single sign-on. Procediment d'autenticació que habilita a l'usuari per accedir a diversos sistemes amb una sola identificació.

start-up Negoci, generalment del món de la innovació i la tecnologia, amb una història de funcionament limitada però amb possibilitats de creixement.

Timestamp Cadena de caràcters que reflecteixen l'hora i data (o alguna d'elles) en la qual va ocórrer determinat esdeveniment.

URI Uniform Resource Identifier. Cadena de caràcters curta que identifica inequívocament un recurs (servei, pàgina, document, etc.).

URL Uniform Resource Locator. Cadena de caràcters que segons un format modèlic i estàndard s'usa per nomenar recursos en Internet per a la seva localització o identificació.

UTC Universal Time Coordinate. Temps de la zona horària de referència respecte a la qual es calculen totes les altres zones del món. Es basa en rellotges atòmics que afegeixen segons a intervals irregulars per compensar la rotació terrestre cada vegada més lenta.

VPN Virtual Private Network. Xarxa que se situa sobre una xarxa d'ordinadors per aconseguir confidencialitat.

Zabbix Zabbix en un programari lliure per monitoritzar i registrar serveis de xarxa, servidors i maquinari.

Bibliografia

[1] ABAD, CRISTINA; TAYLOR, JED; ZHOU, YUANYUAN; SENGUL, CIGDEM; ROEW, KEN i YURCIK, WILLIAM: «Log Correlation for Intrusion Detection: A Proof of Concept». En: Proceedings of the 19th Annual Computer Security Applications Conference, Las Vegas, ACSAC, 2003.

[2] AGOSTI, MARISTELLA i DI NUNZIO, GIORGIO MARIA: «Gathering and mining information from web log files». En: Proceedings of the 1st international conference on Digital libraries: research and development, DELOS'07, pp. 104–113. Springer-Verlag, Berlin, Heidelberg, 2007. http://dl.acm.org/citation.cfm?id=1782334.1782347

[3] ATOS-ORIGIN: «Method for Qualification and Selection of Open Source software (QSOS) version 1.6», 2006.

[4] BITINCKA, LEDION; GANAPATHI, ARCHANA; SORKIN, STEPHEN i ZHANG, STEVE: «Optimizing Data Analysis with a Semi-structured Time Series Database», 2010, p. 7–7.

[5] BURGESS, MARK: «On the Theory of System Administration», 2004.

[6] CALDWELL, MATTHEW: «The Importance of Event Correlation for Effective Security Management», 2002.

[7] CERT-IN: «Guidelines for Auditing and Logging. Version 2», 2008.

[8] CHEE, BRIAN:. «Splunk Simplifies Log-File Searches». InfoWorld; May 1, 2006, p. 44.

[9] CHUVAKIN, ANTON: «Network, Database, and System Log Data Management: The What, Why, and How», 2008.

[10] DAVID, BROUGH; HORWATH, JIM i ZABIUK, JOHN: «What's in the data bucket? Event Correlation and SIEM Vendor Approaches», 2010.

[11] DEPREZ, JEAN-CHRISTOPHE i ALEXANDRE, SIMON: «Comparing Assessment Methodologies for Free/Open Source Software: OpenBRR and QSOS», 2008.

[12] JIANG, WEIHANG; HU, CHONGFENG; PASUPATHY, SHANKAR; KANEVSKY, ARKADY; LI, ZHENMIN i ZHOU, YUANYUAN: «Understanding Customer Problem Troubleshooting from Storage System Logs», 2009.

[13] KARLZÉN, HENRIK: «An Analysis of Security Information and Event Management Systems. The Use of SIEMs for Log Collection, Management and Analysis.», 2009.

[14] KAVANAGH, KELLY M.: «Findings: Evaluate Limitations of Log Management as a Service Offering», 2011.

[15] KENT, KAREN i SOUPPAYA, MURUGIAH: Guide to Computer Security Log Management. NIST (National Institute of Standards and Technology) Special Publication 800-92, 2006.

[16] KOWALL, JOHAH: «Need is Growing for Operations Log File Management», 2011.

[17] KRIZAK, PAUL i DEVICES, ADVANCED MICRO: «Log Analysis and Event Correlation Using Variable Temporal Event Correlator (VTEC)», 2010.

[18] LACROIX, THOMAS; LOUX, VALENTIN; GENDRAULT, ANNIE; GIBRAT, JEAN-FRANÇOIS i CHIAPELLO, HÉLÈNE: «CompaGB: An open framework for genome browsers comparison». BMC Res Notes, 2011, **4**, p. 133.

[19] LI, ZHENMIN; TAYLOR, JED; PARTRIDGE, ELIZABETH; ZHOU, YUANYUAN; YURCIK, WILLIAM; ABAD, CRISTINA; BARLOW, JAMES J. i ROSENDALE, JEFF: «UCLog: A unified, correlated logging architecture for intrusion detection». En: 12th International Conference on Telecommunication Systems - Modeling and Analysis, ICTSM, 2004.

[20] MAIER, PHILLIP Q.: Audit and Trace Log Management. Consolidation and Analysis. Auerbach Publications, 2006.

[21] MARTY, RAFFAEL: Applied Security Visualization. Pearson Education, Inc., 2009.

[22] ——: «Cloud application logging for forensics». En: Proceedings of the 2011 ACM Symposium on Applied Computing, SAC '11, pp. 178–184. ACM, New York, NY, USA, 2011. doi: http://doi.acm.org/10.1145/1982185.1982226.
http://doi.acm.org/10.1145/1982185.1982226

[23] MEARIAN, LUCAS: «IT Looks for New Tools To Exploit 'Big Data'». Computerworld; Apr 4, 2011, 2011, p. 6.

[24] MITRE: Common Event Expression: CEE, A Standard Log Language for Event Interoperability in Electronic Systems., 2008.

[25] ——: Common Event Expression: Architecture Overview Version 0.5, 2010.

[26] NAGAPPAN, MEIYAPPAN: «Analysis of execution log files». En: Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 2, ICSE '10, pp. 409–412. ACM, New York, NY, USA, 2010. doi: http://doi.acm.org/10.1145/1810295.1810405.
http://doi.acm.org/10.1145/1810295.1810405

[27] NICOLETT, MARK i KOWALL, JOHAH: «Log Management for IT Security and IT Operations», 2011.

[28] PETRINJA, ETIEL; SILLITTI, ALBERTO i SUCCI, GIANCARLO: «Comparing OpenBRR, QSOS, and OMM Assessment Models.» En: Pär J. Ågerfalk; Cornelia Boldyreff; Jesús M. González-Barahona; Gregory R. Madey i John Noll (Eds.), OSS, volum 319 de IFIP, pp. 224–238. Springer. ISBN 978-3-642-13243-8, 2010.

[29] SECUROSIS: «Understanding and Selecting SIEM/Log Management. Version 2», 2010.

[30] SVEEN, ATLE FRENVIK: Use of Free and Open Source GIS in Commercial Firms. Tesi doctoral, NTNU - Norwegian University of Science and Technology, 2008.

[31] TAN, JIAQI; PAN, XINGHAO; KAVULYA, SOILA; G, RAJEEV i NARASIMHAN, PRIYA: «SALSA: Analyzing Logs as StAte Machines», 2008.

[32] THURMAN, MATHIAS:. «Buried in SIEM Configuration». Computerworld; Dec 20, 2010, p. 29.

[33] TIFFANY, MICHAEL: «A Survey of Event Correlation Techniques and Related Topics», 2002.

Apèndix A
Formats de log

A.1 Formats de log de sistemes operatius

A.1.1 Unix/Linux

En els sistemes operatius de la família Unix, l'estàndard de log de facto és syslog. El terme syslog és utilitzat per descriure tant el protocol (veure Apèndix B.1) per a l'enviament de missatges com el programa que envia missatges.

El format dels missatges de registre és molt similar, independentment del dimoni que s'usi per registrar-los (en el cas de rsyslog poden aparèixer lleugeres discrepàncies amb altres eines). En el cas de rsyslog, el format per defecte dels missatges és:

```
DataHora hostname programa[PID] missatge.
```

A continuació vegem alguns exemples:

```
2011-05-10T16:03:09.631711+02:00 agd-desktop kernel: [394083.122379] chrome[10865]: segfault at 0
ip 00007fc11ecbcc84 sp 00007fffe5520d70 error 4 in libflashplayer.so[7fc11e93f000+9be000]

2011-05-10T21:53:22.925831+02:00 agd-desktop gnome-session[30498]: WARNING: GSIdleMonitor: Unable
to initialize Sync extension

2011-05-10T21:53:23.040807+02:00 agd-desktop gnome-keyring-daemon[30489]: GLib-GIO: Using the
'memory' GSettings backend.  Your settings will not be saved or shared with other applications.

2011-05-10T23:09:25.341043+02:00 agd-desktop su[31506]: pam_unix(su:auth): authentication failure;
logname=agd uid=1000 euid=0 tty=/dev/pts/3 ruser=agd rhost=  user=root

2011-05-10T23:09:28.045834+02:00 agd-desktop su[31506]: pam_authenticate: Authentication failure

2011-05-10T23:09:28.046433+02:00 agd-desktop su[31506]: FAILED su for root by agd

2011-05-10T23:09:28.046521+02:00 agd-desktop su[31506]: - /dev/pts/3 agd:root
```

A.1.2 Microsoft Windows

Els sistemes operatius de Microsoft escriuen els esdeveniments en diferents fitxers log que es divideixen en aplicació, sistema i seguretat. No resulta fàcil obtenir aquests logs del sistema sense utilitzar programari de tercers. A més és gairebé impossible veure aquests logs si no és en el propi sistema i usant les eines EventViewer o Microsoft Management Console (segons versions) que vénen de sèrie amb Windows.

Microsoft utilitza un mecanisme d'identificadors d'esdeveniment (Event ID) que defineixen unívocament els esdeveniments que un sistema amb Windows pot identificar.

Amb Windows Vista es va redissenyar l'arquitectura d'esdeveniments i traces, usant un format de log amb XML. Amb l'eina Wevtutil es poden veure tots els possibles esdeveniments i la seva estructura. Existeix un gran nombre de tipus d'esdeveniments incloent administratius, operacionals, analítics i depuració.

Propietat	**Descripció**
Source	Programari que va registrar l'esdeveniment, que pot ser tant el nom d'una aplicació com un component de sistema o un gestor de dispositius (driver).
Event ID	Codi numèric identificant un tipus d'esdeveniment.
Level	Classificació de la gravetat de l'esdeveniment. Els següents nivells es generen per logs d'aplicació i sistema: • Information. Indica qualsevol canvi en una aplicació o component, com una operació realitzada amb èxit o la creació d'un recurs. • Warning. Indica que s'ha produït un problema que pot afectar al servei o donar lloc a un problema més greu si no es prenen mesures. • Error. Indica que s'ha produït un problema que podria afectar a la funcionalitat. • Critical. Indica que s'ha produït una fallada i que l'aplicació o el component que va desencadenar l'esdeveniment no es pot recuperar.
User	Nom de l'usuari baix el compte del qual s'ha produït l'esdeveniment.
Operational Code	Codi numèric que identifica l'activitat que l'aplicació estava realitzant quan es va disparar l'esdeveniment.
Log	Nom del log on l'esdeveniment va ser registrat.
Task Category	Sub-component o activitat que publica l'esdeveniment.
Keywords	Conjunt d'etiquetes que es poden utilitzar per filtrar o buscar esdeveniments.
Computer	Nom de l'ordinador en el qual ha succeït l'esdeveniment.
Date and Time	Data i hora en la qual es va registrar l'esdeveniment.

Taula A.1: Propietats comunes en sistemes Windows

Hi ha esdeveniments Windows que es refereixen al procés d'autenticació com a inici (*login*) o fi de sessió (*logout*). En la Taula A.2 s'enumeren els diferents identificadors d'esdeveniments. Per descomptat, tots els identificadors d'esdeveniments canvien en Windows 7/Vista i Windows Server 2008 (veure Taula A.3).

Existeix un problema a l'hora de recol·lectar els esdeveniments de Windows en un format estàndard com syslog. Una forma bastant habitual és utilitzar eines de tercers com *Snare*. *Snare* converteix els esdeveniments de Windows a un format compatible amb syslog, permetent a un administrador especificar quins Event IDs seran enviats a un servidor syslog.

Event ID	Descripció
528	Èxit en iniciar de sessió.
529	Fallada d'inici de sessió: usuari desconegut o contrasenya errònia.
530	Fallada d'inici de sessió: violació de la restricció de temps.
531	Fallada d'inici de sessió: compte deshabilitat.
532	Fallada d'inici de sessió: compte vençut.
533	Fallada d'inici de sessió: l'usuari no pot iniciar sessió en aquest ordinador.
534	Fallada d'inici de sessió: l'usuari no té permisos per iniciar sessió en aquest tipus de màquines.
535	Fallada d'inici de sessió: contrasenya caducada.
536	Fallada d'inici de sessió: component NetLogon no activat.
537	Fallada d'inici de sessió: altres raons.
538	Fi de sessió
539	Fallada d'inici de sessió:compte bloquejat.
540	Èxit en inici de sessió en xarxa.
552	Intent d'inici de sessió usant credencials.

Taula A.2: Esdeveniments d'inici i fi de sessió en Windows

Event ID	Descripció
4624	Èxit quan el compte va iniciar sessió.
4625	Fallada en l'inici de sessió.
4648	Intent d'inici de sessió usant credencials.

Taula A.3: Esdeveniments d'inici i fi de sessió en Windows 7/Vista i 2008

A.2 Formats de log de servidors web

Els únics formats de logs mitjanament estandarditzats es troba en el cas dels servidors web. Aquí veurem els dos més utilitzats.

A.2.1 NCSA

El format de log NCSA[1] està basat en el dimoni (*daemon*) httpd i està acceptat com a estàndard per a molts servidors HTTP.

A.2.1.1 NCSA comú (log d'accés)

El format de log comú de NCSA, normalment denominat log d'accés, conté només dades bàsiques d'accés HTTP. Aquest log conté dades dels recursos sol·licitats i unes poques dades més, però no conté dades del sol·licitant, agent o de *cookies*.

Els camps de format són:

```
host rfc931 username date:time request statuscode bytes
```

- host. L'adreça IP o el nom de host o subdomini del client que va realitzar la petició HTTP.

[1] National Center for Supercomputing Applications.

- rfc931. L'identificador del client; si no hi ha valor, se substitueix per –.
- username. Nom d'usuari utilitzat pel client per a l'autenticació; si no hi ha valor, se substitueix per –.
- date:time. La marca temporal en format [dd/MMM/yyyy:hh:mm:ss +-hhmm] de la petició.
- request. La petició HTTP que conté tres dades importants: la principal és el recurs sol·licitat, també el mètode (GET o POST) i la versió del protocol.
- statuscode. L'estat és un codi numèric que indica l'èxit o fracàs de la petició HTTP (veure http://en.wikipedia.org/wiki/List_of_HTTP_status_codes).
- bytes. Conté el nombre de bytes transferits en la petició HTTP, sense incloure la capçalera.

El següent és un exemple de log amb els camps farcits amb valors:

```
125.125.125.125 - dsmith [10/Oct/1999:21:15:05 +0500] "GET /index.html HTTP/1.0" 200 1043
```

A.2.1.2 NCSA Combinat

El format NCSA combinat és una extensió del format NCSA comú que conté les mateixes dades més tres camps addicionals optatius:

```
host rfc931 username date:time request statuscode bytes referrer user_agent cookie
```

- referrer. La URL que va usar l'usuari per vincular el lloc.
- user_agent. El navegador i la plataforma utilitzada per l'usuari.
- cookie. Es registra cada cookie per al recurs sol·licitat.

A.2.1.3 NCSA Separat

El format de log NCSA separat, denominat format 3-logs, és un format de log en el qual se separa la informació recopilada en tres fitxers separats, en lloc d'un sol fitxer. Els tres fitxers es refereixen sovint com: log d'accés, log de referència i log d'agent.

Els tres logs contenen les mateixes dades que el format NCSA combinat, excepte les dades de les cookies (que no que es registren en aquest format).

A.2.2 W3C Extended

Aquest format de log s'utilitza per Microsoft Internet Information Server (IIS)[2]. El log conté una seqüència de línies amb caràcters ASCII i cada línia és una directiva o una entrada.

Les directives (veure Taula A.4) són registres relatius al propi procés de log i es corresponen amb les línies que comencen amb el caràcter #.

Les entrades consisteixen en una seqüència de camps relatius a una sola transacció. Els camps se separen per espais en blanc; i si un camp no té dades, es posa un –. A diferència d'altres formats, les cookies en el format W3C no van entre cometes; i si hi ha diverses cookies, es delimiten amb el caràcter +.

[2]En les primeres versions de IIS, s'utilitzava una variant de W3C desenvolupada per Microsoft i que era fixa, sense possibilitat de personalització.

Directiva	Descripció
Version: <enter>.<enter>	Versió del format usat.
Fields: [<especificador>...]	Identifica els camps (veure Taula A.5) amb els quals es genera el log.
Software: <cadena>	Identifica el programari que va generar el log.
Start-Date: <data> <hora>	Data i hora en la qual comença el log.
End-Date: <data> <hora>	Data i hora en la qual finalitza el log.
Date: <data> <hora>	Data i hora en la qual es va afegir l'entrada.
Remark: <text>	Comentaris.

Taula A.4: Descripció de directives W3C Extended

Camp	Apareix com a	Descripció
Date	date	Data de l'activitat.
Time	time	Hora en UTC de l'activitat.
Client IP Address	c-ip	Adreça IP del client que va fer la petició.
User Name	cs-username	Nom de l'usuari autenticat que va accedir al servidor. Els usuaris anònims apareixen com –.
Service Name	s-sitename	Servei i nombre d'instància que està executant el client.
Server Name	s-computername	Nom del servidor en el qual es va generar el log.
Server IP Address	s-ip	Adreça IP del servidor en el qual es va generar el log.
Server Port	s-port	Port en el qual el servidor té configurat el servei.
Method	cs-method	Acció de sol·licitud (per exemple, el mètode GET).
URI Stem	cs-uri-stem	Objectiu de l'acció (per exemple, Default.htm).
URI Query	cs-uri-query	La consulta, si hi ha, que el client intenta realitzar.
HTTP Status	sc-status	Codi d'estat HTTP.
Win32 Status	sc-win32-status	Codi d'estat de Windows.
Bytes Sent	sc-bytes	Nombre de bytes que envia el servidor.
Bytes Received	cs-bytes	Nombre de bytes que va rebre el servidor.
Time Taken	time-taken	Temps en mil·lisegons que va requerir l'acció.
Protocol Version	cs-version	Versió del protocol HTTP o FTP que usa el client.
Host	cs-host	Nom del servidor, si té.
User Agent	cs(User-Agent)	Tipus de navegador que usa el client.
Cookie	cs(Cookie)	Contingut de les cookies enviades o rebudes, si hi ha.
Referrer	cs(Referrer)	Últim lloc visitat per l'usuari.
Protocol Substatus	sc-substatus	Codi d'error (sub-estat).

Taula A.5: Descripció de camps W3C Extended

A.3 Formats estàndard

Actualment no hi ha un estàndard de log mínimament acceptat. Cada aplicació o producte implementa els seus formats de logs de manera propietària. En aquesta secció es descriuen els intents que hi ha hagut, així com el format que es perfila com a estàndard de logs, encara no té una gran acceptació.

A.3.1 Intents d'estandardització

Hi ha hagut diversos intents de desenvolupar estàndards de logs i interoperabilitat però, per una raó o per una altra (alguns massa acadèmics o centrats en qüestions concretes), aquests intents no han tingut respatller.

Estàndard	Descripció	Inconvenients
Common Event Format (CEF)	Creat per ArcSight. Basat en parells atribut–valor. Pot aprofitar fitxers plans o syslog	És específic del proveïdor. Té un nombre petit d'atributs (només els necessaris utilitzats pel producte)
Distributed Audit Services (XDAS)	Iniciat en 1998 com un API per Unix. En 2008 Novell va reprendre el treball per fer la v2 i una norma més general.	Centrat gairebé en exclusiva en auditoria. És un API d'Unix.
Intrusion Detection Message Exchange Format (IDMEF)	Orientat a sistemes de detecció d'intrusions i als sistemes que interactuen amb ells.	Orientat gairebé en exclusiva als esdeveniments de detecció. El format és XML sobre BEEP.
Common Intrusion Detection Framework (CIDF)	Iniciat en 1998. Estructura similar al LISP. Protocol i API per a detecció d'intrusió	Centrat específicament en la detecció d'intrusions. Ja no segueix activa.

Taula A.6: Comparativa de diferents estàndards proposats

A.3.2 Common Event Expression (CEE)

Des de fa un temps existeix una iniciativa des de MITRE per intentar crear un estàndard de format per a esdeveniments generats pels diferents dispositius i aplicacions. MITRE, recolzada per tots els grans fabricants de productes de gestió d'esdeveniments, pretén crear un llenguatge únic per intentar evitar els problemes ja vists dels logs. El resultat és *Common Event Expression*, un llenguatge de log estàndard per a la interoperabilitat d'esdeveniments en sistemes electrònics. Els objectius de disseny de CEE eren estandarditzar els logs per simplificar el procés de gestió i mantenir la compatibilitat amb els entorns i productes actuals.

CEE es basa en una combinació de quatre elements: el transport, la sintaxi, la classificació i les recomanacions de logs, que poden ser tractats de forma independent. Per tant, CEE té una arquitectura basada en aquests quatre components:

- **CEE Dictionary and Taxonomy** (CDET). Defineix una col·lecció de camps d'esdeveniments i els tipus que es poden utilitzar dins dels registres d'esdeveniments i també defineix una col·lecció d'etiquetes que es poden utilitzar per classificar els esdeveniments. L'objectiu és proporcionar un llenguatge comú per ajudar a classificar i relacionar els registres pertanyents al mateix tipus d'esdeveniments.

- **Common Log Syntax** (CLS). Indica com es representen les dades de l'esdeveniment i el propi l'esdeveniment . La sintaxi de l'esdeveniment és el que un productor d'esdeveniments escriu i el que un consumidor d'esdeveniments processa.

- **Common Log Transport** (CLT). Proporciona un marc per al transport de logs. Un bon marc no només requereix registre estandarditzat d'esdeveniments; a més hauria de suportar codificació internacional de caràcters, interfícies estàndard per al registre d'esdeveniments i traça de log verificable i fiable.

- **Common Event Log Recommendations** (CELR). Ofereix una sèrie de recomanacions per a desenvolupadors i tècnics de sistemes per decidir quins esdeveniments i camps haurien de ser registrats en certes circumstàncies.

CEE distingeix entre transport i sintaxi. Mentre que la sintaxi és única, pot ser transmesa de diferents formes. Per exemple, la sintaxi podria expressar-se en XML i transportada mitjançant SOAP o SMTP. Algunes opcions de sintaxis i transport són complementàries, mentre que unes altres no funcionen molt bé juntes (com enviar XML sobre syslog o SNMP). S'utilitza Common Log Transport (CLT) per definir els potencials mitjans de transport per a una sintaxi.

Common Log Syntax (CLS) defineix un diccionari d'identificadors sintàctics que es poden usar per comunicar els detalls d'una instància d'un esdeveniment. Ja que no sembla possible crear una sintaxi per a cada possible situació, el diccionari defineix un conjunt universal de termes, juntament amb els seus tipus de dades i utilització. Usant el mateix diccionari s'assegura que els detalls de l'esdeveniment s'inclouen i són consistents.

Existeixen tres possibilitats de transmetre la informació depenent de paràmetres com la velocitat, la facilitat d'ús o l'expressivitat:

- **Velocitat**. Un format de log binari (amb la seva corresponent sintaxi amb camps de grandària fixa en un fitxer binari) és suficient per expressar la informació i és la forma més ràpida per realitzar log i transmetre les dades. No obstant això, no està dissenyat per als humans, i requereix d'aplicacions per codificar i decodificar els logs.

- **Facilitat d'ús**. Les sintaxis en text pla amb delimitadors i parells atribut–valor, com CSV[3], són fàcils de llegir i entendre tant per a humans com per a màquines. A més, aquest tipus de sintaxi ofereix facilitat de transport.

- **Expressivitat**. Les sintaxis basades en estructura, com XML, són potents i permeten representar estructures de dades complexes com a llistes o objectes niats. Els seus inconvenients estan en les limitades opcions de transport compatibles i en la sobrecàrrega que comporten. Tampoc són fàcils de llegir per a humans ni màquines.

Llistat A.1: Missatge de log CEE expressat en XML

```
<CEE xmlns="http://cee.mitre.org">
  <Event>
    <id>example-event-2</id>
    <time>2011-04-01T12:01:00-05:00</time>
    <action>download</action>
    <status>-</status>
    <p_sys_id>host.example.com</p_sys_id>
    <p_prod_id>product</p_prod_id>
    <Field name="tags"><tag>web</tag></Field>
    <Field name="file_name"><str>example.txt</str></Field>
    <Field name="file_data">
      <binary>RmlsZSBDb250ZW50Li4uAAo=</binary>
    </Field>
```

[3]Comma-separated values.

```
    </Event>
    <Augmentation order="1">
      <time>2011-04-01T14:11:53-04:00</time>
      <status>success</status>
      <p_sys_id>relay.example.com</p_sys_id>
      <p_prod_id>cee-relay</p_prod_id>
      <Field name="tags"><tag>hipaa</tag></Field>
    </Augmentation>
  </CEE>
```

Llistat A.2: Missatge de log CEE expressat en JSON i enviat amb syslog

```
<165>1 2011-04-01T17:01:20Z 10.10.0.1 process - example-event-1 cee:{"Event":{"id":"example-event-1",
"time":"t|2011-04-01T17:00:00.123456789Z","action":"g|remove","status":"g|failed","p_sys_id":
"host.example.com","p_prod_id":"cpe:2.3:Vendor:Product:Version:*:*:*:*:*:*","file_name":"example.txt",
"proc_dur":"d|PT.0014S","sess_id":"user1"}}
```

Apèndix B
Transport de log

B.1 syslog

La majoria dels dispositius de xarxa presenten la possibilitat de reexpedir els missatges de logs a un servidor de syslog en una xarxa IP.

Aquest protocol, desenvolupat a principis dels anys 80 per Eric Allman, estava dissenyat per treballar només amb sendmail. Des de llavors, s'ha implementat en la majoria dels dispositius de xarxa i la seva popularitat s'ha incrementat. La documentació actual de syslog està en la RFC5424.

Syslog permet als dispositius la reexpedició de missatges de log a través d'una xarxa IP per al seu emmagatzematge en un servidor remot. Com syslog és multiplataforma, permet que dispositius heterogenis puguin enviar logs a un únic repositori.

Nivell	Codi	Severitat
emerg	0	Situació de pànic
alert	1	Situació urgent
crit	2	Condicions crítiques
err	3	Altres condicions d'error
warning	4	Missatges d'avís
notice	5	Podria ser investigat
info	6	Missatge informatiu
debug	7	Per a depuració només

Taula B.1: Nivells de severitat de syslog (severitat decreixent)

Un missatge syslog es compon de tres camps:

- **PRI**. Aquest camp pot ser de 3, 4 o 5 caràcters limitats per angles (< >). El codi PRI es compon de 2 valors numèrics: el codi de **facilitat** i el codi de **severitat**. Existeixen uns codis estandarditzats per a la facilitat; així "0" es refereix a missatges del kernel, "1" per a missatges a nivell usuari, "2" per a missatges del sistema de correu i diversos més. De la mateixa forma, també existeixen codis de severitat: "0" són emergències o sistema no utilitzable, "1" són alertes i així fins a "7" (veure Taula B.1, i es troba en la RFC3164). El valor de PRI es calcula multiplicant el codi de facilitat per 8 i sumant-li el codi de severitat.

- **HEADER**. La capçalera conté dos camps: la **marca temporal** (TIMESTAMPT) i el **nom del servidor** (HOSTNAME). La marca temporal correspon amb la data i hora local del dispositiu que transmet en format [Mmm dd hh:mm:ss]. El nom del servidor correspon al nom del dispositiu (no ha de contenir espais ni tampoc el nom del domini), l'adreça IPv4 o l'adreça IPv6.

- **MSG**. El missatge té 2 camps: **l'etiqueta** (TAG) i **el contingut** (CONTENT). El primer representa el nom del programa o procés que va generar l'esdeveniment (màxim 32 caràcters). El camp CONTENT és d'utilització lliure i conté el detalls del missatge. Ja que cada procés, programa, aplicació o sistema operatiu va a escriure de manera independent, hi ha poca uniformitat en el contingut dels missatges syslog.

```
<34>Oct 11 22:14:15 mymachine su: 'su root' failed for lonvick on /dev/pts/8
```

L'arquitectura d'un sistema basat en syslog consisteix en uns dispositius que generen missatges, denominats remitents, i altres dispositius que els reben, denominats recol·lectors (també se'ls sol anomenar, incorrectament, servidors syslog). El protocol està dissenyat simplement per transportar missatges d'esdeveniments entre remitent i recol·lector.

Els principals problemes que té syslog són els següents:

- syslog utilitza el protocol UDP (protocol no orientat a connexió), i no s'assegura que els missatges arribin al destinatari.
- Els missatges no estan xifrats i en viatjar per la xarxa en clar són susceptibles a ser vists (per exemple, amb un *sniffer*).
- No hi ha cap autenticació de qui és el remitent.

Existeixen múltiples implementacions de syslog encara que aquí ens centrarem en syslogd, rsyslog i syslog-ng.

B.1.1 syslogd

Syslogd és la base per al transport de missatges de log que utilitzen molts sistemes de gestió de log. L'arquitectura syslog consisteix en tres parts:

- syslogd – el dimoni (*daemon*) de syslog.
- Openlog – biblioteca de rutines que envien missatges a syslogd.
- Logger – comando per enviar missatges de log des de la línia de comandos.

Syslogd és un dimoni que s'executa contínuament en un sistema. Les aplicacions envien els seus missatges a un fitxer especial anomenat /dev/log, on syslogd llegeix els missatges enviats i, segons sigui la seva configuració (syslog.conf), enruta aquests missatges a les destinacions definides. El fitxer syslog.conf és un fitxer de text que controla el comportament de syslogd. Tots els missatges generats amb syslogd porten una marca de temps

B.1.2 syslog-ng

Syslog-ng és una implementació de codi obert de syslog que amplia el model inicial amb un filtratge millorat de missatges, una configuració més flexible i altres característiques afegides. Syslog-ng va ser desenvolupat per Balazs Scheidler en 1998 com un projecte de migració del codi de nsyslogd a Linux.

Entre unes altres, aquestes són algunes extensions a syslogd:

- Marques de temps ISO 8601 amb granularitat de microsegons i informació de zona horària.

- Transport fiable usant TCP.
- Encriptació TLS.

Syslog-ng presenta noves funcionalitats com:

- Enviament de log directament a bases de dades.
- Classificació dels missatges de log entrants i, al mateix temps, extracció d'informació estructurada a partir de missatges syslog sense estructura.

B.1.3 rsyslog

Rsyslog és una versió millorada de syslogd amb llicència GPL. Desenvolupat per Rainer Gerhards en 2004, l'objectiu del projecte rsyslog era substituir directament a syslogd amb un dimoni syslod fiable i amb característiques millorades. La fiabilitat s'aconsegueix utilitzant TCP com a protocol de transport.

syslog	**Versió**	**Plataforma**	**UDP**	**TCP**	**Ports**	**Monitoratge de fitxers**
syslogd	Tots	Sistemes BSD	Si	Si	Si	No
syslogd	Tots	Linux	Si	No	No	No
syslog-ng	2.x	Tots	Si	Si	Si	No
syslog-ng	3.x	Tots	Si	Si	Si	Si
rsyslog	5.6.0	Tots	Si	Si	Si	Si

Taula B.2: Característiques de diferents sistemes syslog

B.2 Alternatives a syslog

Tot i que syslog està pràcticament present en la majoria dels dispositius, no està suportat nativament per Microsoft Windows, encara que existeixen aplicacions de tercers que permeten utilitzar syslog. Com a alternativa a syslog, anem a comentar SNMP.

B.2.1 Simple Network Management Protocol (SNMP)

El *Simple Network Management Protocol* (SNMP) és un protocol estàndard d'internet dissenyat per facilitar la gestió de dispositius en xarxes IP. SNMP està suportat per pràcticament tots els dispositius connectables a una xarxa, incloent enrutadors (*routers*), commutadors (*switchs*), impressores, estacions de treball, servidors, mòdems, sistemes d'alimentació ininterrompuda (SAI), telèfons VoIP, *smartphones*, etc. SNMP pot obtenir informació que va des de l'estat bàsic d'un dispositiu a les estadístiques de tràfic. SNMP utilitza un conjunt relativament simple d'operacions. Aquestes operacions permeten consultar una informació concreta o modificar un paràmetre específic d'un dispositiu. SNMP si que està suportat nativament per MS Windows.

Existeixen tres versions SNMP. SNMPv1 és la primera versió de protocol i està definida en la RFC1157. SNMPv2 va augmentar el tipus d'informació que pot recopilar-se i està definida en RFC3416, RFC3417 i RFC3418. SNMPv3 és la versió més recent de SNMP i inclou un mecanisme de seguretat: mentre que SNMPv1 i SNMPv2 passen la informació en clar, SNMPv3 pot configurar-se per encriptar

tots els paquets del protocol. SNMPv3 es defineix en les RFC3410, RFC3411, RFC3412, RFC3413, RFC3414, RFC3415, RFC3416, RFC3417, RFC3418 i RFC2576.

SNMP té dos components: l'agent i el gestor. Els agents es configuren en els dispositius que es vol monitoritzar. El gestor és el sistema, normalment un sistema de gestió de xarxa[1], que genera peticions d'informació i la rep de diversos dispositius. SNMP també es pot configurar per enviar *traps*, que són alarmes enviades pels dispositius al gestor indicant que alguna cosa no està funcionant ben. La *trap* pot equiparar-se als missatges de syslog. Existeixen set *traps* genèriques, encara que cada dispositiu té capacitat per enviar *traps* més específiques que faciliten la resolució de la fallada.

Nom (nombre)	**Significat**
coldStart (0)	Indica que l'agent s'ha reiniciat.
warmStart (1)	Indica que l'agent s'ha reinicialitzat per ell mateix.
linkDown (2)	Indica que una interfície d'un dispositiu ha caigut.
linkup (3)	Indica que una interfície d'un dispositiu s'ha aixecat.
authenticationFailure (4)	Indica un intent de consultar al dispositiu amb credencials d'autenticació errònies.
egpNeighborLoss (5)	Indica que el veí EGP ha caigut.
enterpriseSpecific (6)	Indica que és una *trap* específica del fabricador

Taula B.3: *Traps* genèriques SMNP

[1] Network Management System (NMS)

Apèndix C
Criteris d'avaluació

C.1 Criteris genèrics

1. Durabilitat
    1.1. Maduresa
        1.1.1. Antiguitat
            0 Menys de 3 mesos.
            1 Entre 3 mesos i 3 anys.
            2 Més de 3 anys.
        1.1.2. Estabilitat
            0 Programari inestable amb nombroses versions i pegats.
            1 Programari estable però versions antiquades.
            2 Programari estable. La noves versions afegeixen funcionalitat.
        1.1.3. Problemes coneguts
            0 Programari amb problemes que ho fan inutilitzable.
            1 Programari sense problemes greus.
            2 Programari amb bona gestió de problemes greus en el passat.
    1.2. Adopció
        1.2.1. Popularitat
            0 Pocs usuaris.
            1 Ús apreciable.
            2 Nombrosos usuaris i referències.
        1.2.2. Comunitat
            0 Sense comunitat o sense activitat.
            1 Comunitat amb activitat.
            2 Comunitat activa amb gran activitat en els fòrums i col·laboradors.
2. Documentació
    0 Sense documentació.
    1 Documentació desactualitzada o escassament detallada.
    2 Documentació actualitzada i adaptada a diferents tipus d'usuaris.
3. Explotabilitat
    3.1. Facilitat d'ús
        0 Difícil d'usar, es necessita coneixement de la funcionalitat.
        1 Uso auster i tècnic.
        2 Interfície d'usuari elaborat i assistit.
    3.2. Administració
        0 Sense funcionalitat per a l'administració.
        1 Funcionalitat d'administració incompleta o millorable.
        2 Funcionalitat d'administració completa i fàcil d'usar.

4. Adaptabilitat tècnica

4.1. Modularitat

0 Programari monolític.
1 Presència de mòduls d'alt nivell.
2 Concepció modular.

4.2. Modificació de codi

0 Manualment.
1 Recompilació possible però complicada sense eines ni documentació.
2 Recompilació amb eines (com make, ANT, ...) i amb documentació incorporada.

4.3. Extensió de codi

0 Qualsevol modificació requereix recompilació.
1 Dissenyat per a extensió estàtica, però requereix recompilació.
2 Basat en plugin, dissenyat per a extensió dinàmica sense recompilació.

C.2 Criteris funcionals

5. Entrada de dades

5.1. Formats

5.1.1. syslog

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.1.2. Windows Event

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.1.3. W3C Extended

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.1.4. NCSA (access log)

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.1.5. CEE

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.1.6. Capacitat d'adaptar-se a nous formats

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.2. Transport

5.2.1. Disposa d'agents per als principals sistemes operatius, dispositius de xarxa i aplicacions més freqüents

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.2.2. Utilitza TCP i UDP per rebre syslog, syslog-ng i variants

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.2.3. SNMP traps

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.2.4. SOAP
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.2.5. Encriptació de dades
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

5.2.6. Autenticació de l'origen
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6. Emmagatzematge

6.1. Traça de l'origen de cada log
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.2. Reconeix i normalitza marques de temps
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.3. Manteniment de log original (inalterat)
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.4. Suport

6.4.1. Sistema de fitxers
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.4.2. Base de dades

6.4.2.1. Relacional
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.4.2.2. XML
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.4.2.3. Orientada a objectes
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.4.2.4. NoSQL
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.4.3. Datawarehouse
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

6.5. Compressió de logs
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

6.6. Arxivament de logs
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7. Cerca i anàlisi

7.1. Recerques bàsiques

7.1.1. Operadors booleans
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.1.2. Selecció de rangs temporals
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.1.3. Navegació en resultats
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.1.4. Salvaguarda de cerques
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.2. Recerques avançades

7.2.1. Subconsultes
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.2.2. Operadors estadístics
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.2.3. Operadors de conjunt
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.2.4. Operadors de correlació
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.2.5. Operadors per agrupament (clustering)
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.3. Informes

7.3.1. Construir informes a partir de les cerques
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.3.2. Opcions de pivotatge i aprofundiment (drill-down)

0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.3.3. Informes programats
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.3.4. Informes en temps real
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.4. Visualització

7.4.1. Expressivitat de la visualització
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.4.2. Múltiples vistes sincronitzades
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.4.3. Diagrames en temps real
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.4.4. Tipus de diagrames
0 Entre 1 i 3 tipus de diagrames.
1 Entre 4 i 7 tipus de diagrames.
2 Més de 8 tipus de diagrames.

7.5. Correlació d'esdeveniments

7.5.1. Correlació manual
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.5.2. Correlació en temps real
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.5.3. Seguiment d'activitat d'usuari
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.6. Quadres de comandament

7.6.1. Posicionar lliurement els objectes en el quadre de comandament
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.6.2. Quadres de comandament millorats
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.6.3. Gestió d'estils
0 Característica no implementada.
1 Característica disponible però limitada o no eficaç.
2 Característica completament aplicada i usable.

7.7. Gestió d'alertes

7.7.1. Gestió i emmagatzematge d'alertes
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.7.2. Regles de gestió d'alertes
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

7.7.3. Enviament d'alertes
- **0** Característica no implementada.
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

C.3 Criteris tècnics

8. Facilitat d'instal·lació
   - **0** La instal·lació requereix molt temps, els scripts d'instal·lació són complexos o estan mal documentats.
   - **1** La instal·lació requereix de molt temps però està suficientment documentada.
   - **2** La instal·lació és fàcil i ben documentada.
9. Facilitat de configuració
   - **0** Requereix temps i és complexa.
   - **1** Requereix temps però no és complicada.
   - **2** Automàtica, l'eina està llista per usar sense necessitat de configuració.
10. Facilitat d'actualització
    - **0** Procés d'actualització inexistent.
    - **1** Procés d'actualització complex o no està funcionant.
    - **2** Procés d'actualització funciona correctament, és senzill o automàtic.
11. Dependència de llibreries externes
    - **0** Moltes dependències (de llibreries externes o programari) que cal instal·lar i mantenir.
    - **1** Poques dependències (de llibreries externes o programari) que són conegudes i fàcils d'instal·lar.
    - **2** El programari no té dependències o és només del navegador web.
12. Interoperabilitat
    12.1. Disponibilitat d'API
    - **0** No hi ha API disponible.
    - **1** API disponible però no és completa o no està ben documentada.
    - **2** API disponible, completa i ben documentada.

    12.2. Serveis web
    - **0** No hi ha tal característica.
    - **1** Capacitats limitades d'integració.
    - **2** Bones capacitats.
13. Seguretat
    13.1. Autenticació
    13.1.1. LDAP
    - **0** No està disponible autenticació per LDAP.
    - **1** Disponible autenticació per LDAP però limitada.
    - **2** Disponible autenticació per LDAP.

13.1.2. Directori Actiu
- **0** No està disponible autenticació per Directori Actiu.
- **1** Disponible autenticació per Directori Actiu però limitada.
- **2** Disponible autenticació per Directori Actiu.

13.1.3. SSO
- **0** Característica no implementada
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

13.1.4. Base de dades
- **0** No està disponible autenticació per base de dades .
- **1** Disponible autenticació per base de dades però limitada.
- **2** Disponible autenticació per base de dades.

13.2. Gestió d'usuaris

13.2.1. Gestió de permisos d'accés
- **0** Característica no implementada
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

13.2.2. Gestió de grups d'usuaris
- **0** Característica no implementada
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

14. Interfície d'usuari

14.1. Personalització de la interfície
- **0** La interfície d'usuari no pot personalitzar-se.
- **1** La interfície d'usuari és difícil de personalitzar.
- **2** La interfície d'usuari és fàcilment personalitzable.

14.2. Internalització
- **0** Un idioma.
- **1** Internacionalització possible, pocs idiomes disponibles.
- **2** Disponible en molts idiomes.

14.3. Basada en navegador
- **0** Característica no implementada
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

14.4. Línia de comandos (CLI)
- **0** Característica no implementada
- **1** Característica disponible però limitada o no eficaç.
- **2** Característica completament aplicada i usable.

Apèndix D
Càlcul de puntuació

Aquest apèndix mostra la taula utilitzada per calcular les puntuacions de cada producte candidat basant-se en els criteris de valoració.

	Splunk	Octopussy	Graylog2	Loggly	Logentries
Criteris genèrics					
1. Durabilitat					
1.1. Maduresa					
1.1.1. Antiguitat	2.0	2.0	1.0	1.0	1.0
1.1.2. Estabilitat	2.0	2.0	2.0	2.0	2.0
1.1.3. Problemes coneguts	1.0	1.0	1.0	1.0	1.0
1.2. Adopció					
1.2.1. Popularitat	2.0	1.0	1.0	1.0	0.0
1.2.2. Comunitat	2.0	1.0	1.0	1.0	0.0
2. Documentació	2.0	1.0	1.0	2.0	1.0
3. Explotabilitat					
3.1. Facilitat d'ús	1.0	1.0	1.0	1.0	1.0
3.2. Administració	2.0	2.0	1.0	1.0	1.0
4. Adaptabilitat tècnica					
4.1. Modularitat	1.0	1.0	2.0	2.0	0.0
4.2. Modificació de codi	0.0	1.0	2.0	0.0	0.0
4.3. Extensió de codi	0.0	1.0	1.0	1.0	0.0
Criteris funcionals					
5. Entrada de dades					
5.1. Formats					
5.1.1. syslog	2.0	2.0	2.0	2.0	2.0
5.1.2. Windows Event	2.0	1.0	1.0	1.0	1.0
5.1.3. W3C Extended	2.0	0.0	1.0	0.0	0.0
5.1.4. NCSA (access log)	2.0	2.0	1.0	1.0	2.0
5.1.5. CEE	1.0	0.0	0.0	0.0	0.0
5.1.6. Capacitat d'adaptar-se a nous formats	2.0	2.0	2.0	1.0	1.0
5.2. Transport					
5.2.1. Disposa d'agents (...)	1.0	0.0	1.0	1.0	2.0
5.2.2. Utilitza TCP i UDP (...)	2.0	2.0	2.0	2.0	2.0
5.2.3. SNMP traps	2.0	2.0	0.0	0.0	0.0
5.2.4. SOAP	0.0	0.0	0.0	0.0	0.0
5.2.5. Encriptació de dades	0.0	0.0	0.0	1.0	1.0
5.2.6. Autenticació de l'origen	0.0	0.0	0.0	0.0	1.0
6. Emmagatzematge					
6.1. Traça de l'origen de cada log	1.0	0.0	0.0	0.0	1.0
6.2. Reconeix i normalitza marques de temps	2.0	1.0	1.0	1.0	1.0
6.3. Manteniment de log original (inalterat)	2.0	1.0	1.0	0.0	0.0
6.4. Suport					
6.4.1. Sistema de fitxers	0.0	0.0	0.0	0.0	0.0
6.4.2. Base de dades					
6.4.2.1. Relacional	0.0	2.0	0.0	0.0	0.0
6.4.2.2. XML	0.0	0.0	0.0	0.0	0.0
6.4.2.3. Orientada a objectes	0.0	0.0	0.0	0.0	0.0
6.4.2.4. NoSQL	2.0	0.0	2.0	2.0	2.0
6.4.3. Datawarehouse	0.0	0.0	0.0	0.0	0.0
6.5. Compressió de logs	2.0	0.0	0.0	0.0	0.0

	Splunk	Octopussy	Graylog2	Loggly	Logentries
6.6. Arxivament de logs	1.0	0.0	0.0	2.0	0.0
7. Cerca i anàlisi					
7.1. Recerques bàsiques					
7.1.1. Operadors booleans	2.0	1.0	1.0	2.0	2.0
7.1.2. Selecció de rangs temporals	2.0	1.0	2.0	2.0	2.0
7.1.3. Navegació en resultats	2.0	2.0	1.0	1.0	1.0
7.1.4. Salvaguarda de cerques	2.0	2.0	2.0	2.0	1.0
7.2. Recerques avançades					
7.2.1. Subconsultes	2.0	1.0	1.0	2.0	0.0
7.2.2. Operadors estadístics	2.0	1.0	0.0	0.0	0.0
7.2.3. Operadors de conjunt	2.0	1.0	0.0	0.0	0.0
7.2.4. Operadors de correlació	2.0	0.0	0.0	0.0	1.0
7.2.5. Operadors per agrupament (clustering)	2.0	0.0	0.0	0.0	0.0
7.3. Informes					
7.3.1. Construir informes a partir de les cerques	2.0	2.0	0.0	1.0	2.0
7.3.2. Opcions de pivotatge i aprofundiment	1.0	0.0	0.0	0.0	1.0
7.3.3. Informes programats	2.0	2.0	0.0	0.0	0.0
7.3.4. Informes en temps real	2.0	2.0	1.0	2.0	2.0
7.4. Visualització					
7.4.1. Expressivitat de la visualització	2.0	1.0	1.0	1.0	1.0
7.4.2. Múltiples vistes sincronitzades	2.0	0.0	0.0	0.0	0.0
7.4.3. Diagrames en temps real	2.0	0.0	1.0	2.0	2.0
7.4.4. Tipus de diagrames	1.0	1.0	0.0	1.0	0.0
7.5. Correlació d'esdeveniments					
7.5.1. Correlació manual	2.0	1.0	0.0	1.0	1.0
7.5.2. Correlació en temps real	2.0	0.0	0.0	1.0	1.0
7.5.3. Seguiment d'activitat d'usuari	1.0	0.0	0.0	0.0	0.0
7.6. Quadres de comandament					
7.6.1. Posicionar lliurement els objectes (...)	1.0	0.0	0.0	0.0	0.0
7.6.2. Quadres de comandament millorats	1.0	0.0	0.0	0.0	0.0
7.6.3. Gestió d'estils	2.0	0.0	0.0	0.0	0.0
7.7. Gestió d'alertes					
7.7.1. Gestió i emmagatzematge d'alertes	2.0	2.0	2.0	0.0	2.0
7.7.2. Regles de gestió d'alertes	2.0	2.0	1.0	0.0	1.0
7.7.3. Enviament d'alertes	2.0	2.0	1.0	0.0	1.0
Criteris tècnics					
8. Facilitat d'instal·lació	2.0	1.0	2.0	2.0	2.0
9. Facilitat de configuració	1.0	1.0	1.0	2.0	2.0
10. Facilitat d'actualització	2.0	2.0	1.0	2.0	2.0
11. Dependència de llibreries externes	1.0	0.0	1.0	2.0	2.0
12. Interoperabilitat					
12.1. Disponibilitat d'API	2.0	1.0	0.0	2.0	1.0
12.2. Serveis web	0.0	0.0	0.0	1.0	0.0
13. Seguretat					
13.1. Autenticació					
13.1.1. LDAP	2.0	2.0	0.0	0.0	0.0
13.1.2. Directori Actiu	2.0	0.0	0.0	0.0	0.0
13.1.3. SSO	1.0	0.0	0.0	0.0	0.0
13.1.4. Base de dades	0.0	0.0	2.0	1.0	1.0
13.2. Gestió d'usuaris					
13.2.1. Gestió de permisos d'accés	2.0	1.0	1.0	1.0	1.0
13.2.2. Gestió de grups d'usuaris	2.0	0.0	1.0	0.0	0.0
14. Interfície d'usuari					
14.1. Personalització de la interfície	1.0	2.0	0.0	0.0	0.0
14.2. Internalització	1.0	2.0	0.0	0.0	0.0
14.3. Basada en navegador	2.0	2.0	2.0	2.0	2.0
14.4. Línia de comandos (CLI)	2.0	0.0	0.0	1.0	0.0

Apèndix E
Cas pràctic

Com a punt final del Projecte, es va a procedir a explicar un petit exemple pràctic per a comprovar com es podria realitzar una exploració de dades molt senzilla, i veure com les eines ens permeten obtenir informació dels logs. Per a aquest exemple, es va a utilitzar el producte que ha obtingut la major puntuació en el procés de valoració: Splunk. I perquè siga més comprensible, s'usarà un fitxer de log procedent d'un servidor web amb Apache.

E.1 Consideracions prèvies

Existeixen pocs conceptes en Splunk però és convenient tenir-los clars. Splunk llig dades d'una font (*source*), com un port o un fitxer; en un servidor (*host*), classifica la font en un tipus de font (*sourcetype*), llavors obté les marques temporals (*timestampt*), divideix la font en esdeveniments (*event*) individuals que poden ser d'una o múltiples línies, i escriu cada esdeveniment en un índex per a poder recuperar-ho més tard mitjançant una cerca (*search*).

Quan es llança una cerca, es recuperen els esdeveniments indexats corresponents, els camps (*fields*) s'extrauen del text de l'esdeveniment i es classifica l'esdeveniment segons la corresponent definició de tipus d'esdeveniment (*eventtype*). Els esdeveniments obtinguts en una cerca poden usar-se, mitjançant el llenguatge de cerca, per a generar informes (*reports*) dins de quadres de comandament (*dashboard*).

Splunk conté diverses aplicacions (*apps*); però la més rellevant, en aquest moment, és l'aplicació "Search", que és la interfície de cerca genèrica. Una vegada iniciada sessió en Splunk, cal seleccionar l'aplicació "Search" i ja es pot començar a cercar.

Per exemple, si cerquem els esdeveniments que contenen errors HTTP 404[1], cal escriure:

```
source="access_combined.log" HTTP 404
```

Els termes de cerca se suposa que porten l'operador AND per defecte. Splunk pot usar els operadors lògics AND, OR i NOT (han d'escriure's en majúscules), així com amb parèntesis per forçar l'agrupació de valors. També es pot utilitzar el caràcter comodí *; per exemple, per obtenir els esdeveniments que tenen els codis d'estat HTTP 4XX i 5XX cal escriure:

```
source="access_combined.log" http (4* OR 5*)
```

Quan Splunk indexa les dades, automàticament afegeix camps o atributs a cada esdeveniment. Això ho fa basant-se en patrons de text freqüents en logs, però també es pot fer manualment afegint regles per a camps addicionals. Per restringir la cerca es poden afegir parells del tipus camp=valor a la cerca.

```
source="access_combined.log" status=404
```

[1] 404 Pàgina no trobada.

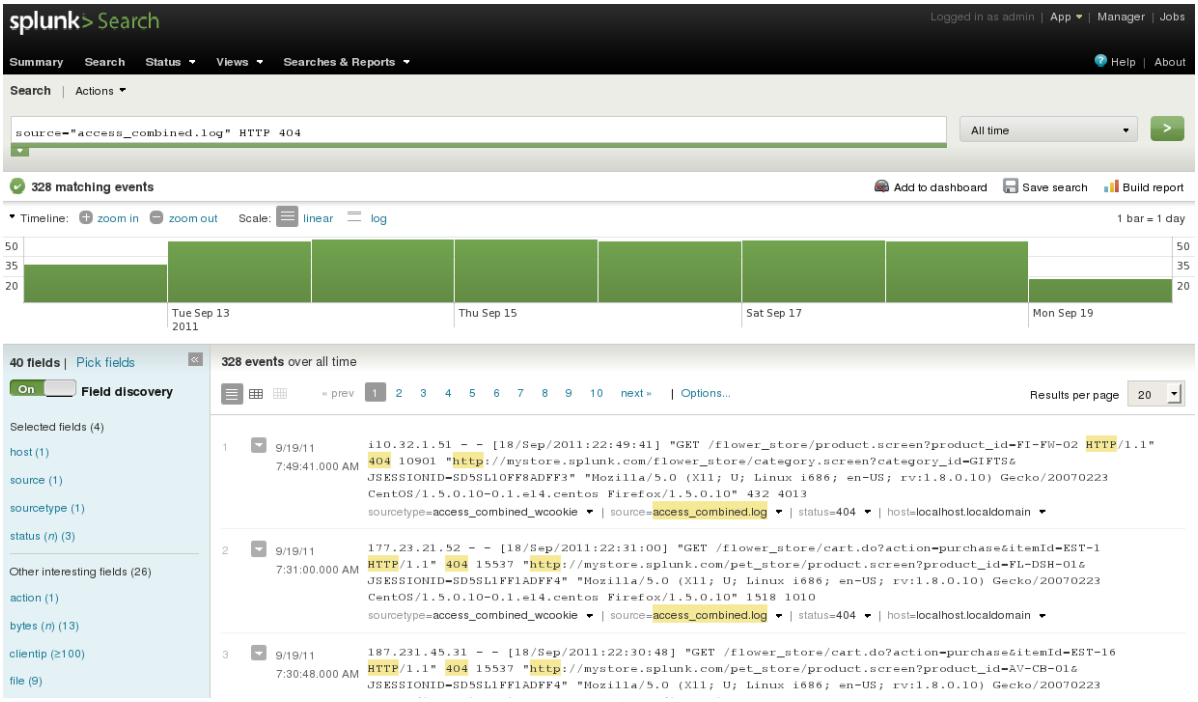

Figura E.1: Captura de Splunk. Pantalla principal de cerca

Existeixen desenes d'operacions que es poden realitzar amb les dades. L'operador | permet que els resultats d'una cerca es transfereixin a altres comandos per filtrar, modificar i agrupar resultats. L'operador "eval" permet calcular un camp nou basant-se en uns altres.

Ara que ja hem realitzat algunes cerques, podem construir informes amb elles. Existeixen tres tipus bàsics d'informes:

- "Values over time" per a diagrames per estudiar els valors d'un camp en un rang temporal.
- "Top values" per als valors del camp més comuns.
- "Rare values" per als valors del camp més estranys.

La pàgina informes permet realitzar un ajust de format del diagrama, guardar-ho, imprimir-ho i exportar els resultats.

E.2 Descobrir informació dins dels logs

Partim d'un log generat per un servidor web d'Apache. Aquest log conté centenars d'entrades que no sabem molt bé què contenen, generalment accessos a pàgines o a aplicacions web. Realitzem una cerca completa sobre aquest log i obtindríem alguna cosa similar a la Figura E.2.

Aparentment no hi ha gens estrany, simplement apareix un diagrama amb el nombre d'accessos durant un rang de temps determinat. Però Splunk ja ha realitzat per nosaltres un descobriment de camps, doncs és capaç de detectar que es tracta d'un log d'accés propi d'Apache. Si premem sobre el camp useragent[2] ens dóna els diferents agents d'usuari (generalment, navegadors) amb els que s'ha accedit al servidor. El resultat hauria de ser alguna cosa semblat a la Figura E.3.

[2]Si no aparegués, habilitar el descobriment de camps i afegir el camp useragent a la llista.

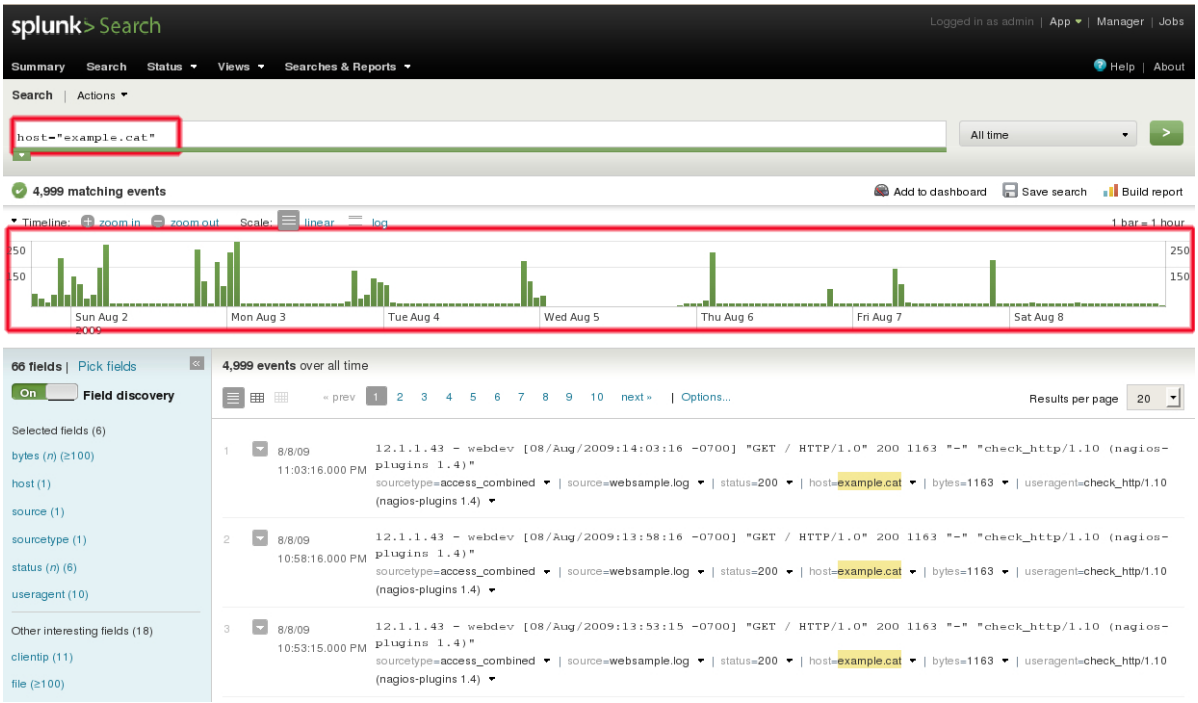

Figura E.2: Captura de Splunk. Cerca bàsica

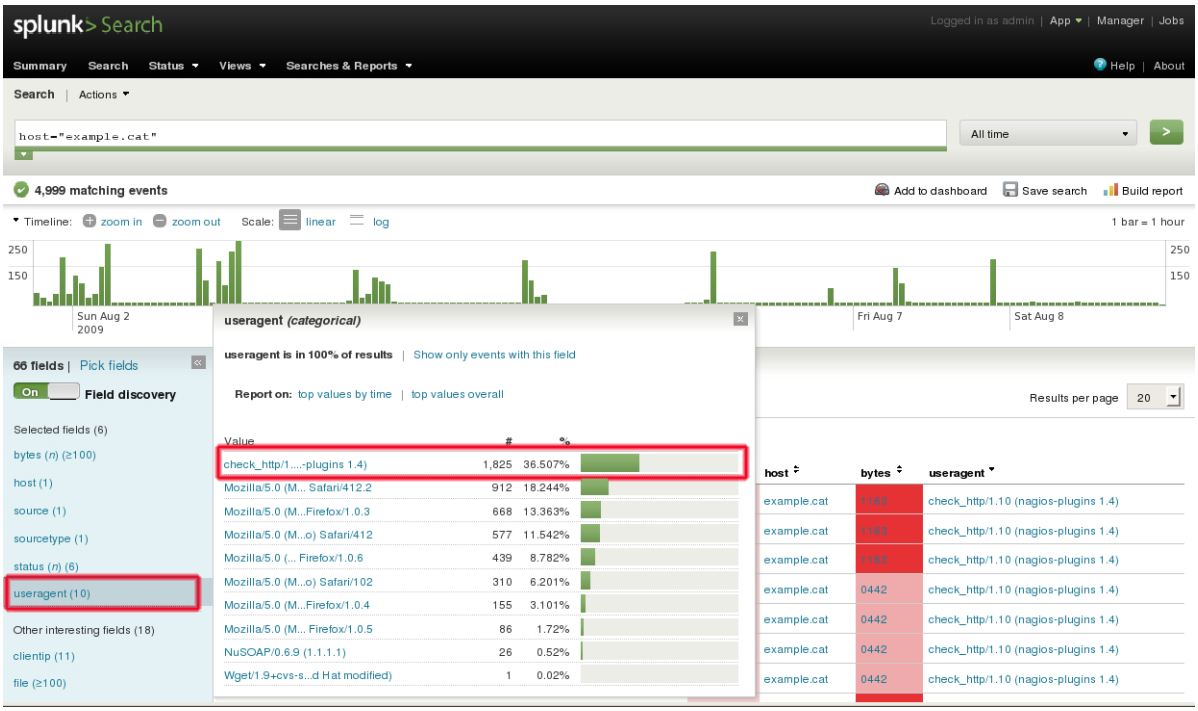

Figura E.3: Captura de Splunk. Cerca bàsica

En vista de la Figura, observem que el 36.5% dels accessos al servidor provenen d'un *plug–in* de Nagios, en lloc de ser un navegador. Si aprofundim més en les dades (veure Figura E.4), observem que tots aquests accessos provenen d'un únic servidor Nagios i que realitza uns 300 accessos al dia i que no generen error. En aquest cas es tracta del nostre sistema de monitoratge que està mal configurat i que,

donat el nostre volum de pàgines, no és necessària tal quantitat d'accessos.

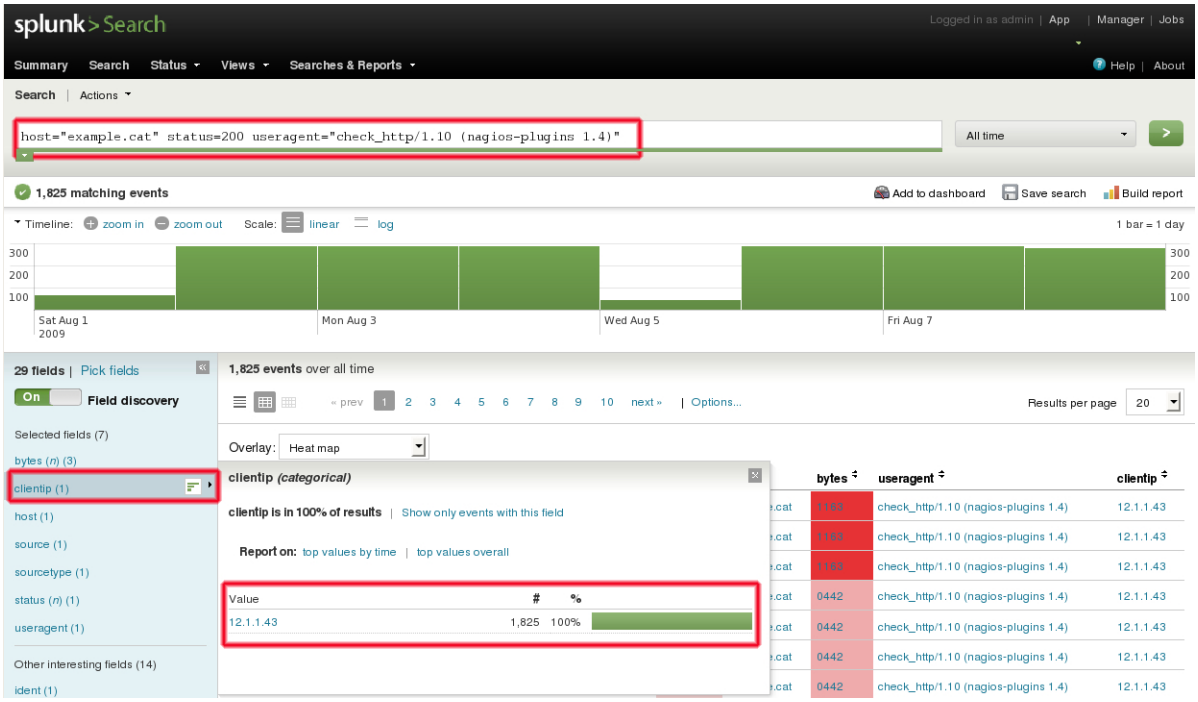

Figura E.4: Captura de Splunk. Cerca bàsica

Amb les dades anteriors podem fer un informe (veure Figura E.5) que mostri diàriament els accessos al servidor web amb els tipus de useragent (és a dir, navegador) apilats (la zona en verd superior és el tràfic generat pel plug-in de Nagios).

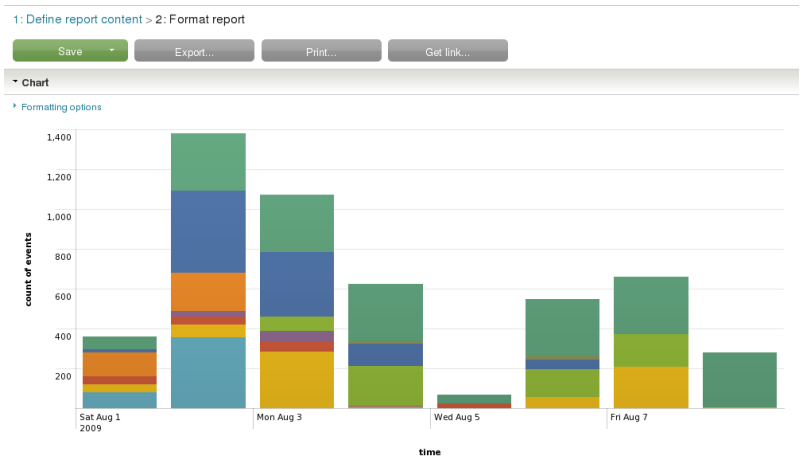

Figura E.5: Captura de Splunk. Informe

Amb aquest exemple s'ha tractat de demostrar que simplement amb una cerca bàsica i les capacitats de visualització resulta molt més senzill obtenir informació dels logs que es generen diàriament, i que aquesta informació pot ser útil per gestionar els entorns d'explotació.