# SCRIPTIE

## Log Management

| | |
|---|---|
| **Student** | Jochem Beekman |
| **Student nummer** | 1614706 |
| **Student email** | Jochem.beekman@student.hu.nl |
| **Bedrijf** | BIT |
| **Bedrijfsbegeleider** | Sander Smeenk |
| **Opleiding** | Systeem en Netwerkbeheer |
| **Afstudeerbegeleider** | Harry Beerlage |
| **Afstudeerperiode** | 5 |
| **Datum** | 04-07-2015 |
| **Versie** | 1.0 |

## Documenteigenschappen

| | |
|---|---|
| Documenttitel | **Scriptie** |
| Datum van uitgave | **04-07-2015** |
| Plaats van uitgave | **Ede** |
| Versie van uitgave | **1.0** |
| Bestandsnaam | **Scriptie-JochemBeekman-1614706** |
| Uitgever | **Jochem Beekman** |
| Contactpersoon | **Jochem Beekman** |
| Mail | Jochem.beekman@student.hu.nl |

## Documentuitgaven

| *Versie* | *Datum* | *Auteur* | *Doelgroep* | *Omschrijving* |
|---|---|---|---|---|
| **0.1** | **25-05-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider** | **Concept versie** |
| **0.2** | **10-06-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider Examencommissie** | **Tweede concept** |
| **0.3** | **18-06-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider Examencommissie** | **Eerste uitgave** |
| **0.4** | **24-06-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider Examencommissie** | **Feedback Harry verwerkt** |
| **0.5** | **02-07-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider Examencommissie** | **Feedback Teun verwerkt** |
| **0.6** | **03-07-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider Examencommissie** | **Feedback Sander verwerkt** |
| **1.0** | **04-07-2015** | **Jochem Beekman** | **Bedrijfsbegeleider Schoolbegeleider Examencommissie** | **Final** |

## Contact gegevens BIT

Galileïlaan 19
6716 BP Ede
The Netherlands
T: +31 (0)318 648 688
F: +31 (0)318 643 334

# Voorwoord

Voor u ligt de scriptie Log Management. De scriptie beschrijft het onderzoek dat is uitgevoerd bij BIT in Ede. Het onderzoek betreft de realisatie van een log management systeem dat vele miljoenen logevents per dag verwerkt en beschikbaar stelt. Met deze scriptie sluit ik mijn opleiding System & Network Engineering aan de Hogeschool Utrecht af.

Na een lange zoektocht naar een uitdagende en leuke afstudeeropdracht ben ik bij BIT terecht gekomen. BIT had aan het begin van mijn zoektocht al mijn interesse en ik ben dan ook zeer blij dat ik bij dit bedrijf mijn afstudeerstage heb kunnen uitvoeren. Het uitvoeren van zo'n groot project was een zeer leerzame en leuke ervaring.

Bij dezen wil ik graag mijn begeleiders, Sander Smeenk en Teun Vink, bedanken voor de gegeven begeleiding en ondersteuning bij dit project. Ik heb met veel plezier aan het project gewerkt en dit komt mede door de leuke collega's en fijne werksfeer.

Verder zou ik Harry Beerlage willen bedanken voor de begeleiding en gegeven feedback op de ingeleverde documenten. Deze feedback heb ik als waardevol ervaren en was een belangrijke input voor het uiteindelijke resultaat.

Ik wens u veel leesplezier toe.

Jochem Beekman

# Managementsamenvatting

In dit document staat het afstudeerproject in al zijn facetten beschreven. De afstudeeropdracht betreft een ontwerp opdracht die is uitgevoerd bij BIT in Ede.

BIT is een zakelijke internet service provider en biedt verschillende diensten aan waaronder hosting, internetverbindingen en colocatie. Het netwerk van BIT bestaat uit vele verschillende servers, devices en services die een grote hoeveelheid logdata genereren. In de huidige situatie worden de logevents op een syslog server opgeslagen. Hierdoor loopt BIT tegen een aantal beperkingen. De grote hoeveelheid logdata zorgt ervoor dat het doorzoeken hiervan moeizaam verloopt en een tijdrovende klus is. Daarnaast wordt er belangrijke informatie over het hoofd gezien als gevolg van de grote stroom aan logdata.

BIT wenst een systeem dat de logdata centraal opslaat, verwerkt en inzichtelijk maakt zodat het doorzoeken van logevents snel en efficiënt verloopt. Het type opdracht is een ontwerpopdracht wat inhoud dat er een functioneel ontwerp, technisch ontwerp en implementatieplan gerealiseerd is. Het functioneel en technisch ontwerp beschrijven de functionaliteiten en de technische aspecten van het systeem. Het implementatieplan beschrijft stap voor stap hoe het log management systeem geïmplementeerd kan worden.

Het is BIT aanbevolen het log management systeem te implementeren volgens het implementatieplan waarbij gebruik wordt gemaakt van de Graylog producten. Graylog bestaat uit drie producten die de logevents ontvangen, verwerken en beschikbaar stellen.

Aan de hand van de eisen en wensen is een functioneel ontwerp gerealiseerd waarin de huidige situatie van BIT, de gewenste situatie en de functionaliteiten van het systeem beschreven staan. Op basis van het functioneel ontwerp is een technisch ontwerp gerealiseerd dat de technische aspecten van het log management systeem beschrijft. Daarnaast bevindt zich hierin een productselectie.

Uit de productselectie is gebleken dat Graylog het product is dat het beste aansluit aan de eisen en wensen van BIT. De selectieprocedure bestaat uit een longlist en shortlist. De longlist bevat vier producten waarvan er twee in de shortlist terecht zijn komen. Uit de shortlist is Graylog als beste getest. Het besluit welke producten uit de longlist in de shortlist worden opgenomen en welk product uit de shortlist geïmplementeerd zal worden is gerealiseerd door een vergelijking. Het vergelijken van de producten is door middel van de decision matrix uitgevoerd.

Vervolgens is Graylog uitgebreid getest door middel van een proof of concept. Hierin is gekeken of het product daadwerkelijk voldoet aan de eisen en wensen van BIT. Samen met de opdrachtgever is een testplan uitgevoerd. De resultaten van de geteste onderdelen zijn positief bevonden.

Aan de hand van het gerealiseerde implementatieplan kunnen de engineers het log management systeem implementeren. Vervolgens kunnen de miljoenen logevents op een gemakkelijke en efficiënte manier doorzocht worden met behulp van de beschikbare zoekfunctionaliteiten. Het uitvoeren van zoekopdrachten kan via de web interface die de zoekresultaten op een inzichtelijke manier presenteert. Daarnaast biedt de web interface de mogelijkheid tot het visueel weergeven van logdata door middel van grafieken of diagrammen.

# Begrippenlijst

| Begrip | Betekenis |
|---|---|
| **API** | Application Programming Interface |
| **CLI** | Command Line Interface |
| **Destination** | Bestemming van logevents. |
| **GROK** | Parsed ongestructureerde logdata. |
| **Log field** | Onderdeel van een log event. |
| **Logdata** | Verzameling van logevents. |
| **Logevent** | Een enkel log bericht. |
| **Log Collector** | De server dat verantwoordelijk is voor het verzamelen verwerken van logevents. |
| **NTP** | Network Time Protocol |
| **Parsing** | Het segmenteren van logevents. |
| **Prunen** | Het verwijderen van data. |
| **RAM** | Random-access memory |
| **Source** | De bron van logevents. |
| **Threshold** | Een drempelwaarde. |
| **Terminal** | Een tool waarmee een device op afstand bedient kan worden. |
| **TCP** | Tranmission Control Protocol |
| **UDP** | User Datagram Protocol |

# Contents

# 1 Inleiding

Dit document beschrijft het afstudeerproject dat is uitgevoerd bij BIT. De uitgevoerde opdracht betreft een ontwerpopdracht waarbij een functioneel ontwerp, technisch ontwerp en implementatieplan gerealiseerd is. De opdracht is uitgevoerd in een periode van 18 weken.

BIT is een internet service provider gevestigd in Ede. Een bedrijf als BIT beschikt over honderden servers en devices die van belang zijn voor de kwaliteit van de diensten die ze leveren. Al deze servers en devices genereren logdata waarin vermeld staat welke gebeurtenissen op een bepaald moment voorgekomen zijn. De tekenen van problemen zijn vaak in een vroeg stadium te signaleren vanuit de gegenereerde logdata. Daarnaast is logdata een belangrijke informatiebron bij het oplossen van storingen.

In de huidige situatie worden de miljoenen logevents, rechtstreeks vanaf de bron, naar een centrale syslog server getransporteerd. In deze situatie loopt BIT tegen een aantal beperkingen. De grote hoeveelheid logdata zorgen er voor dat het een lastige opgave is de logevents te doorzoeken en zo de juiste informatie te vinden. Daarnaast zijn de zoekopdracht complex en neemt dit veel tijd in beslag.

Het doel van dit project is een log management systeem realiseren dat de logevents centraal opslaat, verwerkt en op een inzichtelijke manier beschikbaar stelt aan de engineers. Hierdoor is het mogelijk zonder complexe en tijdrovende zoekopdrachten de juiste informatie te vinden.

## Leeswijzer

Hoofdstuk 2 beschrijft de context waarin het project is uitgevoerd. Hierin staat een beschrijving van het bedrijf en project gerelateerde afdelingen.

In hoofdstuk 3 bevindt zich een beschrijving van de opdracht, aanleiding, probleemstelling en doelstelling. Verder bevat dit hoofdstuk een centrale hoofdvraag met ondersteunende deelvragen.

De verschillende gebruikte methode en technieken die ondersteuning bieden bij het realiseren van het project staan per fase beschreven in hoofdstuk 4. Daarnaast staat hier beschreven wat er per fase is uitgevoerd.

De resultaten van hoofdstuk 4 staan in hoofdstuk 5 beschreven. Hier komen onder andere het functioneel en technisch ontwerp aan bod.

In hoofdstuk 6 staan de hoofd en deelvragen beantwoord. De antwoorden zijn gebaseerd op de resultaten van het onderzoek.

In hoofdstuk 7 staan de aanbevelingen voor BIT opgenomen. Deze aanbevelingen zijn gebaseerd op de conclusies en worden ondersteund door de resultaten van de gebruikte methode en technieken.

Er zijn verschillende bronnen gebruikt bij de realisatie van dit project. Deze bevinden zich in hoofdstuk 8.

Het laatste hoofdstuk bevatten de bijlages. Dit zijn de producten die opgeleverd zijn zoals het functioneel ontwerp, technisch ontwerp, proof of concept en implementatieplan.

# 2 Organisatie

Dit hoofdstuk beschrijft de organisatie waar het afstudeerproject is uitgevoerd. Verder bevindt zich hier een beschrijving van de project gerelateerde afdelingen.

## 2.1 BIT

BIT is opgericht in 1996 en is een zakelijke internet service provider. Het bedrijf bevindt zich in Ede en bestaat uit ongeveer 35 werknemers. BIT levert onder andere de volgende diensten:

- Colocatie
- Virtuele Servers
- Managed Servers
- Backup Storage
- Monitoring
- Shared Hosting
- Internet verbindingen
- Domeinen en DNS

Binnen de doelgroep van BIT vallen bedrijven met hoge kwaliteitseisen. BIT levert voor deze klanten diensten met daarbij garanties op betrouwbaarheid, continuïteit en performance. BIT voldoet aan deze garanties mede door de professionele medewerkers en de uitgebreide, redundante, infrastructuur.

Het pand van BIT bestaat uit een kantoorgedeelte en drie datacenters namelijk BIT-2A, BIT-2B en BIT-2C. BIT-1 bevindt zich op een aparte locatie. De datacenters van BIT behoren tot de top van Nederland. Alle drie de datacenters zijn voorzien van eigen noodstroom en koelvoorzieningen. Daarnaast besteedt BIT maximale zorg aan de kwaliteit van het netwerk. Het netwerk van de datacenters is redundant uitgevoerd en staat in verbinding met Londen, Amsterdam en Frankfurt. De glasvezelpaden zijn bij verschillende leveranciers ingekocht en zijn over het hele traject fysiek gescheiden[1]. Op Afbeelding 2-1 is schematisch het core netwerk afgebeeld.

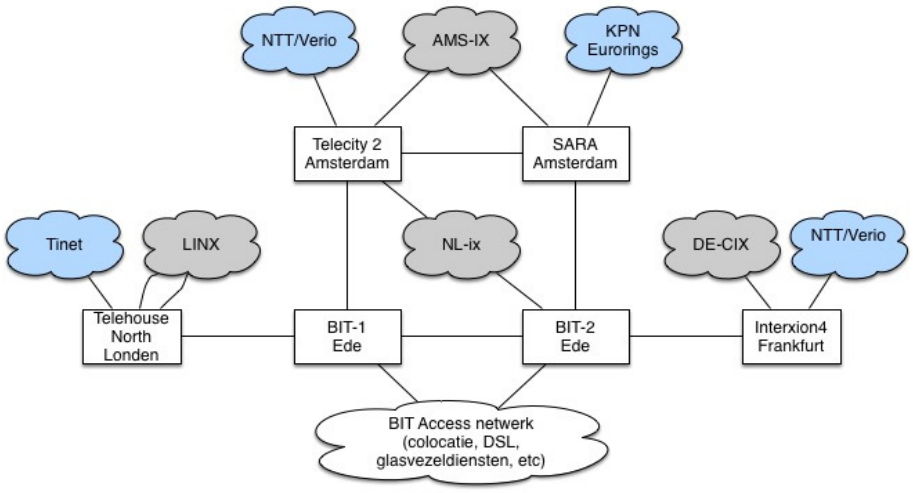


*Figure 2-1Core netwerk BIT*

[1] (Algemeen Datacenters, 2015)

Het zeer grote en veelzijdige netwerk van BIT bestaat ongeveer uit de volgende devices:

- 30 routers, firewalls en DWDM-apparatuur
- 70 switches
- 4 loadbalancers
- 75 servers
- 150 VM's
- 20 andere systemen (UPS'en, gebouwbeheersingssysteem, toegangscontrolesysteem)
- 15 loadbalanced platformen (smtp, mx, imap, maildrop, filter, spamd, http-linux, http-windows, http-BIT, ldap, etc)

Daarnaast beheert BIT tientallen tot honderden servers, firewalls, routers en switches van klanten.

## 2.2 Unix & Windows engineering

De afstudeeropdracht is uitgevoerd op de afdeling Unix & Windows engineering waarvan Sander Smeenk de teamleider is. Deze afdeling bestaat uit een team van vijf mensen en houdt zich bezig met het beheren van de systemen die aangesloten zijn op het netwerk van BIT.

## 2.3 Relatie afdelingen

Niet alleen de medewerkers van Unix & Windows engineering zullen gebruik maken van het log management systeem, maar ook de medewerkers op de afdelingen Network Operations Center (NOC) en Customer Care Sales (CCS).

NOC houdt zich bezig met het beheren van het netwerk van BIT. Dagelijkse werkzaamheden bestaan onder andere uit troubleshooting, vervangen van netwerkcomponenten en het aanpassen van netwerken volgens de wensen van de klanten. De teamleider van NOC is Teun Vink.

CCS is het aanspreekpunt voor klanten, voor service of product gerelateerde vragen. Naast sales activiteiten houd CCS zich bezig met het beantwoorden van vragen en verhelpen van problemen die via een ticketsysteem of telefonisch binnenkomen.

# 3 Opdracht

In dit hoofdstuk bevindt zich een uitgebreide beschrijving van de opdracht. De opdracht omvat het realiseren van een implementatieplan waarmee een log management systeem gerealiseerd kan worden.

## 3.1 Aanleiding

BIT biedt verschillende diensten aan. De klanten die deze diensten afnemen krijgen de garantie dat BIT er alles aan doet om de performance, continuïteit en betrouwbaarheid te waarborgen. De afdelingen CSS, Unix & Windows engineering en NOC houden zich hier dagelijks mee bezig.

Een van de gebruikte tools bij het oplossen of voorkomen van calamiteiten is het raadplegen van logdata. Logdata bevat belangrijke informatie over de gezondheid van devices en services. Daarnaast is het mogelijk dat de oorzaak van een storing is vastgelegd in logdata.

BIT heeft gevraagd onderzoek te doen naar een oplossing die de grote hoeveelheid logdata inzichtelijk en toegankelijk maakt voor de engineers.

## 3.2 Probleemstelling

BIT heeft een groot aantal devices en services die vele logevents genereren. Op dit moment worden de meeste van deze logevents op één server verzameld waar ze geraadpleegd kunnen worden bij complicaties. Bij dit systeem loopt BIT tegen een aantal beperkingen aan:

- De grote hoeveelheid logdata zorgt er voor dat het lastig wordt de juiste informatie snel te vinden.
- Er vindt geen actieve controle van de logevents plaatst waardoor belangrijke informatie gemist wordt.
- De logevents worden niet verwerkt naar grafieken en overzichten.
- Er zijn geen tools beschikbaar waarmee niet-technici toegang hebben tot specifieke logdata.

Deze beperkingen hebben tot gevolg dat de logevents alleen bekeken worden nadat de problemen bekend zijn en er duidelijk is waar deze problemen zich bevinden. Daarnaast kan er veel meer informatie uit logdata gehaald worden die van belang zijn voor het operationeel houden van de devices, servers en services van BIT.

In de huidige situatie worden de logevents van verschillende componenten uit het netwerk van BIT verzameld op een syslog server. Op deze server komen de logevents binnen, maar worden verder niet verwerkt.

## 3.3 Doelstelling

Het doel is om binnen 18 weken een ontwerp en een advies gereed te hebben met daarin opgenomen een implementatieplan dat beschrijft hoe een log management systeem gerealiseerd kan worden. Het log management systeem is een oplossing die logevents (van de vele verschillende devices en applicaties) centraal opslaat, verwerkt en inzichtelijk maakt. Dit zodat er bij calamiteiten snel en doelgericht gezocht kan worden naar de oorzaak van een probleem en zodanig dat er actieve monitoring toegepast kan worden.

Het eindresultaat is een systeem bestaand uit één of meerdere servers waarop de logevents in een database verwerkt worden volgens een vaste structuur. Dit systeem kan vervolgens geraadpleegd worden via de terminal of web interface. Daarnaast kunnen er problemen vroegtijdig gedetecteerd worden doordat de logdata verwerkt kan worden in grafieken en samenvattende overzichten.

## 3.4 Type opdracht

Het type opdracht dat uitgevoerd wordt bij BIT is een ontwerp opdracht. Er is een functioneel en technisch ontwerp gerealiseerd. Aan de hand van deze ontwerpen en het proof of concept, is een implementatieplan geschreven. De resultaten van het onderzoek staan beschreven in hoofdstuk 5.

## 3.5 Centrale onderzoeksvraag

Hoe kunnen de miljoenen logevents die gegeneerd worden door de verschillende devices en services die daarop draaien, centraal opgeslagen en verwerkt worden zodat hier bruikbare, doorzoekbare en inzichtelijke informatie uit voortkomt?

## 3.6 Deelvragen

Onderstaande tabel bevat de deelvragen die ondersteuning bieden op het antwoord op de hoofdvraag.

| Cluster | Deelvraag | Nr. | Subvragen |
|---|---|---|---|
| A | **Welke informatie bevat een log?** | A1 | Wat zijn logs? |
| | | A2 | Uit welke componenten bestaat een log bericht? |
| | | A3 | Hoe worden logs gegenereerd? |
| | | A4 | Wat is de gebruikswaarde van een log? |
| B | **Welke bronnen genereren de input van de logs?** | B1 | Welke devices genereren logs? |
| | | B2 | Welke services genereren logs? |
| | | B3 | Hoeveel logs worden er gegenereerd per 24 uur? |
| C | **Wat zijn de eisen en wensen van BIT?** | C1 | Aan welke eisen en wensen moet het systeem voldoen? |
| | | C2 | Welke functionaliteiten moet het systeem bevatten? |
| D | **Hoe kunnen de logs op een veilige manier centraal opgeslagen worden?** | D1 | Waar moeten de logs opgeslagen worden? |
| | | D2 | Hoe kunnen deze veilig worden getransporteerd over het netwerk? |
| | | D3 | Hoelang moeten deze bewaard worden? |
| | | D4 | Hoe kan dit voldaan worden aan de ISO:27001 norm? |
| E | **Hoe kan de inhoud van de log berichten efficiënt opgeslagen en verwerkt worden in een database?** | E1 | Hoe kunnen de logs volgens een vaste structuur in een database verwerkt worden? |
| | | E2 | Welke tools zijn hiervoor beschikbaar? |
| | | E3 | Hoe kunnen bestaande tools aangepast worden zodat er specifieke logs verwerkt kunnen worden? |
| F | **Op welke manieren kunnen de logs het beste geraadpleegd worden?** | F1 | Hoe kunnen de logs doorzocht worden? |
| | | F2 | Welke tools zijn hiervoor beschikbaar? |
| | | F3 | Hoe kan dit beveiligd worden? |
| | | F4 | Hoe kan dit voldaan worden aan de ISO:27001 norm? |
| G | **Hoe kan het log management systeem beheerd worden?** | G1 | Hoe kunnen de gebruikte producten beheerd worden? |
| | | G2 | Hoe kan het systeem uitgebreid worden? |

# 4 Uitvoering

Dit hoofdstuk beschrijft welke fases doorlopen zijn, wat er in de fase is uitgevoerd en welke methode en technieken zijn toegepast. De resultaten hiervan staan beschreven in hoofdstuk 5.

## 4.1 Initiatiefase

De initiatiefase is de eerste fase van het project. Het doel van deze fase is onder andere bekend worden met de werkwijze van BIT en de afdeling waar het project wordt uitgevoerd.

Verder is in deze fase een Plan van Aanpak gerealiseerd. Hierin staat de opdracht uitgebreid beschreven met daarbij een beschrijving van het bedrijf, de hoofdvraag en deelvragen, op te leveren producten en een planning. Aan de hand van het Plan van Aanpak is vanuit school toestemming gegeven voor het uitvoeren van het project.

## 4.2 Onderzoekfase

De onderzoekfase dient als voorbereiding op het realiseren van het functioneel ontwerp, technisch ontwerp en implementatieplan. De volgende drie onderwerpen: logs, inventarisatie en eisen en wensen staan hier uitgewerkt.

### 4.2.1 Logs

Het is van belang om kennis op te doen over wat logevents precies zijn, hoe ze ontstaan, hoe een log event is opgebouwd en wat de bron van logdata is. Het onderzoek hiervan heeft grotendeels plaatsgevonden op het internet door gebruik van specifieke zoektermen. Er is bijvoorbeeld gezocht op logdata, log management, syslog en linux log. Daarnaast is het boek Logging and Log Management[2] gebruikt als bron van informatie. De opgedane kennis heeft bijgedragen bij het uitbreiden van de lijst met eisen en wensen. Zo is bijvoorbeeld een eis toegevoegd wat betreft de beveilig van het transporteren van logevents.

### 4.2.2 Inventarisatie

BIT beschikt over een uitgebreid netwerk waar vele verschillende devices onderdeel van zijn. De devices die logevents genereren zijn grotendeels geïnventariseerd door de syslog server te onderzoeken. Wanneer een device wordt toegevoegd aan het netwerk van BIT wordt dit zodanig opgeleverd dat device logevents transporteert naar een centrale syslog server. Door na te gaan welke devices logevents transporteren en hoeveel logdata er per 24 uur gegeneerd wordt is een inventarisatielijst gerealiseerd. Het doel hiervan is inzicht krijgen over de hoeveelheid devices en logevents. Dit is noodzakelijk voor het implementeren van een log management systeem omdat deze de stroom aan logevents moet aankunnen en het duidelijk moet zijn waar logevents vandaan komen. Verder heeft de hoeveelheid logevents per 24 uur een grote invloed op de systeemeisen.

[2] (C & K, 2015)

### 4.2.3 Eisen en wensen

De opdrachtgever, Sander Smeenk, heeft aan het begin van het project een uitgebreide lijst met eisen en wensen opgesteld. Deze is vervolgens onderzocht en uitgebreid naar eigen inzicht met functionaliteiten die van belang kunnen zijn voor BIT. Deze toegevoegde functionaliteiten zijn gebaseerd op de resultaten van het vooronderzoek. Een voorbeeld van een toegevoegde eis is het verzenden van logevents over een beveiligde TLS verbinding. Daarnaast zijn de eisen en wensen die aangeleverd zijn specifieker opgesteld. Aan de hand van de lijst is het functioneel ontwerp gerealiseerd en zijn de criteria opgesteld die toegepast zijn in de productselectie.

Op de lijst met eisen en wensen is de MoSCoW methode toegepast. Dit is een methode die elke eis een prioriteit geeft: **M**ust have, **S**hould have, **C**ould have of **W**ont have. In eerste instantie hebben de originele eisen en wensen, ontvangen van de opdrachtgever, een must have gekregen. De meeste toegevoegde eisen en wensen hebben een should of could have gekregen. De prioriteiten hebben invloed op de productselectie. De criteria in de productselectie die een must have betreffen tellen zwaarder mee dan de overige criteria. De aangewezen prioriteiten zijn besproken met de opdrachtgevers.

## 4.3 Ontwerpfase

In de ontwerpfase zijn twee ontwerpen gerealiseerd: een functioneel en technisch ontwerp. Daarnaast is in deze fase de productselectie uitgevoerd. Uiteindelijk zijn de ontwerpen en het gekozen product getest aan de hand van een proof of concept.

### 4.3.1 Ontwerpen

De twee gerealiseerde ontwerpen beschrijven het log management systeem. In het functioneel ontwerp staan onder andere de huidige situatie, gewenste situatie en functionaliteiten die het systeem moet bevatten beschreven. Aan de hand van dit functioneel ontwerp en de eisen en wensen, is het technisch ontwerp gerealiseerd. Hierin bevindt zich een lijst met technische eisen en wensen, een technische beschrijving van de omgeving en een selectieprocedure.

De architectuur van de omgeving bestaande uit de log collector, database server en interface server is gebaseerd op de aanbevelingen uit het boek Logging and Log Management hoofdstuk 20 Planning your own log analysis system.

Naast de eisen en wensen heeft de opgedane kennis over logs en de inventarisatie bijgedragen aan het realiseren van deze twee ontwerpen. Zo bevat het technisch ontwerp bijvoorbeeld een hoofdstuk waarin de systeemeisen staan gespecificeerd. Deze specificaties zijn gericht op de resultaten van de inventarisatie.

### 4.3.2 Selectieprocedure

Er zijn meerdere producten beschikbaar die ingezet kunnen worden als log management systeem. In de selectieprocedure is bepaald welk product het beste voldoet aan de eisen en wensen van BIT. De procedure bestaat uit het opzetten van een longlist met producten waarvan er enkele overblijven en in de shortlist worden opgenomen. Door middel van de decision matrix[3] is bepaald welke producten van de longlist in de shortlist worden opgenomen en welk product uit de shortlist het beste aansluit op de eisen en wensen van BIT.

De producten waaruit de longlist bestaat zijn gekozen door na te gaan welke log management producten er zoal bestaan. Dit onderzoek heeft plaatsgevonden op het internet. Er zijn verschillende applicaties gevonden aan de hand van de volgende zoektermen: log management tool, log application en centralized log tools. Vervolgens zijn vier producten gekozen waarvan voldoende documentatie beschikbaar is, waar een grote community achter staat en wat in grote lijnen de functionaliteiten bied wat aansluit op de eisen en wensen. Daarnaast is gekeken of het product recentelijk een update ondergaan heeft dat aantoont dat er nog developers aan werken.

De decision matrix is een methode waarbij de producten op gegeven criteria een score krijgen. De criteria van de longlist zijn gebaseerd op de eisen en wensen die de must have prioriteit bevatten. De criteria van de shortlist zijn de overige eisen en wensen uit het functioneel ontwerp en de technische eisen en wensen uit het technisch ontwerp. Door de documentatie van de producten te bestuderen is het mogelijk de conclusie te trekken of een product voldoet aan de criteria.

### 4.3.3 Kwaliteitsbewaking

Over het hele traject zijn een aantal momenten opgenomen om de kwaliteit te bewaken. Zo is elk opgeleverd product goedgekeurd door de opdrachtgever. Daarnaast is een proof of concept uitgevoerd en getest aan de hand van een testplan.

De test opstelling bestaat uit dezelfde componenten als de uiteindelijke implementatie. De drie componenten zijn op een test omgeving geïnstalleerd. Het installeren en configureren van de producten is aan de hand van de originele documentatie gerealiseerd[4][5].

Het doel van het proof of concept is bewijzen dat het ontwerp en het product voldoet aan de eisen en wensen van de opdrachtgever. Daarnaast dient het proof of concept als een test waar gekeken is of het product daadwerkelijk in staat is de beschreven functionaliteiten uit de ontwerpen kan uitvoeren. In samenwerking met de opdrachtgever is het testplan doorlopen. De te testen onderdelen zijn opgesteld aan de hand van de lijst met eisen en wensen.

[3] (What is a decision matrix?, 2015)
[4] (Elastic - Docs, 2015)
[5] (Graylog - Docs, 2015)

## 4.4 Implementatiefase

In de implementatiefase is een implementatieplan gerealiseerd. Het doel van dit document is stap voor stap beschrijven hoe het log management systeem geïnstalleerd en geconfigureerd kan worden. Het implementatieplan is gebaseerd op de twee ontwerpdocumenten en de opgedane kennis bij het uitvoeren van de proof of concept.

Het implementatieplan bestaat uit een gedeelte waarin de systemen worden voorbereid op de installatie van de producten en uit een gedeelte waarin de producten daadwerkelijk geïnstalleerd en geconfigureerd worden. Daarnaast bevat het document een compacte beschrijving van de architectuur en systeemeisen.

Het implementatieplan is geschreven voor de engineers en er is hier vanuit gegaan dat degene die het plan doorloopt enige kennis heeft wat betreft Linux en security. Dit omdat degene die het systeem zal implementeren een engineer van BIT is.

## 4.5 Afronden

In de laatste fase is het project afgesloten met de realisatie van een gebruikershandboek en scriptie. Het gebruikershandboek bevat een uitgebreide beschrijving van het log management systeem dat gericht is op het hanteren en beheren van de omgeving. Een gebruiker kan met behulp van het handboek de web interface in gebruik nemen, updates uitvoeren, het systeem uitbreiden en configuraties aanpassen. Het gebruikershandboek is verwerkt op de interne WIKI omgeving van BIT. De interne WIKI omgeving is het documentatiesysteem van BIT.

Naast het gebruikershandboek is in deze fase de scriptie gerealiseerd. De scriptie beschrijft het project in al zijn facetten.

# 5 Resultaten

In dit hoofdstuk staan de resultaten van de uitvoering. Dit omvat de resultaten van het vooronderzoek, gebruikte methode en technieken en de gerealiseerde producten.

## 5.1 Vooronderzoek

In dit deelhoofdstuk staan de resultaten van het uitgevoerde vooronderzoek.

### 5.1.1 Logs

Logdata wordt gegeneerd door verschillende devices en services en bevat de gebeurtenissen die ze ondergaan zijn. Omdat elke gebeurtenis is vastgelegd bevat logdata informatie die het oplossen van calamiteiten kan bevorderen. Ook toekomstige calamiteiten kunnen voorkomen worden door de logdata te analyseren.

Besturingssysteem, applicaties of services slaan logevents op in logbestanden. Elk onderdeel van een systeem dat log berichten opslaat doet dit in zijn eigen logbestand. Vaak bevat een systeem daarom verschillende logbestanden met elk hun eigen type informatie[6].

De log regels die weggeschreven worden naar een logbestand worden geclassificeerd onder de volgende categorieën[7]:

0) **Emergency**: Het betreffende systeem, applicatie of services werkt niet meer.
1) **Alert**: Er moet direct actie ondernomen worden. Er vindt een storing of calamiteit plaats.
2) **Critical**: Een bericht dat geclassificeerd is als critical moet direct onderzocht worden. Critical geeft aan dat er op korte termijn een storing of calamiteit kan ontstaan.
3) **Error**: Een error geeft aan dat er iets niet goed gegaan is. Hier moet naar gekeken worden binnen een bepaalde tijd.
4) **Warning**: Waarschuwingen geven aan dat er elk moment iets fout kan gaan. Bijvoorbeeld een harde schijf die bijna vol is kan een waarschuwing afgeven. Als hier niets mee gedaan wordt gaat dit grotere problemen opleveren.
5) **Notice**: Notice berichten geven aan dat er mogelijk een fout ontstaat. Hier hoeft niet direct naar gekeken te worden, maar moet wel een keer uitgezocht worden.
6) **Informational**: Een bericht met informatie over een systeem. Bijvoorbeeld een bericht met informatie over het aantal gebruikers dat actief is op dat moment.
7) **Debug**: Nuttige informatie over de werking van de applicatie. Deze berichten worden voornamelijk gebruikt door developers om fouten op te sporen in een applicatie.

[6] (Linux Filesystem Hierarchy, 2015)
[7] (The Ins and Outs of System Logging Using Syslog, 2015)

De meeste Unix devices die logging toepassen doen dit volgens het syslog protocol dat beschreven staat in RFC5424[8]. De processen of applicaties die op een systeem draaien transporteren de logevents, via het syslog protocol, door naar een syslog daemon. De syslog daemon is een proces dat de logevents filtert en opslaat op een gegeven locatie. In figuur 5-1 is dit systeem schematisch weergegeven. Hierop is te zien dat de syslog daemon de logevents volgens de configuratiesettings in syslog.conf filtert.

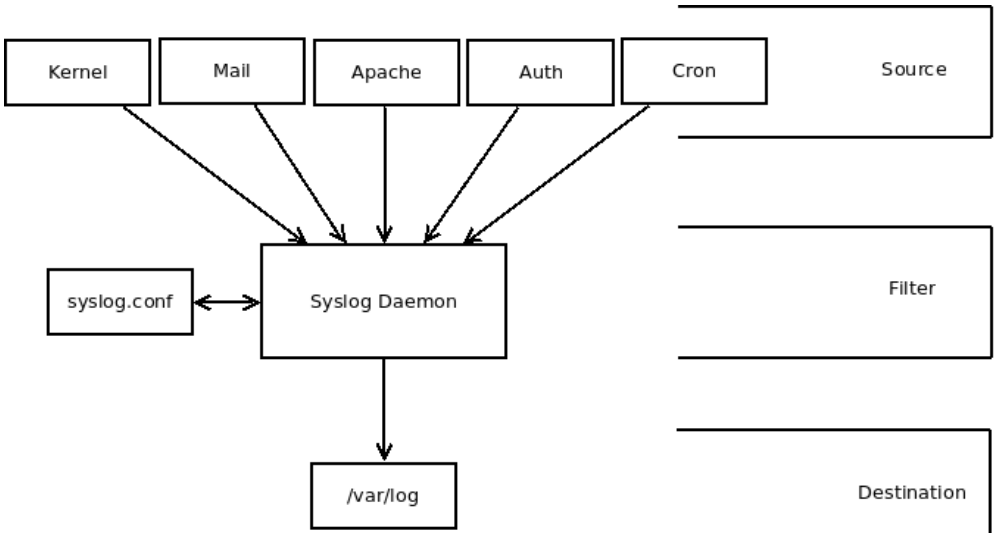


*Figure 5-1Logging proces overview*

Op Windows systemen worden de logevents gefilterd en opgeslagen door middel van de Windows Event Log. Er zijn verschillende applicaties beschikbaar die Windows Event Logs converteert naar syslog berichten volgens het syslog protocol[9].

Een typisch log bericht bevat de volgende drie onderdelen[10]:

- Timestamp
- Source
- Message

[8] (RFC 5424 - The Syslog Protocol, 2015)
[9] (Configuring a Syslog Agent in Windows Server 2012, 2015)
[10] (C & K, 2015)

### 5.1.2 Inventarisatie

De inventarisatie is gerealiseerd door te onderzoeken welke devices log berichten naar de syslog server transporteren. De lijst met devices is vervolgens onderverdeeld in categorieën. De lijst toont aan dat er ongeveer 325 devices in het netwerk van BIT logevents transporteren naar de syslog server.

De log rotation is zo geconfigureerd dat elke dag om 07:00 de logbestanden van de 24u daarvoor gearchiveerd worden in gecomprimeerde bestanden. Hierdoor is het mogelijk te achterhalen hoeveel log berichten er per 24u gegeneerd worden. Het aantal log berichten is berekend door van twee dagen het gemiddelde te nemen.

```
[root@syslog:~] # ls /**/info.log.2.gz | xargs zcat | wc -l
```

Bovenstaande commando telt het aantal regels per gecomprimeerd bestand. De uitkomst van dit onderzoek is een moment opname en kan per dag verschillen. Ook is het mogelijk dat er zicht devices in het netwerk bevinden die nog niet compleet geconfigureerd zijn waardoor deze nog geen logevents transporteren naar de syslog server.

| Categorie | Aantal devices | Logevents per 24u |
|---|---|---|
| **Back-up** | 5 | 26.988 |
| **Cloud** | 12 | 14.712.722 |
| **Database** | 7 | 37.420 |
| **DNS** | 8 | 707.021 |
| **Firewall** | 7 | 263.512 |
| **Web server** | 32 | 12.954.890 |
| **LDAP** | 6 | 448.000 |
| **Loadbalancers** | 3 | 361.474 |
| **Mail** | 39 | 7.815.976 |
| **Monitoring** | 4 | 91.953 |
| **Storage** | 5 | 5.612.570 |
| **Overig** | 73 | 211.633 |
| **Routers** | 11 | 1.571.716 |
| **Spam** | 10 | 5.989 |
| **SQL** | 4 | 14.867 |
| **Switches** | 74 | 14.852 |
| **Wifi** | 19 | 38.248 |
| **Totaal** | **325** | **44.889.831** |

### 5.1.3 Eisen en wensen

De lijst met eisen en wensen geeft aan wat het log management systeem aan functionaliteiten moet bevatten. De opdrachtgever heeft de lijst grotendeels aangeleverd.

| Onderwerp | Nr. | Moscow* | Eisen en wensen |
|---|---|---|---|
| **Zoeken** | 1 | M | Snel en effectief zoeken door logdata. |
| | 2 | M | Zoeken in specifieke tijdsperiodes. |
| | 3 | M | Zoeken op een specifiek type van logevents (info/warning/error/critical/etc.). |
| | 4 | M | Resultaten filteren op verschillende invalshoeken (device/platform/applicatie/etc.). |
| | 5 | M | Correlatie van meldingen van meerdere systemen. |
| | 6 | S | Breakdown per type melding. |
| **Statistieken** | 7 | S | Genereren van diagrammen en grafieken. |
| | 8 | S | Overzichten per melding (jaar/maand/dag). |
| **Toegang** | 9 | M | Via een web interface toegang tot logevents en grafische overzichten. |
| | 10 | C | Klanten via web interface toegang tot eigen logs geven. |
| | 11 | M | Via een CLI toegang tot logevents hebben. |
| | 12 | S | Via een web interface live events ontvangen volgens opgegeven filters. |
| **Security** | 13 | M | Toegangscontrole met username en password. |
| | 14 | M | Inzicht tot logdata op basis van functie medewerker. |
| | 15 | M | Voorkomen dat logevents bewerkt of verwijderd worden. |
| | 16 | M | Gebruik maken van encryptie. |
| | 17 | S | LDAP integratie. |
| **Overig** | 18 | S | Actief monitoren op afwijkingen bijv. door grote toename foutmeldingen. |
| | 19 | C | Meldingen afgeven bij overschrijden threshold. |
| | 20 | C | Integratie met het monitoring systeem van BIT. |
| | 21 | M | Systeem moet schaalbaar zijn. |
| | 22 | S | Verwijderen van logdata aan de hand van bepaalde classificaties. |
| | 23 | M | Zoveel mogelijk gebruik maken van open-source producten. |

* **M**ust have, **S**hould have, **C**ould have, **W**ont have

## 5.2 Functioneel ontwerp

Het functioneel ontwerp beschrijft de huidige situatie, gewenste situatie en de functionaliteiten van het log management systeem.

Uit de inventarisatie is gebleken dat het netwerk van BIT een grote hoeveelheid verschillende devices bevat die ieder logevents genereren. De logevents worden getransporteerd naar een syslog server waar ze vervolgens geraadpleegd kunnen worden. Het raadplegen op de huidige manier neemt veel tijd in beslag en kan complex worden omdat er weinig structuur in zit. Daarnaast zorgt een grote stroom van logevents er voor dat bruikbare informatie over het hoofd wordt gezien.

De opdrachtgever wenst de realisatie van een log management systeem dat de grote stroom aan logdata kan centraliseren, verwerken en inzichtelijk kan maken. Door middel van dit systeem moet het mogelijk zijn de logevents snel en effectief te doorzoeken.

Figuur 5-2 toont de drie componenten waaruit het log management systeem bestaat.

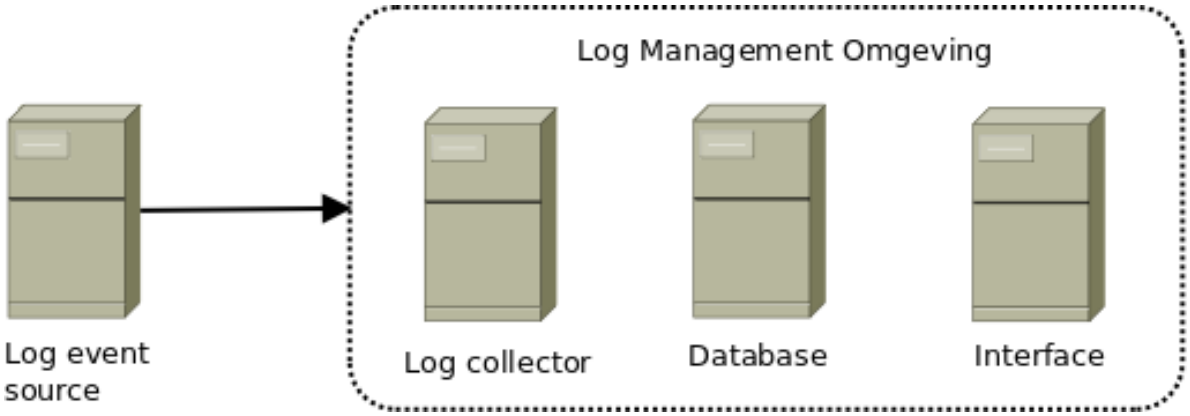


*Figure 5-2Log Management Omgeving*

'Log event source' is de bron van logevents wat in eerste instantie de syslog server zal zijn. De logevents worden vervolgens getransporteerd naar de log collector. Hier worden de logevents verwerkt en verzonden naar de database server. De database server indexeert de logevents in een database waarna deze geraadpleegd kunnen worden. De keuze voor drie componenten is gemaakt omdat het systeem op deze manier schaalbaar is en de workload verdeeld wordt.

Het doorzoeken van logevents wordt mogelijk gemaakt door de verschillende zoekfunctionaliteiten op de web interface. Hier kunnen de logevents snel en efficiënt doorzocht worden. De volledige lijst met zoekfunctionaliteiten staat beschreven in bijlage 2 hoofdstuk 4.2.1 Zoekfuncties.

Omdat de logevents, afkomstig van verschillende bronnen, centraal worden opgeslagen is het mogelijk van meerdere systemen logevents te correleren door gebruik te maken van de beschikbare zoekfunctionaliteiten of de web interface.

Naast zoekfunctionaliteiten is het via de web interface mogelijk verschillende type diagrammen en grafieken te genereren zoals lijngrafieken en cirkeldiagrammen. Hierdoor wordt de kans tot het vroegtijdig signaleren van calamiteiten vergroot. Een grafiek kan aantonen hoeveel logevents, die voldoen aan bepaalde criteria, er op een bepaald moment gegeneerd zijn. Een grote toename kan wijzen op een mogelijk opkomend probleem. Er kunnen ook alarmen afgegeven worden bij een grote toename van logevents of wanneer een bepaalde threshold wordt overschreden.

Niet alleen via de web interface, maar ook via een API is het mogelijk logevents te raadplegen. De API is beschikbaar via de CLI of door middel van scripting. De API ondersteunt meerdere scripting talen zoals Perl, Python of Bash.

Het log management systeem zal gevoelige informatie bevatten. Het is daarom van belang rekening te houden met verschillende security aspecten. Het is bijvoorbeeld niet de bedoeling dat gebruikers van het systeem opzettelijk of per ongeluk data wijzigen of verwijderen. Daarnaast dient de informatie niet publiekelijk toegankelijk te zijn. Door het toepassen van authenticatie is het mogelijk informatie toegankelijk te maken voor bepaalde gebruikers. Autorisatie maakt het mogelijk bepaalde informatie voor bepaalde gebruikers ontoegankelijk te maken door middel van permissions. Naast autorisatie en authenticatie zullen de communicatie lijnen, waar mogelijk, beveiligd worden door middel van encryptie zodat er geen gevoelige informatie onderschept kan worden.

Het is mogelijk dat de stroom aan logdata in de toekomst zal toenemen. Hier is rekening mee gehouden door het systeem schaalbaar te maken. Het systeem zal op een virtuele omgeving geïnstalleerd worden waardoor het mogelijk wordt resources uit te breiden. Verder is er bij het realiseren van de ontwerpen rekening gehouden met horizontale uitbreiding. De drie componenten zullen elk op een eigen virtuele server geïnstalleerd worden. Het is mogelijk om van de componenten een tweede node toe toevoegen wanneer de stroom aan logdata te groot wordt. Zo wordt de workload verdeeld over twee of meerdere nodes en verhoogt dit de beschikbaarheid van de logdata.

## 5.3 Technisch ontwerp

Het technisch ontwerp beschrijft de technische aspecten van de log management omgeving en bevat een selectieprocedure waarin vier producten met elkaar vergeleken zijn. Het product dat het beste aansluit op de eisen en wensen van BIT zal uiteindelijk geïmplementeerd worden. De omgeving bestaat uit drie componenten.

**Log collector**

Op de syslog server is syslog-ng geïnstalleerd als syslog daemon. Door middel van syslog-ng worden de logevents vanaf de syslog server getransporteerd naar de log collector. De communicatie tussen de syslog server en log collector verloopt over TCP op poort 514 aan beide kanten. Hier wordt gebruik gemaakt van TCP omdat dit betrouwbaarder is dan UDP[11].

Naast de syslog server is het mogelijk andere bronnen te configureren zodat deze log events transporteren naar de log collector. Hoewel het aanbevolen is dit over TCP te transporteren kunnen meerdere protocolen gebruikt worden voor transport. Welke protocollen dit zijn is afhankelijk van het gebruikte product.

De log collector zal duizenden logevents per minuut ontvangen, parsen en doorsturen naar de database server. Dit vraagt voornamelijk CPU kracht. Het parsen van logevents is een proces dat een log bericht segmenteert in meerdere onderdelen. Doordat er gebruik wordt gemaakt van het syslog protocol (RFC5424) wordt er standaard gesegmenteerd op timestamp, source, destination, severity level en message[12]. Wanneer het wenselijk is logevents verder te segmenteren is dit mogelijk door custom parsing rules te creëren volgens GROK[13].

Een logevent is in eerste instantie een ononderbroken stuk tekst. Door middel van parsing is het mogelijk structuur aan te brengen door bepaalde onderdelen in fields te segmenteren. Een log event bevat bijvoorbeeld een timestamp dat in een field geparsed kan worden. Dit maakt het filteren op bepaalde tijdsperiodes mogelijk. In het handboek staat beschreven hoe de parsing language GROK toegepast kan worden. GROK geeft logdata structuur door data in fields te parsen.

**Database**

Nadat de logevents geparsed zijn door de log collector komen ze binnen op de database server. Hier worden de logevents geïndexeerd in de database waarna ze geraadpleegd kunnen worden via de web interface of API. De database server zal te maken krijgen met vele indexeer en query operaties. Dit vraagt voornamelijk geheugen en I/O snelheid van de harde schijf.

[11] (TCP vs. UDP, 2015)
[12] (RFC 5424 - The Syslog Protocol, 2015)
[13] (GROK, 2015)

**Interface**

De web interface biedt verschillende zoekfunctionaliteiten en is in staat de logdata visueel te presenteren. De communicatie met de web interface verloopt over een beveiligde HTTPS verbinding.

De systeemeisen van het log management systeem zijn afhankelijk van de gebruikte producten en de hoeveelheid logevents die verwerkt moet worden. Daarnaast hebben de lengte van een log event, in hoeverre deze geparsed wordt en de hoeveelheid logevents per seconde invloed op de systeemeisen. Omdat dit niet vooraf te bepalen is, zijn de opgegeven systeemeisen een richtlijn en betreft geen specifieke minimale eisen.

| Hardware | Log collector | Database server | Interface server |
|---|---|---|---|
| **vCPU** | **4x2GHz** | **2x2GHz** | **2x2GHz** |
| **Memory** | **6GB** | **8GB** | **6GB** |
| **Harddisk** | **80GB** | **300GB** | **80GB** |

Het beveiligen van het systeem is op verschillende manieren mogelijk. Het is aangeraden het systeem in een gecontroleerde omgeving te implementeren zodat er van buitenaf geen toegang mogelijk is, tenzij dit de bedoeling is voor bijvoorbeeld de web interface. De communicatie tussen de drie componenten verloopt over verschillende poorten. De firewall op iedere server server dient zo geconfigureerd te worden dat alleen de nodige server toegang krijgt tot een poort. Een voorbeeld hiervan is de communicatielijn tussen de log collector en database server waarbij de juiste source IP aan beide kanten geconfigureerd moet worden. Wanneer een ander device probeert te communiceren op de betreffende poort wordt dit geblokkeerd door de firewall.

Tijdens het ontwerpen van het log management systeem is rekening gehouden met de ISO:27001 normering[14]. Hier vallen de volgende onderwerpen onder:

- Log retentie
  De log retentie bepaald hoelang logdata bewaard wordt. In de ISO:27001 normering is geen vaste tijd opgenomen. De log retentie zal in eerste instantie zo geconfigureerd worden dat na 30 dagen logevents verwijderd worden. Wanneer het wenselijk is de logdata langer te bewaren zullen de systeemeisen aangepast moeten worden.
- Toegangscontrole
  De omgeving zal gevoelige informatie bevatten dat niet voor iedereen toegankelijk dient te zijn. De web interface zal beveiligd worden door middel van authenticatie. Daarnaast zal autorisatie toegepast worden zodat bepaalde logdata afgeschermd kan worden voor bepaalde gebruikers. De API zal alleen beschikbaar zijn wanneer een SSH verbinding naar de database server is opgezet.

[14] (ISO/IEC 27001:2013, 2015)

- Kloksynchronisatie
  De logevents bevatten een timestamp waarin is aangegeven op welk tijdstip een logevent is gegeneerd. De ISO:27001 normering geeft aan dat de relevantie systemen een accurate tijdsaanduiding moeten hebben. Door gebruik van Network Time Protocol (NTP) op de drie systemen lopen de tijden synchroon. Dit is van belang zodat de logevents juiste timestamps bevatten.

- Integriteit
  Integriteit betekent dat logdata niet gewijzigd of verwijderd mag worden. Het log management systeem kan beveiligd worden zodat dit niet mogelijk is. Via de web interface kan dit door middel van gebruiker permissies toegepast worden.

### 5.3.1 Productselectie

De productselectie bepaalt welk product het beste voldoet aan de eisen en wensen van BIT. De selectie bestaat uit een longlist en een shortlist. De long list bevat vier producten: Graylog[15], Elastic[16], Splunk[17] en Loggly[18].

De producten zijn met elkaar vergeleken door middel van een decision matrix. De decision matrix bevat een lijst met criteria waar op elk criteria een score is toegewezen. De twee producten met de hoogste score zijn nader onderzocht en verder met elkaar vergeleken in de shortlist.

Tabel 5-1 toont van elk product uit de longlist de totale score. Hieruit is gebleken dat Graylog en Elastic het hoogste scoren en daarom in de shortlist terecht komen. De scores liggen bij elkaar in de buurt omdat de vier producten ieder ongeveer dezelfde functionaliteiten bieden. Het grote verschil is dat Graylog en Elastic beiden open-source en gratis producten zijn.

| | Graylog | Elastic | Splunk | Loggly |
|---|---|---|---|---|
| **Total** | **111** | **105** | **102** | **102** |

*Table 5-1Longlist Decision matrix*

De producten Graylog en Elastic zijn nader onderzocht en zijn met elkaar vergeleken. In bijlage 3 hoofdstuk 8 Producteigenschappen staan de producten beschreven. Tabel 5-2 toont de score van de shortlist. Er is twee keer een decision matrix ingevuld omdat het anders een te lange en onoverzichtelijke lijst zou worden. De criteria in de eerste decision matrix zijn opgesteld aan de hand van de overgebleven eisen en wensen uit het functioneel ontwerp. De tweede decision matrix bevat criteria opgesteld aan de hand van de technische eisen en wensen. Hieruit is gebleken dat Graylog het beste past bij de functionele en technische eisen en wensen van BIT.

[15] (Open Source Log Management with Graylog, 2015)
[16] (Elastic · Revealing Insights from Data, 2015)
[17] (Operational Intelligence, Log Management, Application Management, Enterprise Security and Compliance, 2015)
[18] (Log Management | Cloud Log Management Service | Loggly, 2015)

| | ALTERNATIVES | |
|---|---|---|
| | **Graylog** | **Elastic** |
| **Decision matrix 1** | **99** | **81** |
| **Decision matrix 2** | **47** | **39** |
| **Total** | **146** | **120** |

*Table 5-2Shortlist Decision matrix*

De volledige tabellen met toegewezen scores per criteria bevinden zich in bijlage 3 Technisch ontwerp hoofdstuk 4.3 Selectie.

Graylog is een open-source log management platform dat uit drie producten bestaan: graylog-server, Elasticsearch en graylog-web-interface. Eerst en laatst genoemde producten worden onderhouden door Graylog zelf. Elasticsearch is een open-source database applicatie die onderhouden wordt door Elastic.

Graylog-server is de log collector dat de logevents ontvangt, parsed en transporteert naar de database server waarop Elasticsearch is geïnstalleerd. Graylog-web-interface zal geïnstalleerd worden op de interface server en bied een uitgebreide en veilige web interface.

### 5.3.2 Architectuur

Figuur 5-3 toont de architectuur van Graylog. De logevents komen vanaf de bron binnen op de graylog-server. Vervolgens worden de logevents getransporteerd naar de database server waarna ze geraadpleegd kunnen worden. Graylog-server slaat de configuratie van de web-interface op in MongoDB dat op dezelfde server draait. De web-interface communiceert met de graylog-server en niet direct met de database. Wanneer gebruik wordt gemaakt van de API wordt er direct met de database gecommuniceerd. Een uitgebreide beschrijving van Graylog bevindt zich in bijlage 3 hoofdstuk 8 Producteigenschappen.

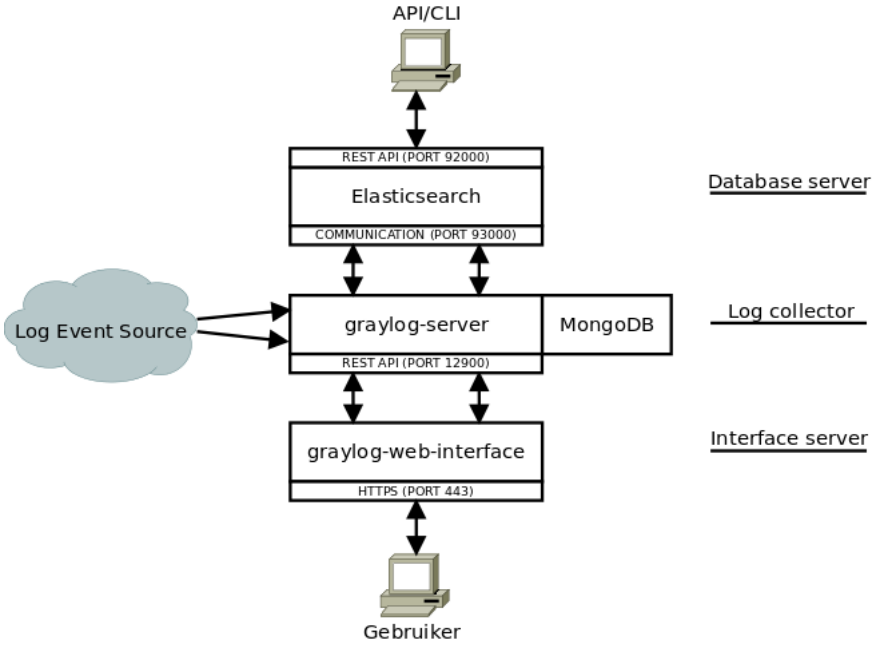


*Figure 5-3Architectuur Graylog*

Graylog biedt mogelijkheid tot beveiliging van de communicatielijnen. De volgende verbindingen zullen beveiligd worden door middel van encryptie:

- Log source ↔ graylog-server (TCP/TLS)
- graylog-server ↔ graylog-web-interface (HTTPS)
- Gebruikers ↔ graylog-web-interface (HTTPS)

## 5.4 Proof of concept

Het proof of concept toont aan dat het product en ontwerp voldoet aan de eisen en wensen van de opdrachtgever. Daarnaast is het een test om te achterhalen of de beschreven functionaliteiten in de documentatie van Graylog ook daadwerkelijk toepasbaar zijn. Verder is het proof of concept een handige tool om na te gaan of de systeemeisen voldoende zijn.

De test opstelling is minder uitgebreid geconfigureerd dan de uiteindelijke implementatie zal zijn. Zo is er bijvoorbeeld geen encryptie toegepast tussen de bron van logevents en de log collector.

De test opstelling bestaat in principe uit dezelfde opstelling als de implementatie zal zijn. De opstelling bevat drie servers met daarop de Graylog log management producten. De log events worden vanaf de syslog server getransporteerd naar de log collector over TCP, maar zonder encryptie. De testopstelling verwerkt ongeveer 22 miljoen logevents.

Samen met de opdrachtgever is een testplan doorlopen. Het testplan bevat verschillende te testen onderdelen die zijn opgezet aan de hand van de lijst met eisen en wensen. Het testplan is over het algemeen succesvol doorlopen. Er zijn echter drie onderdelen die niet getest konden worden, omdat dit bij het realiseren van test omgeving geconfigureerd had moeten worden. Deze onderdelen zijn nader onderzocht en daaruit is gebleken dat de onderdelen wel degelijk mogelijk zijn te implementeren.

## 5.5 Implementatieplan

In het implementatieplan staat uitgebreid, stap voor stap, beschreven hoe het systeem geïmplementeerd kan worden. Aan de hand van dit plan kan iedere engineer het systeem realiseren. Naast het configureren bevat het plan een compacte beschrijving van de architectuur en systeemeisen.

Het implementeren van het log management systeem bevat het configureren van drie systemen namelijk de log collector, database server en interface server. De eerste stap is de drie systemen configureren zodat er vervolgens Graylog op geïnstalleerd kan worden. Graylog en Elasticsearch zijn afhankelijk van een aantal dependencies die eerst geïnstalleerd en geconfigureerd moeten worden. Daarnaast dienen de volgende services geconfigureerd te worden: NTP, Firewalling en MongoDB. Nadat de systemen geconfigureerd zijn kunnen de producten geïnstalleerd worden. Het laatste hoofdstuk bevat een lijst met bruikbare informatieve bronnen over Graylog en Elasticsearch.

## 5.6 Gebruikershandboek

In dit handboek staat beschreven hoe de gebruikers het systeem kunnen hanteren en beheren. Een groot onderdeel van dit handboek is een beschrijving van de web interface waar iedere pagina staat beschreven. Daarnaast behandelt het handboek de zoekfunctionaliteiten, GROK, updaten en het uitbreiden van de omgeving.

Naast de web interface staat beschreven hoe logdata via de API geraadpleegd kan worden. De API is beschikbaar via de command line interface of door middel van scripting. Er is een python script in opgenomen dat laat zien hoe een query uitgevoerd kan worden en hoe de output naar eigen wens opgezet kan worden.

Zowel graylog-server als Elasticsearch kunnen horizontaal geschaald worden. Wanneer dit wenselijk is moeten er een aantal aanpassingen aan de architectuur van de omgeving uitgevoerd worden. In het gebruikershandboek staat beschreven hoe beide producten, Graylog en Elasticsearch, uitgebreid kunnen worden met een tweede node. Dit is overigens geen stappenplan zoals het implementatieplan, maar een globale omschrijving van het proces.

Naast het schalen van de omgeving is het updateproces beschreven. Beide producten ondergaan regelmatig updates die de functionaliteiten uitbreiden. Elasticsearch is door middel van een repository geïnstalleerd en is dus eenvoudig te updaten via upgrade command. De graylog producten moeten echter op een andere manier geüpdatet worden. Beide producten dienen opnieuw gedownload te worden en de configuratie bestanden moeten worden overgezet. Bij een update staat op de website van het product vermeld of een product zomaar geüpdatet kan worden, of dat er speciale handelingen uitgevoerd moeten worden.

# 6 Conclusies

In dit hoofdstuk staan de deelvragen en hoofdvraag beantwoord.

## 6.1 Deelvragen

De hoofdvraag is onderverdeeld in deelvragen die in dit hoofdstuk beantwoord worden.

**Welke informatie bevat een log?**

Wanneer gesproken wordt over logs doelt men op de logdata die gegeneerd wordt door verschillende applicaties of services. Logdata wordt verwerkt in logbestanden die de gebeurtenissen van applicaties of services bevatten. De logevents kunnen lokaal in een bestand worden opgeslagen, maar kunnen ook over het netwerk naar externe locaties verzonden worden.

De applicatie of service is vaak zo ontworpen dat het mogelijk is logging te configureren. Een log event, ongeacht de bron, is veelal op dezelfde manier opgebouwd. Veel voorkomende onderdelen van een log event zijn een timestamp, source en message.

Omdat de services of applicaties iedere gebeurtenis verwerkt in een logbestand, bevat dit bruikbare informatie bij het oplossen van storingen. Indien duidelijk is welke handelingen een device of service uitvoert op het moment van, of momenten voor, een storing kan achterhaald worden wat de oorzaak is. Logdata bevat niet alleen bruikbare informatie voor het oplossen van storingen, maar ook voor verschillende calamiteiten zoals volle schijfruimte, hack pogingen of vermiste mailberichten. Logdata geeft daarnaast inzicht op de gezondheid van een systeem of netwerk.

**Welke bronnen genereren de input van de logs?**

Er zijn verschillende bronnen die in staat zijn logdata te genereren zoals Windows, Unix, routers, switches, firewalls, etc. Op Windows of Unix based devices is het mogelijk dat er meerdere services op draaien die ook logevents genereren. Een Ubuntu server met daarop Apache zal naast system logdata ook Apache logdata genereren. Het netwerk van BIT bevat voornamelijk Unix based devices met daarop verschillende services. Die devices transporteren logevents naar een centrale syslog server. Door deze syslog server te onderzoeken in het voor onderzoek is gebleken dat er ongeveer 325 devices in het netwerk van BIT bevindt. Gezamenlijk genereren ze ongeveer 44 miljoen logevents per 24 uur Hoofdstuk 5.1.2 Inventarisatie bevat de volledige inventarisatielijst.

**Wat zijn de eisen en wensen van BIT?**

De opdrachtgever heeft een lijst met eisen en wensen aangegeven aan het begin van het project. Dit is naar eigen inzicht uitgebreid door te kijken welke eisen en wensen eventueel van belang kunnen zijn voor BIT. Het uitbreiden van de lijst met eisen en wensen heeft plaatsgevonden tijdens het vooronderzoek. De resultaten van het vooronderzoek heeft mede bijgedragen aan de aanvullende eisen en wensen. Vervolgens hebben meerdere medewerkers van verschillende afdelingen input gegeven op de lijst. De lijst met eisen en wensen bevindt zich in hoofdstuk 5.1.3.

**Hoe kunnen de logs op een veilige manier centraal opgeslagen worden?**

Het is mogelijk de logevents over een beveiligde verbinding te transporteren naar het log management systeem door gebruik van TCP en TLS. TCP is een betrouwbare transport protocol dat controleert of de pakketjes ook daadwerkelijk zijn aangekomen. TLS zorgt voor een encryptie over de communicatielijn waardoor informatie niet gelezen kan worden door een derde partij. Daarnaast biedt TLS een vorm van authenticatie door te controleren of de ontvangende partij ook daadwerkelijk is wie hij zegt te zijn[19].

Bij het ontwerpen en het realiseren van het implementatieplan is rekening gehouden met de ISO:27001 normering. Hierin staan richtlijnen voor wat betreft de beveiliging van het system. De volgende onderwerpen hebben een raakvlak met de opslag van logdata:

- Retentie
- Klok synchronisatie
- Integriteit

In het Technich Ontwerp hoofdstuk 5.3 staan bovenstaande onderwerpen toegelicht.

**Hoe kan de inhoud van de log berichten efficiënt opgeslagen en verwerkt worden in een database?**

De logevents worden door de graylog-server efficiënt verwerkt en opgeslagen. Dit houdt in dat er vele, soms duizenden, logevents per minuut geparsed en opgeslagen worden. Het parsen van logevents houdt in dat de ononderbroken log berichten in onderdelen gesegmenteerd worden. De gesegmenteerde onderdelen worden in fields opgeslagen. Een enkel log event bestaat na het parsing proces uit meerdere fields.

Graylog-server benoemt het onderdeel dat de logevents parsed een Extractor. Er zijn verschillende Extractors beschikbaar, maar een populaire is GROK[20]. GROK is een verzameling van regular expressions en is in staat informatie uit een log bericht in een field te verwerken, ongeacht de bron. Een voorbeeld hiervan is een log bericht dat een source IP bevat. Dit IP adres kan in een aparte field verwerkt worden en maakt het mogelijk zoekopdrachten uit te voeren op source IP adressen.

[19] (RFC 5246 - The Transport Layer Security (TLS) Protocol Version 1.2, 2015)
[20] (Using Grok patterns to extract data, 2015)

**Op welke manieren kunnen de logs het beste geraadpleegd worden?**

Wanneer de logevents in de database geïndexeerd zijn is het mogelijk de logdata te raadplegen. De beste manier van raadplegen is via de graylog-web-interface die op de interface server draait. De web interface bevat verschillende zoekfunctionaliteiten die het zoeken door logdata efficiënt en gemakkelijk laat verlopen.

Naast de web interface is het via de API mogelijk logdata te raadplegen. De API van Elasticsearch kan benaderd worden mits er een SSH verbinding met de database server is opgezet. De API kan door middel van de command line interface geraadpleegd worden of door middel van scripting.

De web interface van Graylog bevat de mogelijkheid tot authenticatie en autorisatie. Door middel van authenticate is het mogelijk de interface toegankelijk te maken voor bepaalde gebruikers. Autorisatie maakt het mogelijk gebruikers toegang te verlenen tot bepaalde logdata. Dit maakt het mogelijk bepaalde logdata af te schermen voor bepaalde gebruikers.

**Hoe kan het log management systeem beheerd worden?**

Het beheren van het log management systeem wordt behandeld in het gebruikershandboek dat verwerkt wordt op de interne WIKI omgeving van BIT. Het handboek beschrijft onder andere het beheer via de web interface, het updateproces en het schalen van de omgeving. Daarnaast staat hierin beschreven hoe GROK toegepast kan worden zodat er custom parsing rules aangemaakt kunnen worden.

Het beheren van de omgeving is grotendeels via de web-interface mogelijk. Het is bijvoorbeeld mogelijk een input te configureren waarmee bepaald wordt op welke manier de logevents zullen binnenkomen. Daarnaast is het onder andere mogelijk users, GROK patterns en alarmen te configureren. Ook geeft de web interface inzicht in de gezondheid van de omgeving.

Het gebruikershandboek beschrijft niet alleen het beheer van de omgeving, maar ook het gebruik hiervan. Er bevindt zich een uitgebreide beschrijving van de web interface en API in het handboek.

## 6.2 Hoofdvraag

Hoe kunnen de miljoenen logevents die gegeneerd worden door de verschillende devices, en services die daarop draaien, centraal opgeslagen en verwerkt worden zodat hier bruikbare, doorzoekbare en inzichtelijke informatie uit voortkomt?

Het netwerk van BIT is een uitgebreid en divers netwerk dat, gebleken uit de inventarisatie, ongeveer 325 devices bevat die gezamenlijk per dag ongeveer 44 miljoen logevents genereren. Dit zorgt voor een grote stroom aan informatie waarvan het doorzoeken momenteel moeizaam verloopt en er veel bruikbare informatie over het hoofd wordt gezien.

De oplossing betreft een log management systeem dat de logevents, afkomstig van verschillende bronnen, centraliseert, verwerkt en toegankelijk maakt voor de engineers. Aan de hand van de eisen en wensen is een functioneel en technisch ontwerp gerealiseerd dat beschrijft hoe het log management systeem zal werken, welke functionaliteiten het systeem bevat en welke producten gebruikt zullen worden.

Het log management systeem bestaat uit drie componenten namelijk de log collector, database server en interface server. De log collector ontvangt, parsed en transporteert de logevents naar de database server. Op de database server worden de logevents geïndexeerd en kunnen ze vervolgens geraadpleegd worden. Het raadplegen van logevents is mogelijk vanaf de web interface of door middel van de API.

Uit de productselectie is gebleken dat Graylog het beste aan de eisen en wensen van BIT voldoet. Graylog is een open-source oplossing die bestaat uit drie producten: graylog-server, Elasticsearch en graylog-web-interface. Samen vormen de drie producten een log management omgeving. Graylog-server ontvangt en verwerkt de inkomende logevents centraal in Elasticsearch. Vervolgens is het mogelijk de logevents te raadplegen vanaf graylog-web-interface op de interface server.

Graylog-server ontvangt en parsed de inkomende logevents. Dit proces bevordert de zoekfunctionaliteiten, omdat hierdoor de ononderbroken log berichten gesegmenteerd worden in meerdere fields. Het is mogelijk te zoeken op bepaalde waardes van een field en de zoekresultaten kunnen gefilterd worden op fields.

De graylog-web-interface stelt verschillende zoekfunctionaliteiten beschikbaar. Hier is het mogelijk binnen bepaalde onderdelen van een gesegmenteerde log event te zoeken, filteren of sorteren. Doordat de logevents centraal in een database worden opgeslagen is het mogelijk logevents afkomstig van meerdere devices te correleren. Daarnaast is het mogelijk data visueel te presenteren aan de hand van diagrammen, grafieken en overzichten.

Dit alles maakt het log management systeem een efficiënt systeem waar de logevents centraal worden opgeslagen en op een overzichtelijke manier gepresenteerd worden.

# 7 Aanbevelingen

Dit hoofdstuk beschrijft de aanbevelingen voor BIT. De aanbevelingen zijn gebaseerd op de antwoorden op de hoofd en deelvragen en de opgedane kennis bij het uitvoeren van dit onderzoek.

Naar aanleiding van de kwestie en op basis van de antwoorden op de hoofdvraag en deelvragen is het aan te raden een log management systeem te implementeren dat bestaat uit een combinatie van Graylog en Elasticsearch. De gekozen producten zijn voortgekomen uit de productselectie en voldoen aan de eisen en wensen van BIT.

Aan de hand van het proof of concept is gebleken dat Graylog daadwerkelijk in staat is de functionaliteiten te leveren zoals beschreven in het functioneel en technisch ontwerp. Aan de hand van een testplan zijn de criteria, opgezet volgens de eisen en wensen, voldoende getest in samenwerking met de opdrachtgever.

Het log management systeem kan geïmplementeerd worden volgens het implementatieplan. Hierin staat stap voor stap beschreven hoe de systemen en producten geconfigureerd kunnen worden. Wanneer het systeem is geïmplementeerd kan dit aan de hand van het gebruikershandboek beheerd en benut worden.

Het is aanbevolen het log management systeem in eerste instantie in een gecontroleerde omgeving te implementeren. Dit omdat het proof of concept geen inzicht geeft op de betrouwbaarheid van het product op een langer termijn. Verder is het verstandig de syslog server te behouden tot dat zeker is dat het log management systeem betrouwbaar en veilig blijkt.

Logdata kan gevoelige informatie bevatten die niet door derde partijen bekeken mag worden. Door gebruik te maken van TLS over TCP is dit te voorkomen omdat de logevents encrypted over het netwerk worden getransporteerd. Het is aangeraden gebruik te maken van TLS over TCP.

Beide producten worden regelmatig geüpdatet door de developers. De functionaliteiten, betrouwbaarheid en snelheid worden hierdoor constant verbeterd. Wanneer een update beschikbaar is wordt dit aangegeven op de website van Graylog of Elasticsearch. Het is aanbevolen de producten regelmatig te updaten volgens de beschrijving in het gebruikershandboek.

Het log management systeem is een belangrijk onderdeel voor het oplossen van storingen of calamiteiten. Het is van belang dat het systeem bereikbaar is. Daarnaast is het belangrijk dat er geen logevents verloren gaan. Gedurende tijd dat de log collector niet operationeel is kunnen er geen logevents verwerkt worden. Wanneer de database een langere tijd niet beschikbaar is zullen er ook logevents verloren gaan. De log collector bevat een functionaliteit dat logevents lokaal opslaat in een journal tot de database server weer beschikbaar is. Dit is echter een tijdelijke noodoplossing omdat de schijfruimte snel zal vol raken. Het is belangrijk de log collector en database server te monitoren zodat bij uitval direct actie ondernomen kan worden om zo de schade te beperken.

# 8 Bronvermelding


*About BIT*. (2015, Februari 10). Opgehaald van BIT.nl: https://www.bit.nl/en/about-bit/general-about-bit

*Algemeen Datacenters*. (2015, Maart 09). Opgehaald van BIT - De beste techniek, de beste mensen: https://www.bit.nl/nl/datacenters/algemeen-datacenters

*Apache Lucene - Query Parser Syntax*. (2015, April 06). Opgehaald van Apache Lucene: https://lucene.apache.org/core/2_9_4/queryparsersyntax.html

*BIT - Alles voor internetgebruik*. (2015, Januari 25). Opgehaald van BIT: http://www.bit.nl

C, P., & K, S. (2015). *Logging and Log Management.* Verenigde Staten: Syngress Media U.S.

*Configuring a Syslog Agent in Windows Server 2012*. (2015, Maart 16). Opgehaald van Windows Networking: http://www.windowsnetworking.com/articles-tutorials/windows-server-2012/configuring-syslog-agent-windows-server-2012.html

*Docs*. (2015, Maart 23). Opgehaald van Elastic: https://www.elastic.co/guide/index.html

*Docs*. (2015, Maart 23). Opgehaald van Graylog 1.0: http://docs.graylog.org/en/1.0/

*Elastic · Revealing Insights from Data*. (2015, Juni 05). Opgehaald van Elastic: https://www.elastic.co/

*GROK*. (2015, Mei 01). Opgehaald van Google Code: https://code.google.com/p/semicomplete/wiki/Grok

*ISO/IEC 27001:2013*. (2015, Juni 03). Opgehaald van ISO 27001 Security: http://www.iso27001security.com/html/27001.html

*Linux Filesystem Hierarchy*. (2015, Maart 19). Opgehaald van The Linux Documentation Project: http://www.tldp.org/LDP/Linux-Filesystem-Hierarchy/html/var.html

*Log Management | Cloud Log Management Service | Loggly*. (2015, Maart 07). Opgehaald van Loggly: http://www.loggly.com/

M, S. (2015). *Leren Communiceren.* Houten: Noordhoff Uitgevers B.V.

*Open Source Log management with Graylog*. (2015, Maart 16). Opgehaald van Graylog: https://graylog.org/

*Open Source Log Management with Graylog*. (2015, Juni 03). Opgehaald van Graylog: https://www.graylog.org/

*Operational Intelligence, Log Management, Application Management, Enterprise Security and Compliance*. (2015, April 06). Opgehaald van Splunk: http://www.splunk.com/

*Revealing Insights from Data*. (2015, Maart 17). Opgehaald van Elastic: https://elastic.co

*RFC 5246 - The Transport Layer Security (TLS) Protocol Version 1.2*. (2015, Mei 05). Opgehaald van Internet Engineering Task Force: https://tools.ietf.org/html/rfc5246

*RFC 5424 - The Syslog Protocol*. (2015, Maart 18). Opgehaald van The Internet Engineering Taskforce: https://tools.ietf.org/html/rfc5424

*RFC 793 - Transmission Control Protocol*. (2015, Mei 05). Opgehaald van The Internet Engineering Task Force: https://tools.ietf.org/html/rfc793#section-2.6

*TCP vs. UDP*. (2015, Mei 21). Opgehaald van Skullbox: http://www.skullbox.net/tcpudp.php

*The Ins and Outs of System Logging Using Syslog*. (2015, Maart 27). Opgehaald van Sans: http://www.sans.org/reading-room/whitepapers/logging/ins-outs-system-logging-syslog-1168

*The Syslog Protocol*. (sd). Opgehaald van The Internet Engineering Taskforce: https://tools.ietf.org/html/rfc5424

*Using Grok patterns to extract data*. (2015, Juni 03). Opgehaald van Graylog documentation: http://docs.graylog.org/en/latest/pages/extractors.html#using-grok-patterns-to-extract-data

*Using Grok patterns to extract data*. (2015, Juni 09). Opgehaald van Graylog Documentation: http://docs.graylog.org/en/1.0/pages/extractors.html#using-grok-patterns-to-extract-data

*What is a decision matrix?* (2015, Mei 10). Opgehaald van http://www.rfptemplates.technologyevaluation.com/what-is-a-decision-matrix.html

# Bijlagen

1) Plan van aanpak
2) Functioneel ontwerp
3) Technisch ontwerp
4) Proof of concept
5) Implementatieplan
6) Evaluatie