

design your future

Vernieuwen van de on-premise infrastructuur en voorbereiden van migratie naar de cloud

VIVES Technology
bachelor in de elektronica-ICT
afstudeerrichting ICT

campus Brugge

Ben Boydens

academiejaar 2021-2022

Vernieuwen van de on-premise
infrastructuur en voorbereiden
van migratie naar de cloud

Renewing of the on-premise
infrastructure and preparing for
migration to the cloud



Vernieuwen van de on-premise infrastructuur en voorbereiden van migratie naar de cloud

door Ben Boydens

Mentors

Deze eindverhandeling werd geschreven onder begeleiding van volgende mentoren:

- Philip Van Isacker
 - ICT & Senior Developer
- Nico De Witte
 - Projectmedewerker/Lector - VIVES Hogeschool
 - nico.dewitte@vives.be

Over de auteur

Ben Boydens is iemand die van jongs af aan al gefascineerd was door computers daarom dat hij een studie richting heeft gekozen in de ICT.

Ben is iemand die zeer praktisch is en graag dingen maakt. Hij is ook geduldig en heeft een goed gevoel voor teamspirit. Tijdens het weekend is hij scouts leiding en gaat ik graag eens weg met zijn medeleiding.

Dankwoord

Graag zou ik enkele mensen bedanken die me hebben geholpen tijden mijn stage en bachelorproef.

Eerst zal ik Philip Van Isacker bedanken om mij tijdens de stage op te leiden en hulp te bieden tijdens de bachelorproef. Ook wil ik de mensen van Dataline bedanken om mij te verwelkomen in hun bedrijf en het mogelijk te maken om mijn bachelorproef daar te doen.

Dan wil ook nog een dankwoordje geven aan mijn Vives-mentor Nico De Witte voor zijn feedback en snelle antwoorden.

Ten slotte wil ik nog eens dankjewel zeggen tegen de Vives leraren die me hebben geholpen om tot hier te geraken.

Digitale versie

De digitale versie van deze thesis kan worden geraadpleegd via <https://thesis-dataline-ben-2022.netlify.app>.

Bachelorproef aangeboden tot het behalen van het diploma van Bachelor in de elektronica-ICT. Deze eindverhandeling was een examen. De tijdens de verdediging geformuleerde opmerkingen werden niet opgenomen."

Abstract

Elk bedrijf wil graag een IT-omgeving die veilig is en klaar voor de toekomst. Maar om dit te gaan realiseren hangen er meestal zware kosten aan. Er is geen enkele goede manier om iets te doen alles heeft zijn voor- en nadelen op het vlak van IT. Daarom is een goede analyse nodig om te weten wat de volgende stap is.

Een stap naar de toekomst kan zijn om gebruik te maken van Cloud computing. Het concept ervan bestaat al een tijdje maar nu beginnen bedrijven zich er echt voor in te zetten. Steeds meer kleine bedrijven gaan naar de cloud om van de voordelen te profiteren zoals schaalbaarheid en toegankelijkheid.

Een ander belangrijk concept is data storage van applicaties. Deze moeten een zo hoog mogelijke up-time hebben zodat deze altijd beschikbaar zijn. Maar als iets misloopt met de storage dan kan het zeer lang duren om de applicatie terug werkende te krijgen. Een gekende manier om fout tolerantie te gaan introduceren op het vlak van storage is met een vSAN. Dit is een virtueel netwerk van opslag en het zorgt dat je storage zo goed als altijd beschikbaar is.

Om een plan op te stellen voor de IT-omgeving van een bedrijf moeten alle applicaties bekeken worden. Per applicatie moet beslist worden of deze naar de cloud moet of best lokaal blijft op een vSAN.

In deze thesis wordt synchronisatie van een lokale omgeving naar de cloud voorzien. Er wordt een prijsanalyse gedaan op applicaties zoals Confluence, Jira en email om te migreren naar de cloud.

Verschillende types vSAN worden besproken en Starwind vSAN wordt getest. Deze resultaten worden gebruikt om andere opstellingen te vergelijken.

Abstract - English version

Every company wants an IT environment that is secure and ready for the future. But there are heavy costs for realizing this. There is no single right way to do something, everything has its pros and cons when it comes to IT. That is why a good analysis is necessary to know what the next step is.

A step towards the future could be to use Cloud computing. The concept has been around for a while, but now companies are really starting to commit to it. More and more small businesses are turning to the cloud to take advantage of its benefits such as scalability and accessibility.

Another important concept is data storage of applications. Apps should always have the highest possible uptime so that they are constantly available. But if something goes wrong with the storage then it can take a long time to get the application back to work. A well-known way to implement fault tolerance of storage is with a vSAN. This is a virtual network of storage and it ensures that your storage is almost always available.

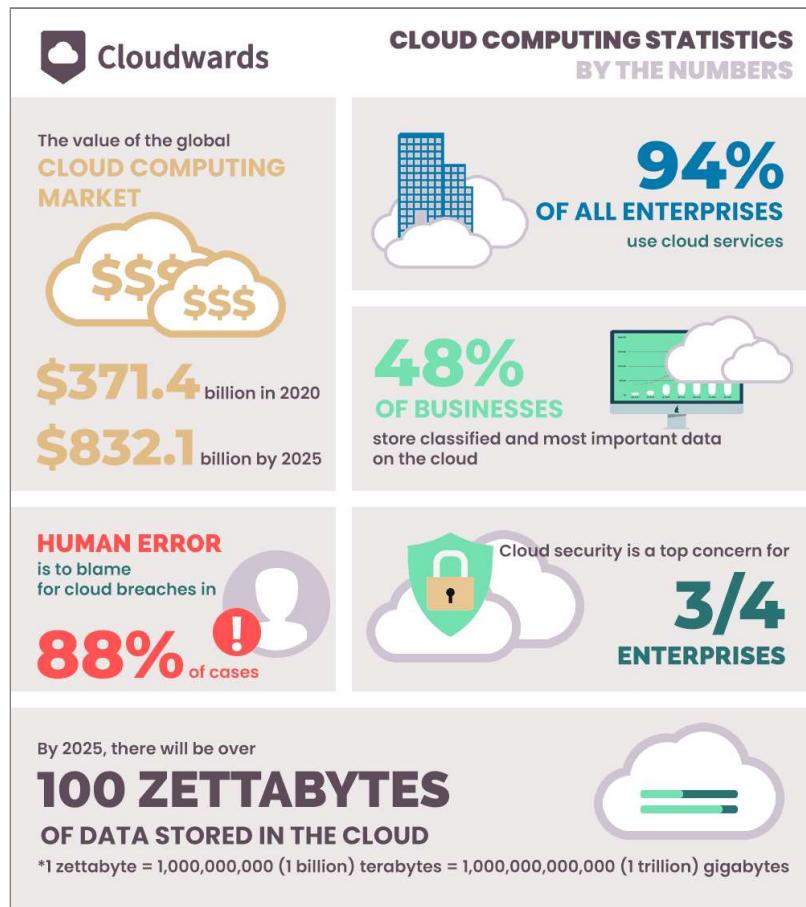
To create a plan for a company's IT environment, all applications need to be looked at. A decision must be made for each application. Will this application move to the cloud or should it remain local on a vSAN.

In this thesis, synchronization is provided from a local environment to the cloud. A price analysis is done on migrating apps such as Confluence, Jira and email to the cloud.

Different types of vSAN are discussed and Starwind vSAN is tested. These test result serves as a baseline to compare other setups.

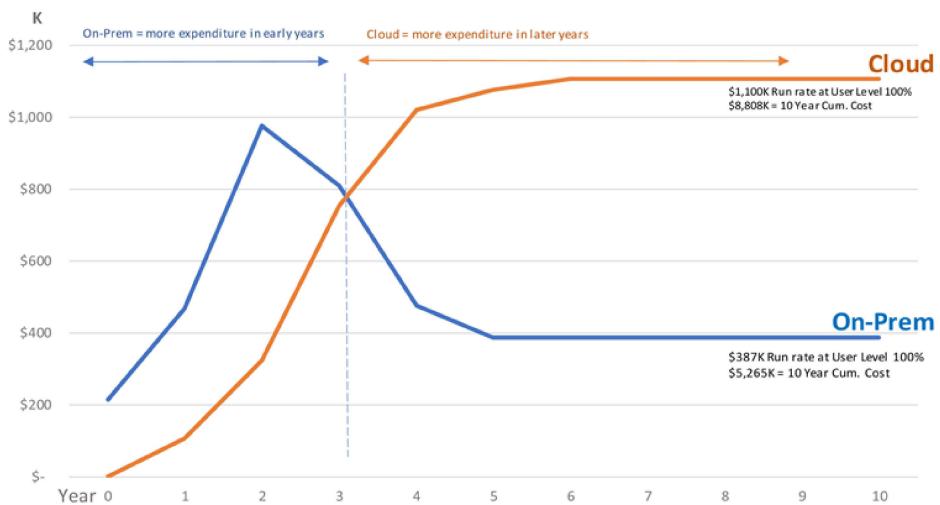
Introductie

Heel wat bedrijven gebruiken nog lokale servers om hun applicaties en data op te slaan. Dit geeft hen veel controle over wat er gebeurt met hun IT omgeving, maar heeft ook zijn nadelen. Zoals het feit dat het bedrijf een team moet samen stellen om dit op te onderhouden. Ook kan het zeer kostelijk zijn om uit te breiden en nieuwe apparatuur aan te kopen. Een alternatief hiervoor is **Cloud computing** en het wordt steeds populairder. **Sumina, 2022**↗.



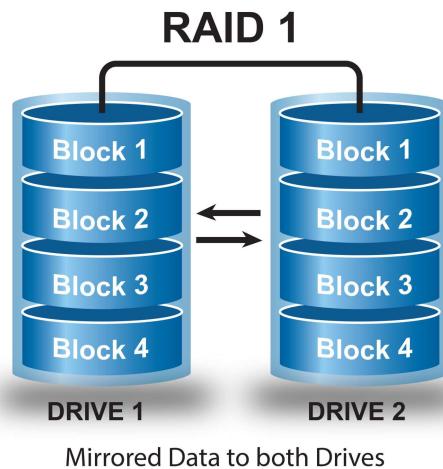
Een groot voordeel van Cloud computing is dat het zeer flexibel is. Werken in de cloud is veel simpeler om de omgeving up-to-date te houden en te gaan uitbreiden. Stel dat er nood is aan extra bandbreedte of storage, dan kan de cloud direct aan die vraag voldoen, in plaats van een complexe (en dure) update van de IT-infrastructuur te ondergaan.

Aan de andere kant kan cloud computing duur zijn, elke maand moet er een som betaald worden en na een lange periode lopen die kosten op. Wanneer er lokaal gewerkt is dit niet, daarom dat lokaal werken op lange termijn goedkoper is. Het is dus zeer belangrijk om te kijken welke applicaties mogen blijven en welke beter naar de cloud vertrekken. In grafiek is er een vergelijking van de kost tussen de 2 te zien. **Fisher, 2018**↗



Een ander belangrijk aspect is fouttolerantie. Het kan altijd gebeuren dat een applicatie of service plots niet meer werkt. In dat geval moet er een alternatief zijn om snel de services terug up and running te krijgen. Er moet dus vermeden worden dat er een **Single Point Of Failure** bestaat. Zodat wanneer één iets kapot gaat er direct een alternatief is die de taak kan overnemen. Een manier om applicaties en data fouttolerant te maken zou kunnen zijn om alles meerdere malen te gaan opslaan. Dit noemt men in het Engels ook **Redundancy**(overbodigheid).

Bijvoorbeeld met een RAID configuratie kan data worden geduplicateerd over meerdere harde schijven. Dit zorgt dat de er altijd een kopie wordt gemaakt.



Er zijn 3 opties waartussen een bedrijf kan kiezen om hun applicaties en data op te slaan.

- On Premise
- De Cloud
- Een hybride oplossing

Elke van deze opties heeft zijn voor- en nadelen. Wat is nu de beste oplossing qua prijs, gebruikersgemak en beveiliging? Dat is wat ik zal proberen te beantwoorden in deze analyse.

On premise

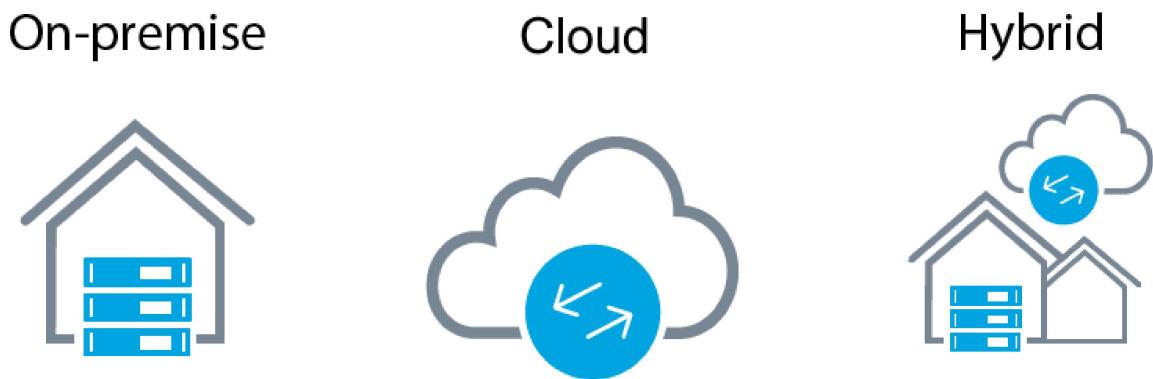
Alle data en applicaties worden opgeslagen op lokale servers in het bedrijf. Meestal wordt dit afgesloten van de buitenwereld en zijn de toepassingen enkel beschikbaar op de werkvlloer. Dit is gemakkelijk omdat alles heel dichtbij staat en je veel vrijheid hebt wat je doet met de servers.

De Cloud

De cloud of cloud computing is een dienst die aangeboden wordt door een bedrijf om software of hardware ter beschikking te stellen over het internet. De cloud provider zorgt voor het onderhoud en kosten van de fysieke apparatuur zoals de fysieke apparatuur zoals de servers. Gebaseerd op de noden van je bedrijf kun je dan een aantal servers huren.

Hybride Oplossing

Met een hybride oplossing wordt er gedeeltelijk in de cloud gewerkt en ook on premise. Het kan soms zijn dat er een applicatie draait in de cloud maar verbonden is met een server on premise.



Dataline Solutions nv

Dataline is een bedrijf dat actief is in de print industrie en een ontwikkelaar van software. Multipress is één van hun applicaties, het digitaliseert het volledige process van drukkerijen. Van het eerste contact met de klant tot het controleren van de magazijn, en zelf de facturen opmaken. Dataline focust ook op het helpen van de klanten en probeert zich steeds uit te breiden. Multipress is de meest populaire MIS/ERP software voor de print industrie in heel Europa.

Dataline zelf heeft in hun hoofdkantoor in Loppem een hele IT infrastructuur. Daarbij horen een heleboel applicaties die gebruikt worden door de werknemers. Sommige applicaties die worden

niet meer ondersteund en moeten naar de cloud. Daarom moet er onderzocht worden welke applicatie's best naar de cloud verplaatsen en welke best nog on premise blijven.

Dit is zeer belangrijk voor Dataline voor de toekomst en de kostprijs van de IT infrastructuur. Hoe moet hun infrastructuur veranderen om up-to-date te blijven. Cloud computing kan ook de kostprijs verlagen en gebruikersgemak verhogen.

Doelstelling

Dit is een complex probleem waar een analyse gemaakt moet worden van de huidige IT omgeving, kijken hoe die kan worden verbeterd en een plan opstellen voor de toekomst.

In de eerste fase bekijken worden de verschillende applicaties die Dataline heeft bekeken. Alle servers worden opgeliist die in de huidige omgeving gebruikt worden. Specificaties van de servers worden bekeken zoals het aantal CPU's, de hoeveelheid RAM en de hoeveelheid storage. Uit deze lijst wordt er gekeken welke applicaties naar de cloud moeten en welke die beter on premise zouden blijven.

In de tweede fase wordt er onderzoek gedaan naar elke applicatie en er wordt gekeken naar de mogelijkheid om naar de cloud te gaan. Verschillende cloud providers worden bekeken en er wordt een onderzoek gedaan op wat de beste mogelijkheid is qua prijs efficiëntie. Ook wordt er gekeken naar het migratie process naar de cloud. Hoe moet de data naar de cloud gebracht worden en kunnen er problemen optreden.

In de derde fase wordt er gekeken naar de applicaties die lokaal moeten blijven. Voor die applicaties moet er ook gekeken worden naar de storage ervan zodat die applicaties fout tolerant zijn. Er zijn hier vele mogelijke aanpakken voor, welke is geschikt voor Dataline en past in hun budget?

Ten slotte word een test opstelling gemaakt op vlak van storage. Dit om te kijken of deze storage optie wel goed is voor de prijs.

Huidige omgeving

In de eerste fase kijken wordt de huidige omgeving van Dataline bekeken. De applicaties en services die nu in gebruik zijn worden opgeliijst. Er wordt een lijst opgesteld van alles die naar de cloud mag en alles die juist lokaal moet blijven.

Alle applicaties en services worden uitgevoerd op virtuele machines. Dit maakt het toekennen van resources gemakkelijk en zorgt dat het besturingssysteem kan gekozen worden. Per virtuele machine wordt er een aantal vCores, RAM en Storage toegekend. vCores is het aantal CPU cores die de virtuele machine denkt dat hij heeft. vCores komen niet 1 voor 1 overeen met fysieke cores, maar eerder met draden (threads).

Hieronder bevindt zich een oplijsting van alle applicaties, de virtuele machines die gebruikt worden en de resources van ervan.

Confluence

Confluence is een content management systeem van Atlassian. Het is simpel gezegd een plek waar mensen documenten kunnen aanmaken en delen met elkaar. Het wordt door iedere werknemer gebruikt en bevat belangrijke documenten. De versie die nu gebruikt wordt van Confluence is de Server editie. Atlassian zelf wil afstappen van de lokale server editie en daarom verplichten ze de gebruikers om naar de cloud over te stappen.

Er moet bekeken worden hoe Confluence zal gebracht worden naar de cloud zodat de gebruikers een transparante overgang ondervinden en dat alle huidige data veilig bewaart word.

Servers

Server	CPU	RAM	Storage
Confluence	4 cores	28 GB	256 GB

Active Directory

Active directory is directory service van Microsoft die gebruikt wordt om gebruikers te gaan beheren in een domein. Wanneer een nieuwe werknemer toekomt in Dataline wordt er voor hem of haar een nieuw account aangemaakt in Active Directory. Dit zorgt dat mensen toegang hebben tot applicaties en zich overall kunnen inloggen.

Active directory dient als een centrale database om gebruikers in op te slaan. Heel veel applicaties maken er gebruik van om gebruikers te gaan authenticeren.

Servers

Er zijn 2 windows Servers die instaan voor het beheer van Active Directory in Dataline. Hier zijn er 2 van op elk kantoor van Dataline.

Server	CPU	RAM	Storage	Locatie
DC1	2 cores	4 GB	32 GB	Loppem
DC2	2 cores	4 GB	32 GB	Limmen (Nederland)

Mail server

Dataline heeft een eigen mail server en een mail filter server. De mail server is een GFI Kerio connect mail server. Dit is een mail oplossing voor kleinere bedrijven die niet te veel geld willen investeren voor een mail oplossing. De mail filter server wordt gebruikt om binnengekomen e-mails te gaan filteren. Het doet dit op basis van een aantal regels en de geschiedenis van de e-mails die binnen zijn gekomen. Het heeft een black list en een white list van email adressen.

Een mogelijkheid is om over te stappen op Microsoft Outlook. Deze oplossing heeft al een ingebouwde mail filter.

Servers

Server	CPU	RAM	Storage
Mail server	4 cores	8 GB	700 GB
Mail filter	4 cores	6 GB	20 GB

Telefonie servers

De telefonie servers zijn zeer belangrijk voor Dataline. Deze worden gebruikt om contact op te nemen met de klanten en om support aan te bieden. De telefonie servers draaien zelf om virtuele machines met een gratis licentie van EXSi (VMWare). Er is een probleem met deze servers is dat een **single point of failure** is. Wanneer er iets fout loopt met de servers dan moeten deze hersteld worden.

Dit gebeurt door te kijken naar de back ups van die servers en dan zo de server te gaan herstellen. Dit is ten eerste zeer traag en er zijn niet veel mensen in Dataline die weten hoe dit moet. Dus er moet hier zeker een oplossing gezocht worden om de storage van deze servers fout tolerant te maken.

File server

De file server wordt voornamelijk gebruikt in Dataline door het marketing team. Daar wordt er gebruikt gemaakt van zeer grote bestanden die moeten gedeeld worden via deze server. De grote van deze server is 6,5 TB.

Er is een mogelijkheid om de server gedeeltelijk naar de cloud te migreren. Dit zou de fileserver makkelijker toegankelijk maken aangezien mensen dan ook thuis toegang zouden hebben. De file server zelf is een SAMBA file share die gekoppeld wordt met LDAP. De koppeling met LDAP zorgt dat mensen zich eerst moeten inloggen met hun gebruikers account.

Backup server

Op de back up server worden incrementele back ups genomen van de virtuele machines, databases, mail server, etc. Dit zou ook eventueel gedeeltelijk naar de cloud kunnen gebracht worden. De bandbreedte van de verbinding kan wel een probleem zijn.

Development Servers

Bij het ontwikkelen van software komen ook development servers. Op deze servers wordt software getest en gebuild. Er worden vele server gebruikt. In totaal zijn er:

- 4 development servers
- 2 staging servers
- 2 build servers

Er zijn ten slotte ook nog Git, SonarQube en devops servers. Al deze servers zouden lokaal moeten blijven dus ze brengen naar de cloud is geen optie. Er kan wel gekeken worden naar de storage van al deze servers en hoe dat het best aangepakt wordt.

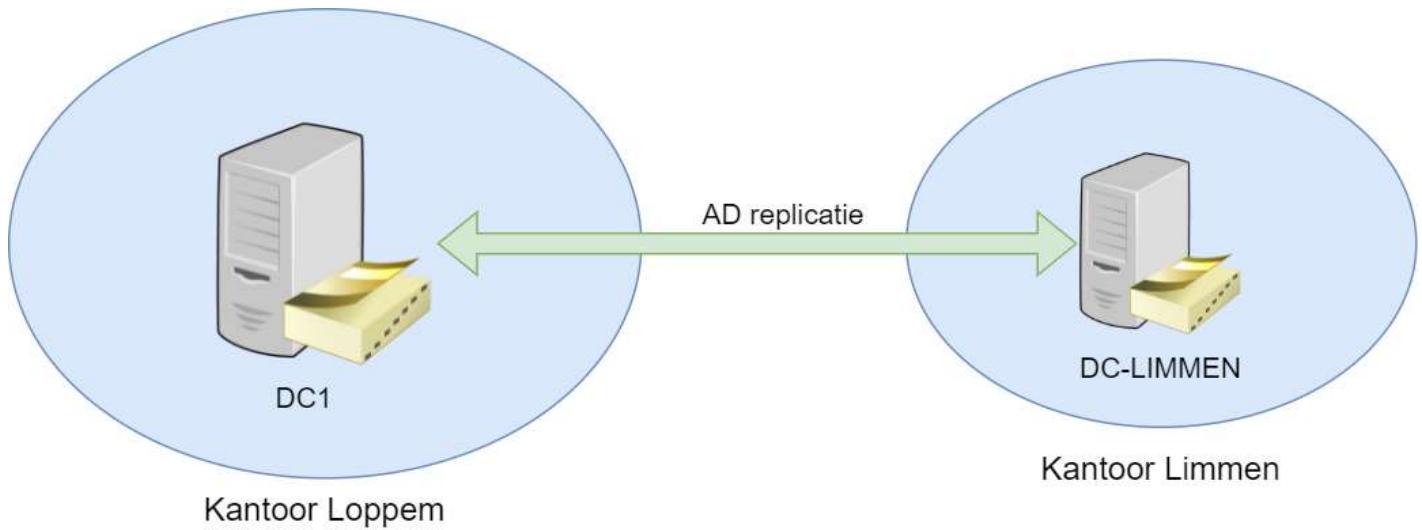
Active Directory

Active Directory is de naam die wordt gegeven aan de directory service van Microsoft voor een Windows domein. Het komt inbegrepen bij meeste Windows Servers als een verzameling van services. Een Windows server die Active Directory draait staat ook bekend als een **domeincontroller**.

Domeincontroller

De domeincontroller (dc) is een Windows Server waar alle gebruikers en computers worden opgeslagen. Het zal de gebruikers en computers gaan authenticeren en autoriseren in een Windows domein. Bijvoorbeeld als iemand inlogt met een windows computer die lid is van het domein dan zal de domeincontroller de ingegeven gebruikersnaam en wachtwoord controleren.

Het grootste voordeel hiervan is dat alles op 1 plek wordt opgeslagen waardoor het gemakkelijk is voor IT personeel om alles te onderhouden. Wanneer er meerdere domeincontrollers zijn dan zullen die onderling gesynchroniseerd worden zodat ze altijd in sync zijn met elkaar. Dataline heeft nu in totaal 2 domeincontrollers, één staat in Loppem en de andere staat in Limmen (Nederland).



De 2 domeincontrollers staan op 2 geografisch verschillende locaties. Een andere naam voor locatie die wordt gebruikt in context van domeincontrollers is **Site**. Hier zal er synchronisatie tussen 2 verschillende Sites gebeuren en dit wordt ook wel een **intersite verbinding** genoemd. Wanneer 2 dc's zich op dezelfde site bevinden, wordt het een **intrasite verbinding** genoemd.

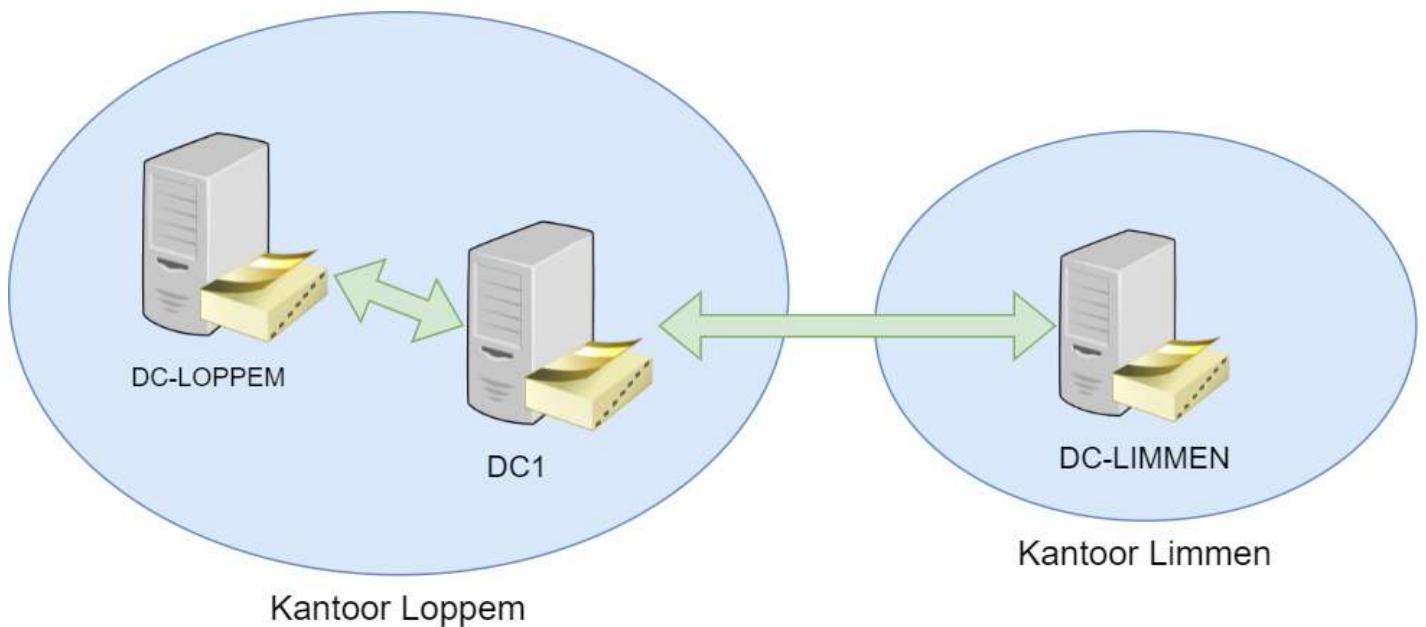
Services

Een domeincontroller staat in voor heel wat taken binnen een domein, zo zijn er een heleboel ingebouwde services waarvoor een domeincontroller kan dienen. Een paar voorbeelden hiervan zijn:

- Gebruikers en computers beheren
- Gebruikers authenticeren
- Security policies afdwingen op het domein
- DNS server
- DHCP server

Nieuwe Domeincontroller

Een windows server versie van 2016 of later is nodig om de synchronisatie van de domeincontroller naar de cloud mogelijk te maken. Domeincontroller DC1 voldoet hier niet aan en daarom moet er een nieuwe domeincontroller geïnstalleerd worden. De domeincontroller krijgt de naam **DC-LOPPEM**. Deze domeincontroller zal in het begin samen werken met DC1 en na een bepaalde periode zal DC1 uitgeschakeld worden en neemt DC-LOPPEM de taak over.

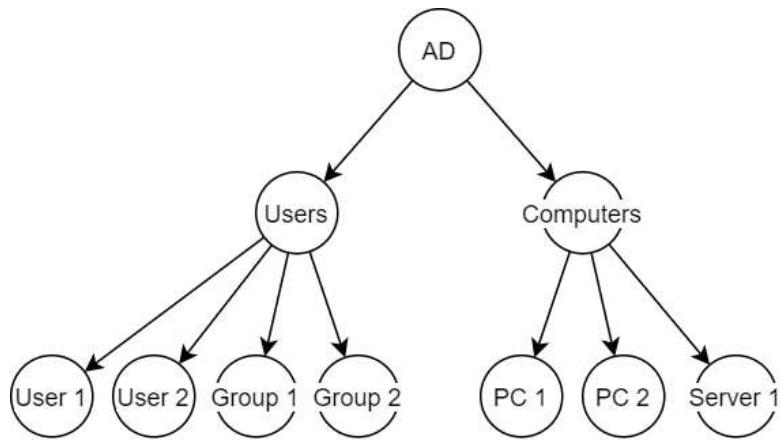


OU-structuur

Een Organizational Unit (OU) is een manier om Active Directory objecten te gaan groeperen. Het wordt gebruikt om een hiërarchische structuur te creëren binnen het domein, vergemakkelijkt het beheer en kan gebruikt worden om de verschillende afdelingen en locaties voor te stellen. OU's kunnen andere OU's bevatten en Microsoft adviseert het gebruik ervan.

Huidige structuur

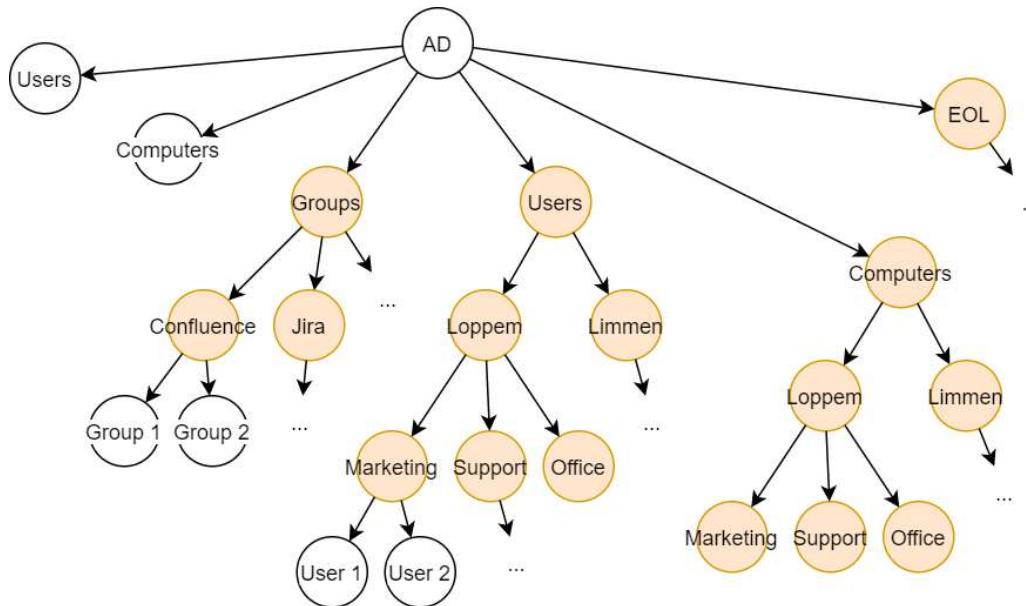
De huidige structuur van Dataline werkt nog niet met OU's en er zou liefst overgestapt worden naar een duidelijke OU-structuur. Nu wordt nog een oude structuur gebruikt die er als volgt uitziet.



In bovenstaande figuur kan er worden gezien dat 2 mappen gebruikt worden, namelijk 'Users' en 'Computers'. Dit zijn 2 standaard directories die in elke domeincontroller aanwezig zijn. Users wordt gebruikt voor gebruikers en groepen terwijl computers dient voor computers en servers. Dit is een oude manier van werken waar alle objecten samen zitten en er is geen structuur zit in de indeling.

Nieuwe structuur

De nieuwe manier van werken zal OU's gebruiken om objecten te groeperen per soort, per afdeling en per locatie. Op die manier is het gemakkelijker werken en zijn er meer mogelijkheden om gebruikers te beheren. De nieuwe structuur ziet er als volgt uit. De oranje nodes stellen OU's voor.



Er zal een OU zijn voor enkel groepen en die zal nog eens verdeeld worden onder de verschillende applicaties. Er is een OU voor gebruikers die wordt opgesplitst per locatie en dan per afdeling. Op deze manier wordt het direct duidelijk waar en op welke afdeling een gebruiker werkt. Voor de OU Computers wordt dezelfde structuur gevolgd.

De objecten die niet meer gebruikt worden komen in de groep EOL (End-of-life) terecht. Dit kan handig zijn want het gebeurt vaak dat een werknemer weg gaat en dan later terug komt.

Dit is niet de volledige structuur, maar het geeft een beeld van wat het zou moeten worden.

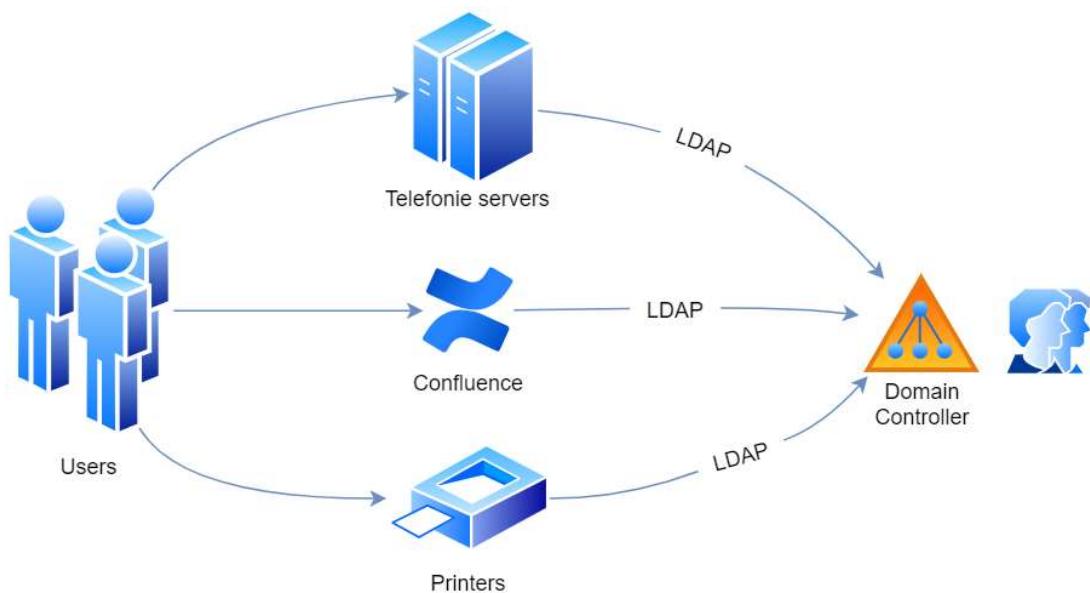
Problemen

Er moet overgestapt worden naar de nieuwe structuur maar dit brengt enkele problemen met zich mee. Vele applicaties en services gebruiken nog de oude structuur en kunnen mogelijk niet meer functioneren als ze overschakelen op de nieuwe structuur. De reden hiervoor is omdat vele applicaties enkel in de Users directory zoeken naar gebruikers en groepen.

De manier waarop die applicaties gaan communiceren met de domeincontroller is aan de hand van het **LDAP protocol**.

LDAP

LDAP of **Lightweight Directory Access Protocol** is een netwerkprotocol voor het opvragen en onderhouden van een directory service over TCP/IP. Het is een open standaard die door veel applicaties ondersteund wordt. Apps en services van Dataline gebruiken het als een interface om gegevens op te vragen van Active Directory.



Ter illustratie worden de printers van Dataline eens bekeken. Die hebben de optie om een document in te scannen en door te mailen naar een werknemer. De printer zelf zal een LDAP query gaan uitvoeren op de domeincontroller om alle gebruikers op te halen. Die query zal er als volgt uitzien.

- **LDAP query root:** CN=Users,DC=dataline,DC=eu
- **LDAP query:** (objectclass=user)(objectCategory=person)

De **root** is de plek waar de query zal uitgevoerd worden en hier wordt dit de Users map. Dat wil zeggen dat enkel objecten gevonden worden die in de Users map zitten of in een submap ervan. De **query** zelf zal zoeken naar alle gebruikers objecten.

Nieuwe Queries

Om de applicaties klaar te maken voor de nieuwe structuur moet bij elke applicatie de **query root** aangepast worden, zodat ook objecten niet in de Users map ontdekt kunnen worden. De **LDAP query** zelf heeft geen verandering nodig want er wordt nog steeds gezocht naar dezelfde soort objecten.

Bij de printers wordt de query dan uiteindelijk:

- **LDAP query root:** OU=Dataline Users,DC=dataline,DC=eu
- **LDAP query:** (objectclass=user)(objectCategory=person)

Voorbeelden

Om een beter zicht te krijgen hoe LDAP queries in elkaar zitten worden nog een enkele voorbeelden gegeven die van pas kunnen komen.

Gebruikers zonder email

(objectClass=user)(objectCategory=person)(!mail=*)

Deze query zoekt naar user objecten die behoren tot de 'Person' categorie. Dit zal alle accounts terug geven die gelinkt zijn aan een persoon (geen admin of service accounts). Een user object heeft een veld 'mail' die het mail adres bevat. Met `!mail=*` worden alle gebruikers die een mail adres hebben gezocht en met het uitroepingsteken worden de selectie geïnverteerd.

Alle administrators accounts

(objectClass=user)(objectCategory=person)(adminCount=1)

Enkel administrators krijgen in het `adminCount` veld een waarde van 1.

Alle lege groepen oplijsten

(objectCategory=group)(!member=*)

Hier wordt gezocht naar alle objecten die tot de 'Group' categorie behoren. Met `!member=*` wordt er gezocht naar de groepen die geen waarde hebben voor het member veld.

Alle gebruikers lid van de groep Marketing

(objectclass=user)(MemberOf=CN=Marketing,CN=Users,DC=dataline,DC=eu)

Hier wordt gezocht naar alle user objecten die lid zijn van een groep. Het `MemberOf` veld heeft als waarde een ander Active Directory object. Om dit te doen moet het pad ingegeven worden van dat

object, in dit geval wordt dit `CN=Marketing,CN=Users,DC=dataline,DC=eu` . Deze waarde komt overeen met het een object Marketing in de Users map op het domein `dataline.eu`.

Microsoft Azure

Dataline zelf heeft al een Office 365. Dat is een online cloud omgeving die allerlei services zal aanbieden. Het wordt gebruikt om bijvoorbeeld Office licenties voor Word, Excel en PowerPoint toe te kennen aan gebruikers. Deze gebruikers worden opgeslagen in een Azure AD omgeving.

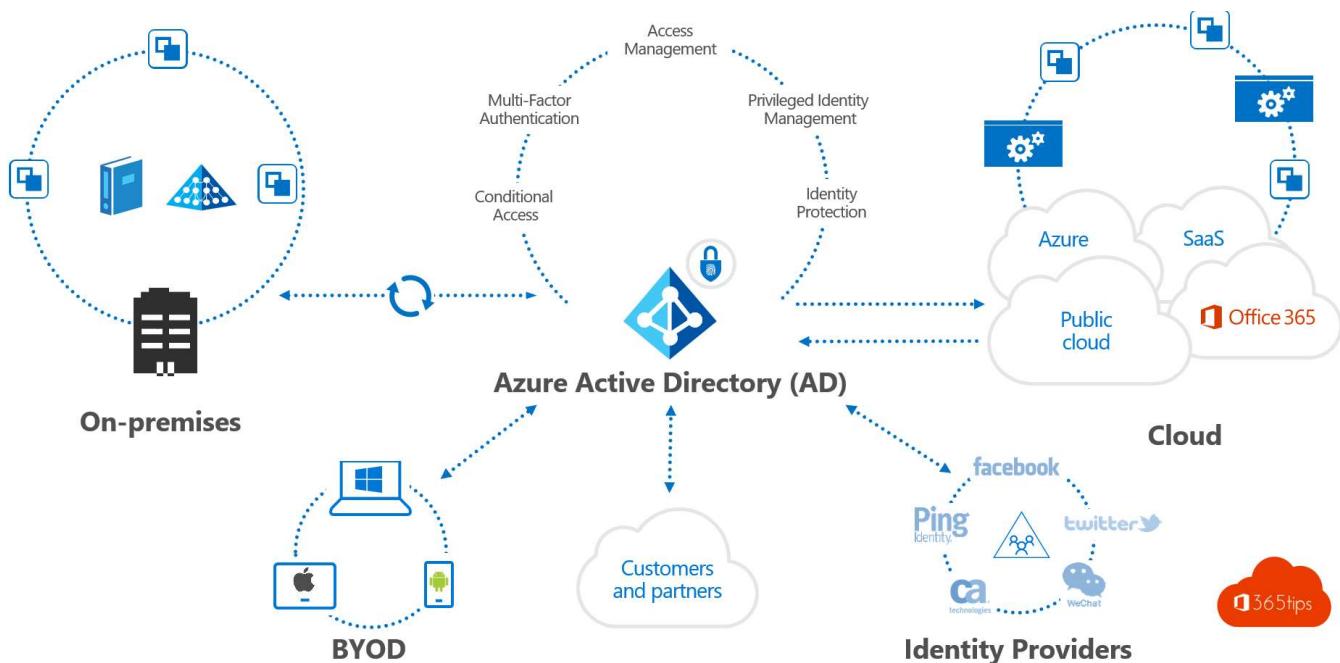
Met de huidige implementatie hebben gebruikers dan 2 accounts:

- Een Azure AD account in de cloud
- Een on-premise Active Directory account

Deze 2 accounts hebben aparte wachtwoorden en dit kan lastig zijn voor gebruikers. Om de huidige omgeving te brengen naar de cloud moeten deze accounts gesynchroniseerd worden. Eenmaal gesynchroniseerd kunnen de accounts in Azure AD worden gebruikt om de gebruikers te authenticeren in de cloud. Deze accounts kunnen dan worden geïntegreerd met andere cloud applicaties.

Azure AD

Azure Active Directory (Azure AD) is een identiteits- en toegangsbeheerservice in de cloud. Het helpt om werknemers toegang te geven tot externe resources zoals Office 365 en andere cloud toepassingen.



Zoals je kan zien in de figuur dient Azure AD als een centrale plek die dient om de gebruiker accounts overal te verspreiden. Wat wij moeten gaan realiseren is de pijl tussen de **on-**

premises en Azure AD. Microsoft geeft ons een heleboel manieren om dit probleem aan te pakken. Hier worden de opties overlopen om tot de juiste conclusie te komen.

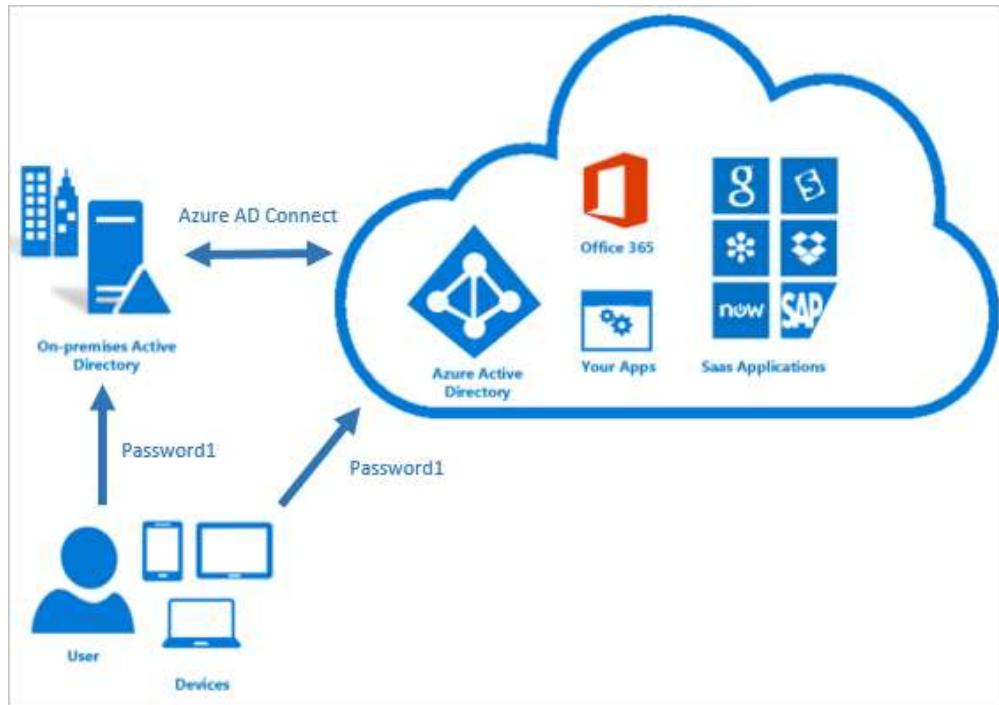
Bepalen van Hybrid Identity

Om gebruikers te gaan identificeren is een oplossing nodig die werkt on premise maar ook in de cloud. Microsoft noemt zo'n oplossing een **Hybrid Identity**. Het zorgt dat gebruikers zich kunnen gaan authenticeren en autoriseren zowel in de cloud als on premise. Er zijn 3 verschillende methoden om een hybrid identity te gaan implementeren:

- Password hash sync
- Pass-through authentication
- Federation

Password hash synchronization

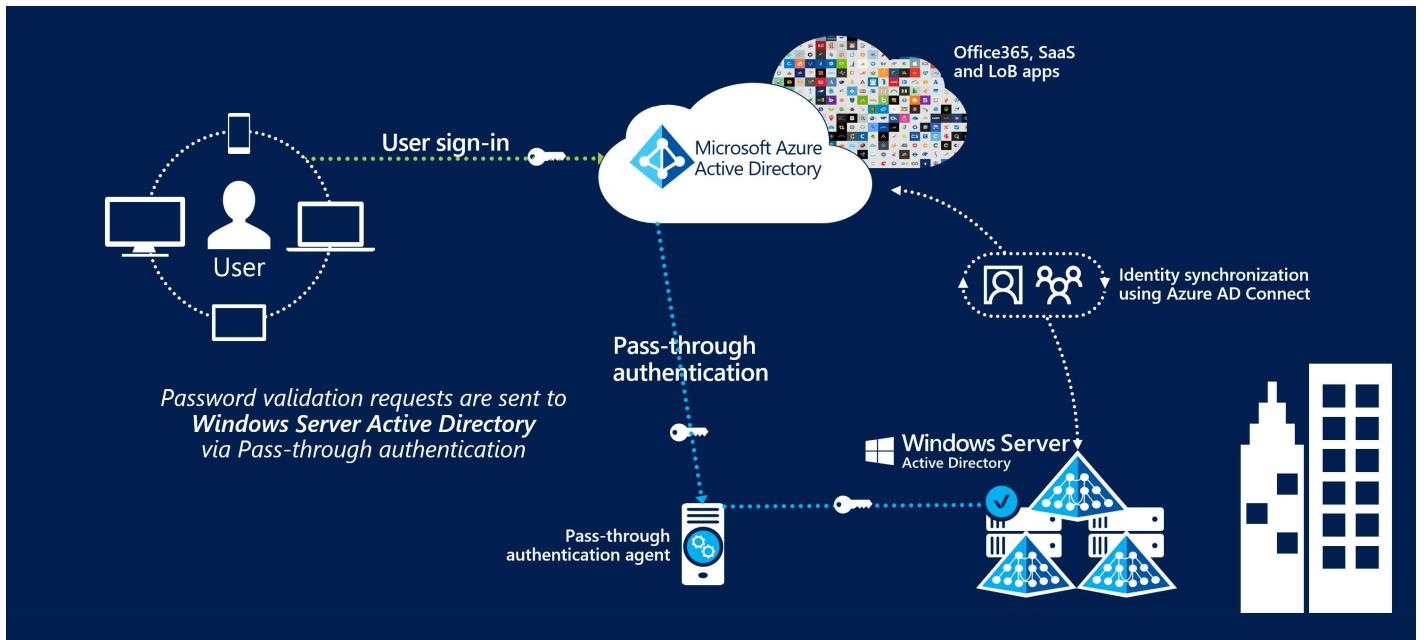
De hash waarde van het wachtwoord in Active Directory zal gesynchroniseerd worden met een hash die wordt opgeslagen in de cloud. Zo kunnen gebruikers inloggen in de cloud met hetzelfde wachtwoord. Dit is de standaard methode om authenticatie te gaan doen en het is ook de gemakkelijkste manier.



Password hash sync heeft ook een optie **Leaked credential detection**. Microsoft werkt samen met dark web onderzoekers en law enforcement agencies om gelekte credentials te vinden. Als Microsoft merkt dat er wachtwoorden van jouw organisatie tussen zitten dan wordt er een melding gegeven.

Pass-through authentication

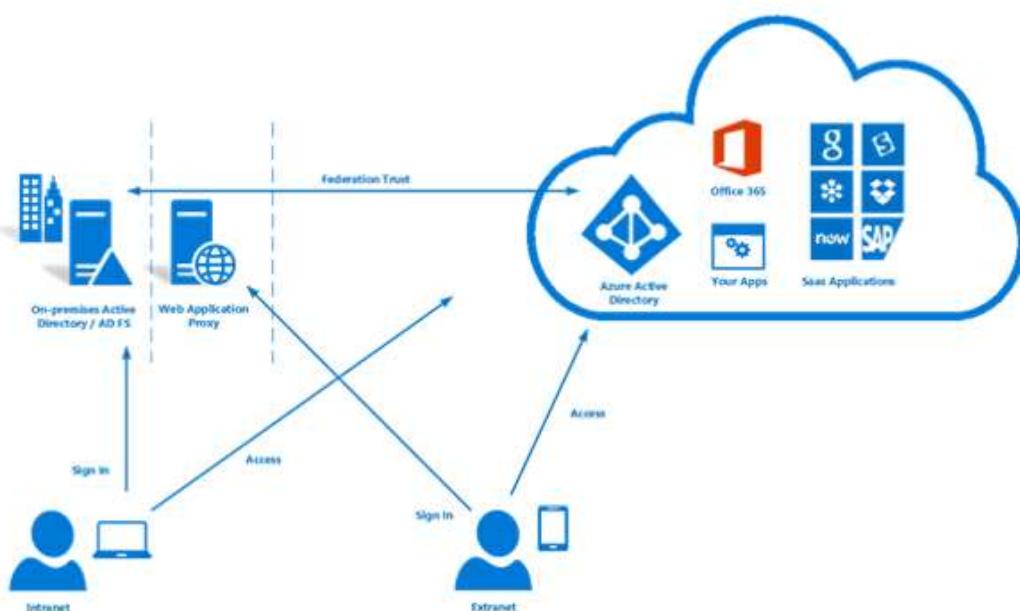
Met Pass-through authentication gebeurt de authenticatie niet meer in de cloud maar on-premise. Wanneer een gebruiker probeert in te loggen via de cloud wordt er een verbinding gemaakt met de Active Directory die lokaal op het kantoor staat.



Een voordeel hiervan t.o.v. password hash sync is dat alle authenticatie nu gebeurt op de lokale domain controller. Dit zorgt dat lokale security en password policies toegepast kunnen worden zelf als de gebruiker inlogt via de cloud.

Federation

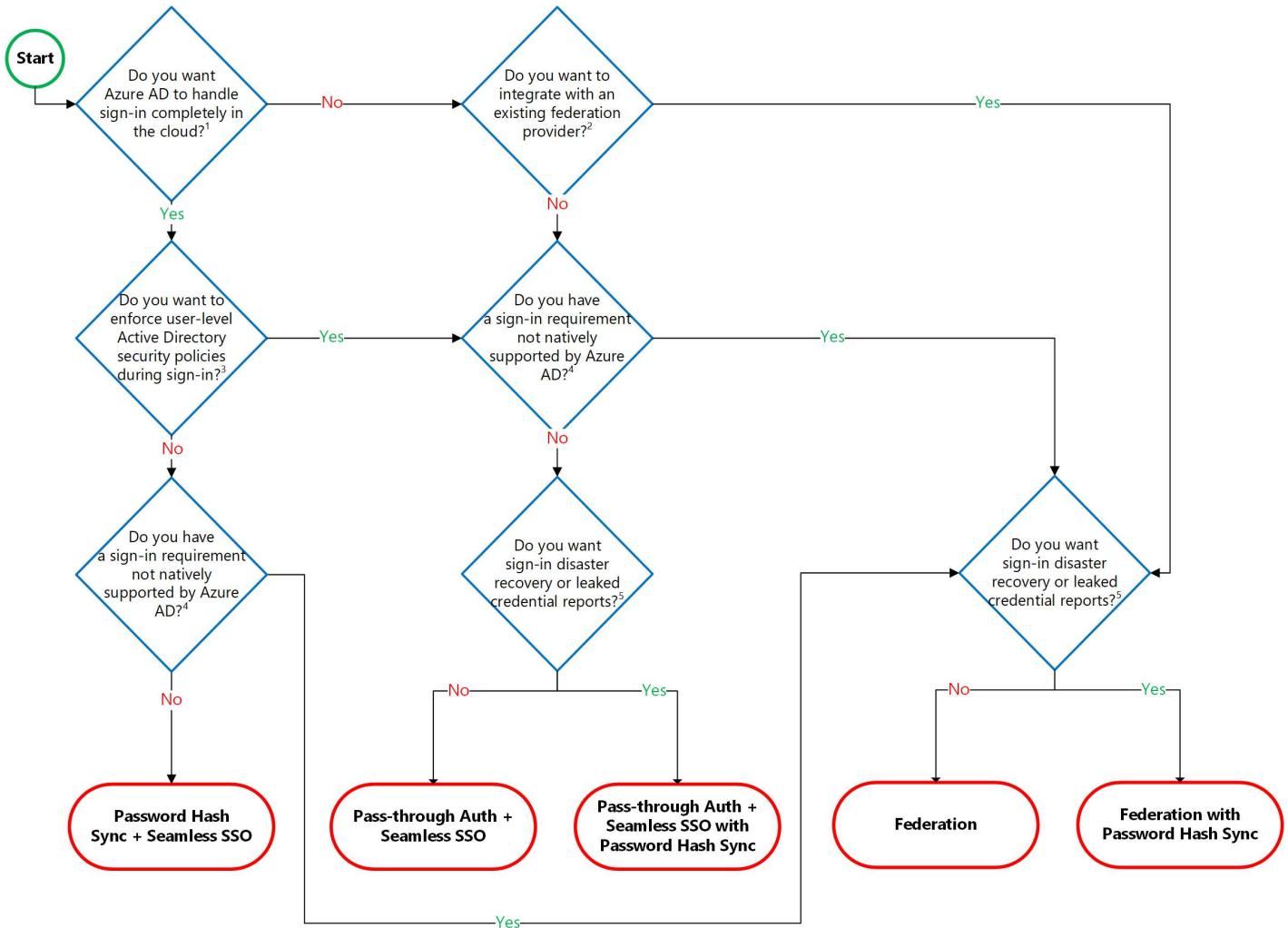
Deze aanpak gebruikt een aparte vertrouwde server om de authenticatie te gaan doen. Mensen die van buiten af toegang willen tot de cloud moeten zich eerst authenticeren bij die server. Mensen die lokaal proberen in te loggen kunnen gewoon gebruik maken van de on-premise active directory.



Het grote voordeel van Federation is dat er enorm veel vrijheid is om de authenticatie te gaan doen. De andere 2 methoden zijn standaard methoden van microsoft die beperkte functionaliteit hebben. Zo kun je bijvoorbeeld met federation smartcard authenticatie implementeren wat niet mogelijk is met de andere methoden.

Hybrid Identity Keuze

Om ons te helpen met de juiste keuze te maken heeft Microsoft een flowchart opgesteld.



Microsoft geeft ons 5 mogelijke oplossingen om mee te werken.

- Password Hash Sync
- Pass-through Auth
- Pass-through Auth + Password Hash Sync
- Federation
- Federation + Password Hash Sync

Password hash Sync kan gecombineerd worden met de andere authenticatie methoden. Zo kan de **Leaked Credentials Detection** van password hash sync combineren met Pass-through authentication en Federation. De combinatie met Password hash sync heeft nog een voordeel en

dat is dat het kan gebruikt worden als backup methode voor moest er iets mislopen. Dit wordt ook **Sign in disaster recovery** genoemd.

Keuze

Federation zal zeker geen optie zijn voor Dataline. Het is lastig om op te zetten en is overbodig, een andere methode zou beter passen. De Leaked credentials report zou een feature zijn die van pas zou kunnen komen. Daarom dat onze keuze zeker Password Hash Sync moeten bevatten.

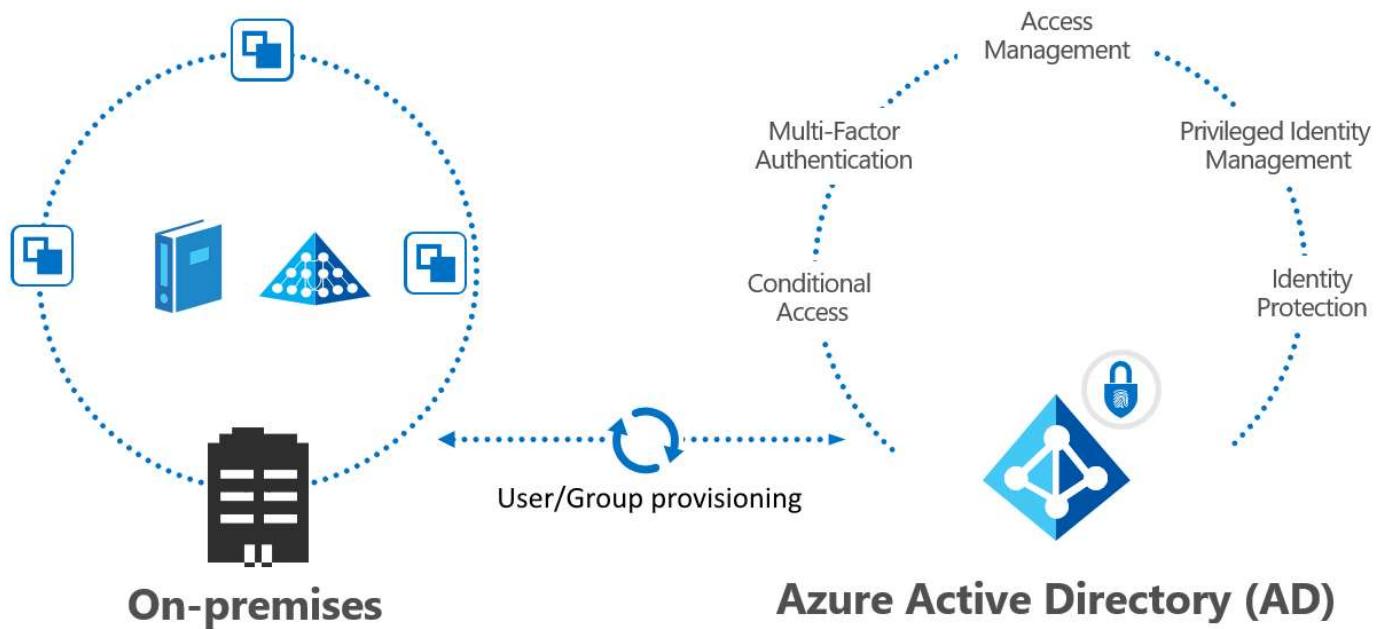
Dat geeft ons nog 2 keuzes:

- **Password Hash Sync**
- **Pass-through Auth + Password Hash Sync**

Pass-through Auth + Password Hash Sync zo ideaal zijn maar het laat ons niet toe om de Cloud Sync Agent te gebruiken. Dit is iets wat later nog aan bod komt maar daarom gaat onze keuze naar **Password Hash Sync**.

User en group provisioning

Dataline wil de Active Directory synchroniseren met Azure AD. Dit noemt provisioning en het is een process die automatisch gebruikers gaan aanmaken, verwijderen en up to date houden.



Er zijn 2 opties om automatic user en groep provisioning te gaan doen.

Azure AD connect sync

Veel agent heeft goede support en is robust, zeker een mogelijke optie. Kan moeilijk zijn om te configureren en kostelijk om te onderhouden. Heeft ook een grote investering nodig op vlak van

infrastructuur (sterke server nodig voor synchronisatie).

Azure AD connect cloud sync

Nieuwste optie support, zeer snel en makkelijk op te zetten. Hoge availability. Is lightweight dus geen nood aan een sterke server voor de synchronisatie.

Besluit

Er is niet echt een goede server om de Connect Sync server te runnen en de extra features van Connect Sync zijn niet echt nodig. De belangrijkste feature was password writeback en deze wordt ondersteund door beide. Daarom gaat de voorkeur naar de Cloud Sync methode. Een volledige lijst met alle verschillen tussen de 2 kunt u vinden in de Microsoft Docs [hier](#)

Soft- en Hardmatch

Om de bestaande account van Office 365 te synchroniseren met de Active Directory account moet er een manier zijn om deze gesynchroniseerd te krijgen met elkaar. Maar hoe weet de synchronisatie agent welke accounts overeen komen met elkaar? Dit gebeurt aan de hand van een **Soft- of Hardmatch**.

De Office 365 accounts zijn belangrijk omdat die gebruikt worden om licenties aan gebruikers toe te kennen (bv Word). Als er iets misloopt bij het matchen van de 2 accounts zal er een nieuw account aangemaakt worden. Dat zorgt ervoor dat de licentie niet zal toegekend zijn voor die persoon. Het is dus belangrijk dat dit goed verloopt.

Softmatch

Bij een soft match gaat er gekeken worden naar 2 attributen van een gebruiker.

- **proxyAddresses**
- **userPrincipalName**

Het attribuut **proxyAddresses** bestaat uit meerdere delen. Hier zal er enkel naar het SMTP gedeelde gekeken worden wat neer komt op de email van de gebruiker.

De **userPrincipalName** komt neer de systeem representatie van een gebruiker in een email formaat. Meestal komt dit overeen met het email address van de gebruiker maar niet altijd!

Hardmatch

Bij een hard match gaat er gekeken worden naar een enkel attribuut.

Namelijk **sourceAnchor/immutableID**. Dit is een soort identifier die uniek is voor elke gebruiker.

Idfix

Om zeker te zijn dat er geen problemen zouden voorkomen bij het synchroniseren raad Microsoft aan om de idFix tool te gebruiken. Het zal mogelijke problemen gaan opsporen die kunnen optreden bij synchronisatie naar de cloud. Het checkt voor duplicates, missing attributes, en rule violations.

Uitvoering

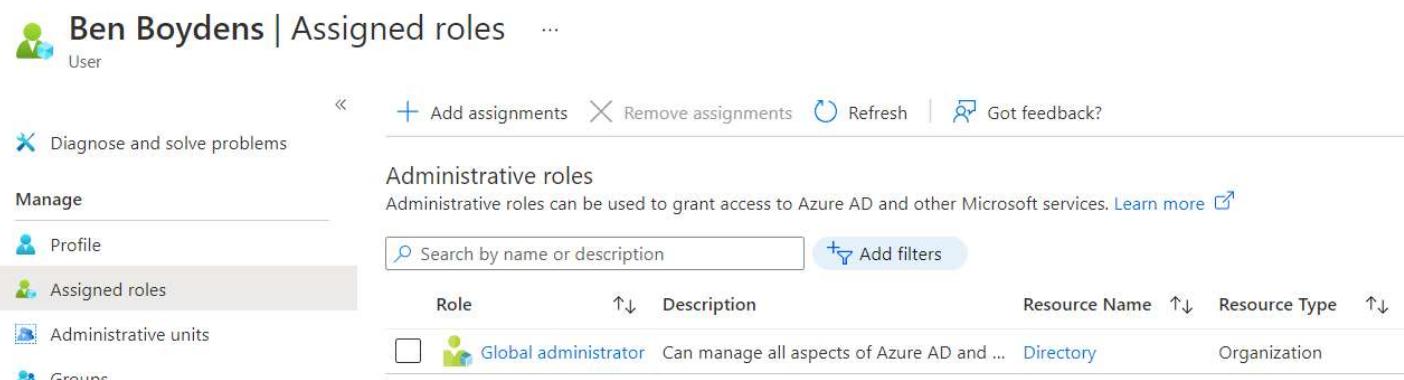
Dan wordt het tijd om de cloud sync agent te gaan installeren op de nieuwe domain controller. Dit gebeurt gewoon via een installer die je kan downloaden van de Azure AD web interface. Er wordt gecontroleerd voor problemen met de gebruikersaccounts met de idFix tool. De tool vindt geen problemen, dus kan de synchronisatie starten.

Problemen Softmatch

Voor dat de synchronisatie start laat Azure AD ons toe om eerst een enkele gebruiker te synchroniseren als test. Er wordt een poging gedaan om een enkele gebruiker te synchroniseren, maar er loopt iets mis. In figuur kan gezien worden dat er een nieuwe gebruiker wordt aangemaakt en dat de oude gebruiker blijft bestaan zonder dat die gesynchroniseerd is.

Search users		Add filters					
55 users found							
Name	User principal n...	User type	Directory synced	Account enabled	Identity issuer	Company name	
<input type="checkbox"/>  BB Ben Boydens	ben.boydens@datali...	Member	No	Yes	datalineholding.onmicrosoft.com		
<input type="checkbox"/>  BB Ben Boydens	ben.boydens9789@...	Member	Yes	Yes	datalineholding.onmicrosoft.com		
<input type="checkbox"/>  BV Ben Boydens	ben.boydens9789@...	Member	No	Yes	datalineholding.onmicrosoft.com		
<input type="checkbox"/>  BV Ben Boydens	ben.boydens9789@...	Member	

De reden dat hier is misloopt is omdat het gebruiker account een admin role heeft toegekend.



The screenshot shows the 'Assigned roles' section for a user named Ben Boydens. It lists three entries, each with a checkbox, a user icon, and the role name. The first two entries are 'Global administrator' and the third is 'Cloud administrator'. The third entry is highlighted with a grey background.

Role	Description	Resource Name	Resource Type
<input type="checkbox"/>  Global administrator	Can manage all aspects of Azure AD and ...	Directory	Organization
<input type="checkbox"/>  Cloud administrator	Can manage cloud services and ...	Cloud services	Organization

Microsoft zegt zelf:

Azure AD Connect isn't allowed to soft match a user object from on-premises AD with a user object in Azure AD that has an administrative role assigned to it. [Link](#)

Dit wordt gedaan voor de veiligheid omdat het matching van gebruikers automatisch gebeurt en je dus geen controle hebt over wie het zal matchen. Je **wilt** zeker niet per ongeluk een admin role gaan toekennen aan een gebruiker die het niet nodig heeft. Dus voor veiligheidsredenen laten ze het niet toe.

Dit is snel opgelost door de role tijdelijk weg te doen van elke gebruiker. Na dit te doen werkte de synchronisatie zoals verwacht.

Confluence

Confluence is een content management systeem gemaakt door Atlassian. Confluence is de centrale hub waar werknemers hun documenten kunnen delen met elkaar. Atlassian wilt afstappen van hun server producten en overstappen naar een cloud oplossing. Deze overstay moet goed gebeuren omdat Confluence een heleboel belangrijke en essentiële documenten die Dataline nodig heeft om te kunnen werken.

Huidige omgeving

De huidige omgeving van Confluence moet klaar gemaakt worden om naar de cloud te gaan. De huidige omgeving is hier nog **niet** klaar voor. Dit geeft een paar redenen:

- Confluence groepen moeten gelinkt zijn met groepen in AD
- De huidige permissies zijn chaotisch en moeten opgekuisht worden

Groepen

Confluence heeft de optie om groepen te gaan gebruiken van in Active Directory. Nu is er een mengeling van confluence groepen en AD groepen. Er moet overgestapt worden naar enkel AD groepen. Dit is een handige feature want wanneer een werknemer lid moet worden van een groep dan kan dit op de domain controller gedaan worden en hoeft dit niet apart te gebeuren in Confluence.



In de afbeelding hierboven is groep A een Confluence groep terwijl groep B een Active Directory groep is.

Permissions

Permissies kunnen toegekend worden aan **groepen** of aan **individuele gebruikers**. De Confluence omgeving van Dataline heeft nu een heleboel individuele permissies die weg moeten. De beste manier om te werken is door permissies te gaan toekennen per groep, omdat het zo eenvoudig is om permissies te geven en afnemen. Ook wanneer Confluence naar de cloud gaat dan is het nodig dat permissies alleen via groepen worden uitgedeeld.

In Confluence zijn permissies **additief**. Als iemand lid is van meerdere groepen (A, B, C, etc), dan is wat ze doen de som van de machtigingen die zijn verleend aan Groep A + Groep B + Groep C.

Space permissions

Admins van een space dan kunnen een heleboel permissies toekennen voor die space, zoals wie wat kan zien en wat ze ermee kunnen doen. De space permissies worden toegekend aan alles in de space.

Permissies in een space kunnen gegeven worden aan individuele gebruikers en groepen:

- **Group permissions** zijn toegepast op alle mensen die lid zijn van de groep
- **User permissions** zijn toegepast op een individuele gebruiker

Pagina Restricties

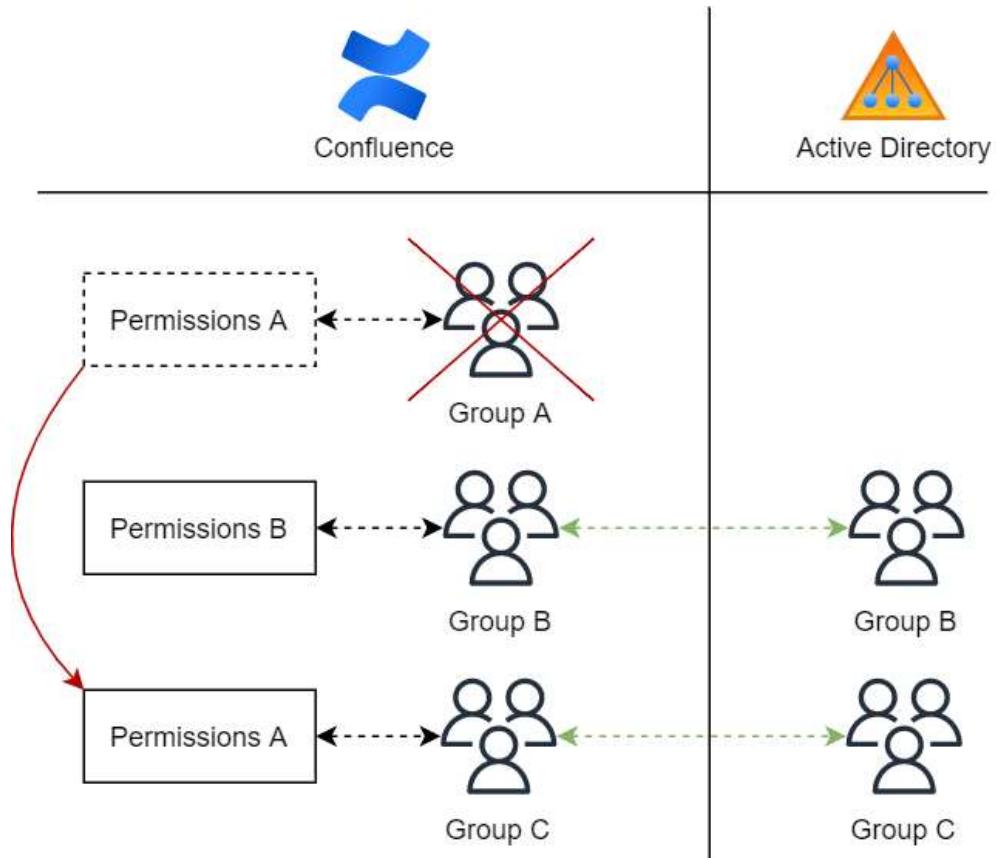
Pagina restricties werken net iets anders dan Space permissies. Pagina's kunnen standaard gezien en bewerkt worden, maar je kan dit beperken voor bepaalde users of groepen als dat nodig is.

Elke pagina in Confluence bevindt zich in een space, en space permissies geven de admin de mogelijkheid om de zichtbaarheid van alles in de space te beheren. Zelf de mogelijkheid om permissies aan te passen van een pagina's wordt gecontroleerd door de *restrict pages* space permissie.

AD groepen

Het doel is om alle groepen in Confluence Active Directory groepen te maken. Nu zijn er nog een aantal groepen die enkel in Confluence zitten en die groepen hebben bijhorende permissies. Het is niet mogelijk om een nieuwe AD groep te linken met een bestaande

Confluence groep. Daarom moeten de permissies van de Confluence groep overgezet worden naar een nieuwe AD groep.



De permissies van groep A worden overgezet naar de Active directory groep C. De groep A wordt dan verwijderd vanuit Confluence. Het overzetten van de permissies moet gebeuren aan de hand van een SQL query omdat Atlassian deze functionaliteit nog niet geïmplementeerd heeft.

Confluence Producten

Confluence stopt support voor de standaard server producten. Dat is de versie die Dataline nu voorlopig nog gebruikt. De support hiervoor stopt in 2024. Er zijn in totaal 3 producten van Confluence.

- Server editie
- Datacenter editie
- Cloud editie

De server editie stopt in 2024 en Atlassian verplicht mensen om over te stappen naar de Datacenter of Cloud editie. De cloud editie meerdere programma's om tussen te kiezen:

	Free	Standard	Premium	Enterprise
Prijs per user	0	\$5,50	\$10,50	Jaarlijkse Factuur
Prijs per maand (50 users)	0	\$275	\$525	Jaarlijkse Factuur
Max aantal gebruikers	10	20.000	20.000	20.000
Storage	2 GB	250 GB	unlimited	unlimited

Dan hebben is er ook nog de Datacenter versie die \$27.000 kost per jaar. Daarin zit alle functionaliteit en kan er nog lokaal on-premise gewerkt worden. Atlassian zegt zelf het volgende

Met onze Data Center-producten kun je profiteren van de flexibiliteit om te implementeren op een infrastructuur naar keuze. Dit is de beste keuze voor degenen met unieke of complexe operationele vereisten of die verder willen opschalen dan onze huidige cloudbruikersniveaus. Als je upgradet heb je volledige controle over gegevensbeheer, beveiliging en compliance, en over hoe je uptime en prestaties beheert. **We raden Data Center aan voor degenen die strengere vereisten hebben en nog niet kunnen overstappen naar cloud.**

Mijn eerste indruk is dat de Datacenter editie niet voor ons zal zijn aangezien het duur is en enkel nodig is als je de vereisten niet hebt om naar de cloud over te stappen. De logische keuze lijkt te gaan naar de standard cloud editie, aangezien die goedkoop is en juist genoeg storage zal hebben. Onze huidige confluence server heeft namelijk 256 GB aan storage.

Overstap naar de cloud

Migreren gebeurt met de Confluence Cloud migration Assistant. Dit is een applicatie die het gemakkelijk maakt om de data, gebruiker en groepen van Confluence te brengen naar de Cloud. Voor Jira is er een andere applicatie beschikbaar namelijk de Jira Cloud Migration Assistant. Die werkt gelijkaardig als de Confluence migration assistant.

Migration Assistant home

This tool will help you assess and create your migration plan. You'll then be able to test and migrate to a cloud-hosted site.

ASSESS

 **1. ASSESS YOUR APPS**
Decide which apps you need to bring to cloud. The assessment must be 100% complete before you can migrate your app data.

PREPARE

 **2 . PREPARE YOUR APPS**
Connect to a cloud site and install your apps before you migrate to cloud.

MIGRATE

 **3. MIGRATE YOUR DATA**
Migrate users, groups, spaces and apps to Confluence Cloud in stages, or all at once.

ABOUT YOUR PRODUCT

2 Groups
15 Users
4 (14 pages, 390 KB) Spaces
7 Apps

ADDITIONAL RESOURCES

[Preparing for migration](#)

View our best practice guides for more information on app migration security, migration testing, and preparation.

[Confluence Cloud Migration Assistant](#)

View our guides and learn how to prepare for your move.

Assess your apps

In onderstaande afbeelding staan alle applicaties (plugins) die geïnstalleerd zijn op Confluence. Per applicatie staat er dan of de app kan gemigreerd worden en indien dit mogelijk is het pad die moet gevuld worden. Er zijn een heleboel applicaties die amper gebruikt worden en die waarschijnlijk gewoon weg mogen. De belangrijkste applicaties zullen de volgende zijn: Handy Macro's, PocketQuery, Reporting en Scaffolding.

[X Close](#)[Give feedback](#)[Need help? ▾](#)

Assess your apps

The statuses you assign in this table guide your app migration. The Use Alternative status allows you to continue through the flow to select alternative apps to install.

User installed apps (18) *	Status	Exists in cloud	Appears on	Viewed by	Can be migrated	Notes
Angular JS integration...	No decision made		Not applicable*	Not applicable*		<input type="text" value="Enter your notes here"/>
Avono Read Confirmatio...	No decision made		33 pages	22 users		<input type="text" value="Enter your notes here"/>
Balsamiq Wireframes ...	No decision made	View differences	47 pages	12 users	View path	<input type="text" value="Enter your notes here"/>
Confluence Dutch (Be...	No decision made		Not applicable*	Not applicable*		<input type="text" value="Enter your notes here"/>
Confluence Source E...	No decision made		Not applicable*	Not applicable*		<input type="text" value="Enter your notes here"/>
Easy Confluence Tran...	No decision made		Not applicable*	Not applicable*		<input type="text" value="Enter your notes here"/>
Handy Macros for Co...	No decision made	View differences	1647 pages	62 users	Automated path Stage 2	<input type="text" value="Enter your notes here"/>
Lucidchart Plugin	No decision made	View listing	4 pages	0 users	Contact vendor	<input type="text" value="Enter your notes here"/>
PocketQuery	No decision made	View differences	4519 pages	52 users	View path	<input type="text" value="Enter your notes here"/>
Reporting	No decision made	View listing	3639 pages	55 users	View path	<input type="text" value="Enter your notes here"/>
Scaffolding	No decision made	View differences	2874 pages	50 users	Automated path Stage 1	<input type="text" value="Enter your notes here"/>

Merk op dat er bij 'can be migrated` verschillende mogelijkheden zijn gebaseerd op hoe vlot het migratie process zal gebeuren.

	Beschrijving
	Applicatie kan niet gemigreerd worden aangezien er geen alternatief is voor in de Cloud.
	Applicatie kan gemigreerd worden naar de cloud maar niet volledig automatisch. Er zijn nog bepaalde dingen waar rekening moet mee gehouden.
stage 1	Applicatie in stage 1 hebben een ongekende of lage migratie success rate. Met andere woorden er is een grote kans dat het fout loopt. Bij problemen moet er contact opgenomen worden met de app vendor.
stage 2	Applicaties in stage 2 hebben een hoge success rate voor migratie naar de cloud.

BELANGRIJK!

Er zijn een aantal Queries van **PocketQuery** die niet ondersteund worden in de cloud! Die plugin is nodig om data van externe systemen op te lijsten in confluence. PocketQuery wordt nu door 4519 pagina's in Confluence gebruikt. Het kan gemigreerd worden naar de Cloud maar wel niet automatisch. Er zijn nog enkele queries die gebruikt worden die niet zullen werken in de cloud. Daarom moet er nog even gewacht worden met de migratie totdat de queries vervangen zijn door functionele cloud alternatieven.

Automatic User/Group Provisioning

Automatic User/Group provisioning is het process die automatisch gebruikers en groepen gaan aanmaken, verwijderen en up to date houden. Nu haalt Confluence de gebruikers en groepen van de lokale domain controller, maar als er verhuist wordt naar de cloud hoe zal dit dan gebeuren? Atlassian geeft ons de mogelijkheid om dit via **Azure AD**, maar er moeten wel aan een aantal prerequisites voldaan zijn.

Voorwaarden

Om de automatic user/group provisioning te activeren zijn er enkele voorwaarden:

- An Azure AD tenant
- A user account in Azure AD with permission to configure provisioning
- An Atlassian Cloud tenant with an **Atlassian Access subscription**
- A user account in Atlassian Cloud with Admin permissions

Om automatic user provisioning te gaan implementeren is er een abonnement nodig is voor **Atlassian Access**. Atlassian Access is een apart programma die extra kost bovenop de confluence cloud subscriptie. Hieronder kan je de prijzen zien van de verschillende abonnementen:

Prijs per user	100 users	500 users	1000 users
Confluence	\$5,50	\$5,50	\$5,50
Confluence + Access	\$9,50	\$7,44	\$6,79

Jira

Een ander product van Atlassian is Jira. Dit wordt ook gebruikt in Dataline en het is een issue tracking programma die bugs tracked en zorgen voor agile project management. De producten die Atlassian heeft van Jira zijn de volgende:

- Jira Cloud
- Jira Server
- Jira Datacenter

Gelijkwaardig met Confluence stopt de support voor de server versie in 2024. De cloud editie is de logische keuze aangezien de Datacenter edition zeer duur is en gericht is op grote bedrijven die nog niet klaar zijn om naar de cloud over te stappen. De cloud editie komt in volgende versies:

	Free	Standard	Premium	Enterprise
Prijs per user	0	\$7,50	\$14,50	Jaarlijkse Factuur
Prijs per maand (50 users)	0	\$375	\$725	Jaarlijkse Factuur
Max aantal gebruikers	10	20.000	20.000	20.000
Storage	2 GB	250 GB	unlimited	unlimited

Voor ons is de standard versie de beste keuze aangezien deze het goedkoopst is en er genoeg features in zitten

Mail Server

Zoals vele bedrijven heeft Dataline hun eigen mail server. Die mail server zou ook naar de cloud kunnen worden gebracht. Er moet moet gekeken worden naar wat de opties zijn. Zal dezelfde software gebruikt worden in de cloud of is het beter dat er overgestapt wordt naar een ander alternatief? Dat zijn vragen die in dit hoofdstuk beantwoord zullen worden.

Kerio Connect

Nu wordt er gebruikt gemaakt van een Kerio Connect server. Kerio Connect is een product van GFI software die veilige mail services en kalender diensten voor een lage prijs zal aanbieden. Lokaal op Dataline zijn er 2 servers namelijk een lokale Kerio Connect Mail server en een Mail filter server.

Server	Specs
Mail server	4 cores, 8 gb RAM, 700 gb storage
Mail filter	4 cores, 6 gb RAM, 20 gb storage

Pricing

Het aantal licenties die nu gebruikt worden zijn in totaal 100. Er worden zo veel licenties aangekocht om te voldoen aan de huidige werknemers en de toekomstige. De prijs van een lokale Kerio Connect server is enorm goedkoop maar dit is natuurlijk zonder de onderhoudskosten van de servers. Hieronder vindt u overzicht van de pricing.

	Small (10-49)	Medium (50-249)	Large (>250)
Prijs per gebruiker per jaar	€32,50	€29,50	€26,50

Met 100 gebruikers en een Medium abonnement zou de prijs neer komen op een totaal van **€2950 per jaar**. Dat is niet de prijs inbegrepen van de mail filter server en de kosten om die server zelf draaiende te houden. Maar voor 100 licenties is dit wel enorm goedkoop.

Kerio Cloud

GFI biedt ons de mogelijkheid aan om naar een gehoste omgeving over te schakelen. GFI zelf zal wel geen hosting doen, dit gebeurt via een partner bedrijf. GFI stelt hiervoor 2 bedrijven voor in de regio België/Nederland.

- [Tuxis Internet Engineering](#)
- [vBoxx](#)

Tuxus Internet Engineering

Tuxus biedt ons 2 mogelijkheden om Kerio Connect in te cloud te hosten. Er is een SaaS licentie die zelf nog zou gehost moeten worden en een cloud licentie. De cloud licentie is hier de logische keuze aangezien alles inbegrepen is voor een prijs van €5,25 per gebruikers per maand.

Features

- Altijd up-to-date
- 25 Gbyte opslag per gebruiker inbegrepen (voor 50 gebruikers is dat 1,25TB totale opslag)
- Dagelijkse back-ups
- Advanced anti-virus, advanced anti-SPAM, Exchange Active Sync en archivering

Een abonnement voor 50 gebruikers zou komen op een prijs van **€3150 per jaar**.

Vboxx

VBox is een andere partnerbedrijf aangeraden door GFI zelf. VBox biedt ons 3 programma's aan.

	Small	Big	Business
Prijs per gebruiker per maand	€5,00	€7,50	€6,50
Opslag per gebruiker	3GB	10GB	10GB

Het small programma heeft enorm weinig Storage. Het komt neer voor 50 gebruikers krijgen 150 GB opslag. Aangezien de mail server nu 700GB opslag heeft wil zeggen dat Dataline hier niet gaat mee toekomen.

Features

- Daily Back-ups
- SSL Beveiligd
- Spam Filter
- Anti-Virus

Voor een small programma voor 50 gebruikers zou het neer komen op **€3000 per jaar**. Een Business programma voor 50 gebruikers zou neer komen op **€3900 per jaar**.

Migratie

Aangezien deze opties ook gebruik maken van Kerio Connect zal de migratie makkelijker zijn en zullen alle mails, agenda en contacten ook gemigreerd worden. Kerio connect heeft een migration service om dit automatisch te gaan doen. [Hier](#) kun je meer info vinden hier omtrent.

Welke data wordt gemigreerd?

- Alle mailboxes gemaakt in Kerio Connect
- Alle emails, kalenders, contacten, taken, en notes
- Alle users email filters in Kerio Connect Client
- Public folders (calendars, contacts, tasks, en notes)

Welke data wordt NIET gemigreerd?

- Passwords
- Aliases
- Resources
- Mailing lists
- Server settings

Wachtwoorden worden niet gemigreerd en dat wil dus zeggen dat gebruikers een nieuw wachtwoord krijgen voor in de cloud! Er zal een CSV bestand gemaakt worden tijdens de migratie met alle wachtwoorden in.

Microsoft Office

Dataline heeft al een Office 365. Om de mailbox te activeren moet dan via Office 365 een licentie toegekend worden aan alle gebruikers die nood hebben aan een mailbox. Dit zou een goede oplossing kunnen zijn.

Pricing

	Prijs gebruiker/maand
Microsoft 365 Business Basic	€5,10
Microsoft 365-apps voor bedrijven	€8,80
Microsoft 365 Business Standard	€10,50
Microsoft 365 Business Premium	€18,90

Jammer genoeg heeft Microsoft niet echt een abonnement die enkel email bevat dus moeten alle extra features er ook bij genomen worden. Aangezien **Basic** het goedkoopste is en ons toegang heeft tot een mailbox van **50 GB per gebruiker** is dit de beste optie.

Voor 50 gebruikers zou ons dat **€3060 per jaar** kosten. Dit is voorlopig de beste prijs kwaliteit cloud oplossing aangezien er een heleboel features bijkomen.

Migratie

Aan de hand van een **IMAP migratie** kunnen de mailboxen overgezet worden naar Microsoft Outlook. Andere zaken zoals kalenders en contacten worden niet mee gemigreerd. [Hier](#) is een stappenplan van Microsoft om naar de Cloud te migreren via een IMAP enabled server.

Enkele dingen om rekening mee te houden:

- Enkel emails worden gemigreerd (geen notities, kalenders, ...)
- Maximum 500.000 items kunnen gemigreerd worden
- De grootste email die kan gemigreerd worden is 35MB

Conclusie

Als conclusie worden de verschillende prijzen en voordelen van alle opties even op een rij gezet.

Optie	Prijs per 50 gebruikers per jaar	Voordelen
Tuxus	€3150	Gebruikt Kerio in de cloud waardoor migratie simpeler zal zijn
Vboxx	€3900	Gebruikt Kerio in de cloud waardoor migratie simpeler zal zijn
Outlook	€3060	Goedkoopste, 50 GB storage + nog extra features, betrouwbaar

Hier lijkt verplaatsen naar Microsoft Outlook een logische keuze. Werknemers hebben nu al een Office 365 account waardoor de licentie toekennen simpel is. Het is ook de goedkoopste optie die nog extra features heeft zoals online web-versies van Word en Powerpoint.

Indien dat dit niet in het budget zou passen dan wordt de huidige on premise mail server best behouden.

Data Storage

Het beheren van data is zeer belangrijk voor elk bedrijf, data verlies kan leiden tot serieuze gevolgen en hoge kosten. Daarom moet het absoluut vermeden worden. Een goed beheer van data is nodig om de IT-infrastructuur in goede banen te leiden.

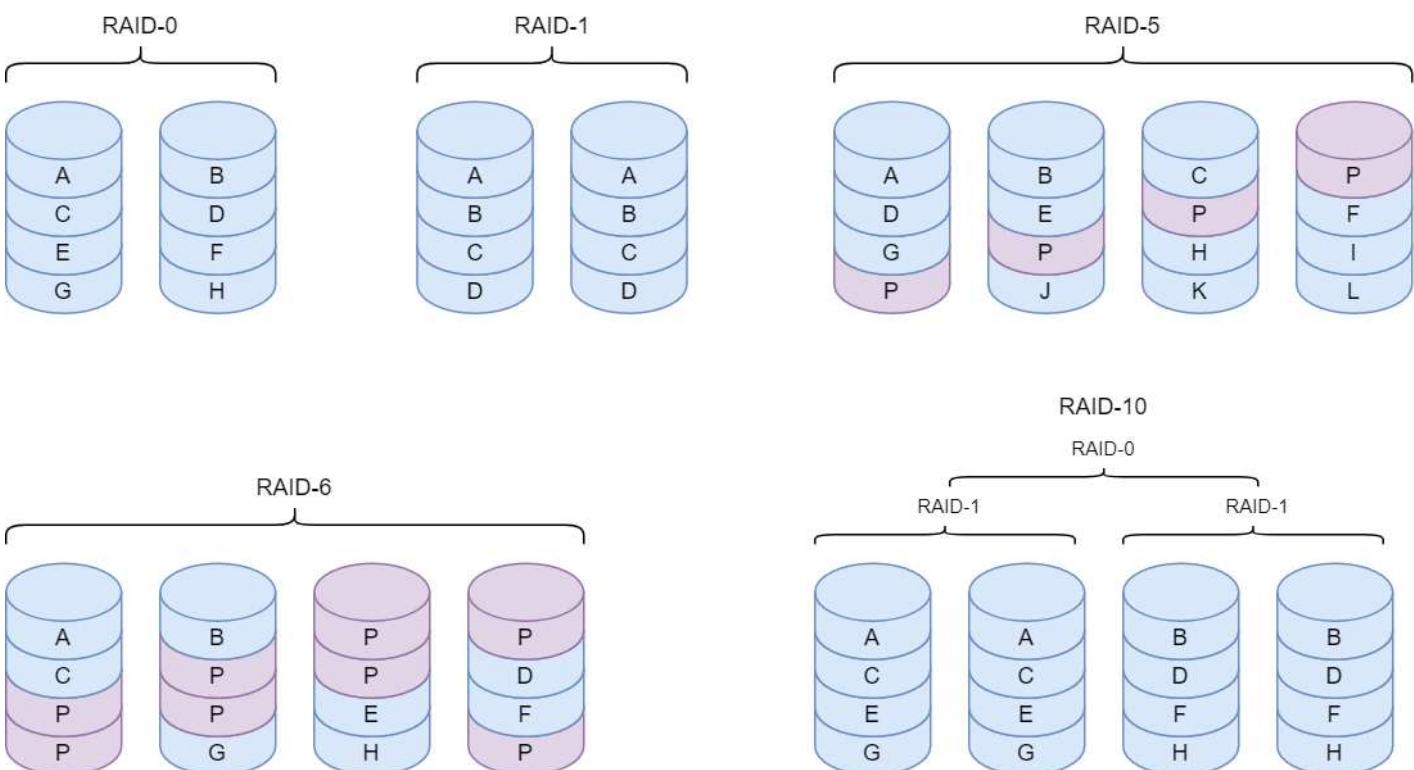
Methoden zoals RAID en backups worden gebruikt om data te gaan beveiligen, maar deze hebben ook hun nadelen. Zoals bijvoorbeeld de hersteltijd van RAID setups of het feit dat backups maken lastig kan zijn. Dit wordt later in dit hoofdstuk nog besproken.

Wanneer er gesproken wordt over data storage dan mogen virtuele machines niet vergeten worden. Dataline maakt gebruik van virtuele machines om verschillende applicaties en services uit te voeren. De data van de virtuele machines moet beschermd worden zodat de applicaties altijd up and running zijn.

RAID

RAID of redundant array of independent disks wordt gebruikt om harde schijven te gaan combineren in 1 enkel volume. Het doel ervan is om reads-/writes te versnellen of om data te beschermen. Dataline gebruikt RAID om data van hun servers te gaan beschermen en versnellen.

RAID Types



RAID 0

Deze manier wordt ook **striping** genoemd en dient om storage te gaan versnellen. De data wordt verspreid over 2 harde schijven en dit zorgt ervoor dat je data tegelijk van beide schijven kan ophalen. Merk op dat deze aanpak de kans op data verlies juist zal vergroten, aangezien er slechts 1 schijf moet falen om de data te verliezen. Deze methode wordt daarom enkel gebruikt voor niet belangrijke data die versneld moet worden.

RAID 1

Deze manier wordt ook **mirroring** genoemd. Er wordt een kopie opgeslagen van de data op 2 verschillende harde schijven. Dit zorgt ervoor als een enkel schijf faalt je nog steeds een kopie hebt van je data. Wanneer beide schijven falen zal je wel data verlies hebben.

RAID 5

Met RAID-5 zal de data verspreid worden over 3 of meer opslag apparaten. Naast de data worden ook pariteitsblokken opgeslagen. Deze pariteitsblokken zorgen dat wanneer er iets misloopt je de data kan herstellen. Deze aanpak neemt minder plek in dan RAID-1 maar geeft dezelfde bescherming. Aangezien de data zal verspreid worden over meerdere harde schijven zal data inlezen sneller gebeuren.

Een nadeel van RAID-5 is dat data schrijven traag zal zijn, omdat ook telkens de pariteit moet berekend worden.

RAID 6

Met RAID-6 wordt de data verspreid over 4 of meer opslag apparaten. Deze methode werkt op dezelfde manier als RAID-5 maar je zal 2 pariteitsblokken opslaan. Dit zorgt dat RAID-6 twee schijven kan verlezen. Aan de andere kant maakt dit het schrijven van data nog trager dan RAID-5 omdat nu 2 pariteitsblokken moeten berekent worden.

RAID 10

Deze methode combineert de voordelen van RAID-1 en RAID-0. Striping en mirroring wordt toegepast op de data. Hiervoor zijn er minstens 4 opslag apparaten nodig.

Software- vs Hardware RAID

Er zijn 2 manieren om RAID te gaan implementeren in Software of met hardware. Bij Hardware RAID zal er een extra fysieke component nodig zijn in de computer. Deze component wordt ook de **RAID controller** genoemd. Bij software is er geen controller nodig en zal het operating system (OS) RAID implementeren.

Hier worden de voor- en nadelen van beide methoden eens bekeken.

Software RAID	Hardware RAID
Komt samen met OS (goedkoper)	Heeft een RAID controller nodig (duurder)
Zet een last op CPU (trager)	Werkt onafhankelijk van de CPU (sneller)
Schijven zijn niet hot swappable	Schijven zijn wel hot swappable

Hot swappable wil zeggen dat schijven kunnen verwisseld worden zonder het systeem af te sluiten.

Failures To Tolerate

Een belangrijk aspect van elke RAID configuratie is hoeveel opslagplek ze innemen en hoeveel fouten ze kunnen tolereren. Het aantal schijven dat een RAID configuratie kan verliezen noemt men in het Engels ook **failures to tolerate (FTT)**.

Hieronder worden de FTT, de gegevensgrootte en de benodigde capaciteit vergeleken van de verschillende RAID configuraties.

RAID configuratie	FTT	Gegevensgrootte	Benodigde capaciteit
RAID 0	0	100 GB	100 GB
RAID 1	1	100 GB	200 GB
RAID 1	2	100 GB	300 GB
RAID 5	1	100 GB	133 GB
RAID 6	2	100 GB	150 GB
RAID 10	1 (soms 2)*	100 GB	200 GB

* Wanneer er 2 fouten gebeuren in hetzelfde RAID 1 paar heb je wel data verlies.

Virtuele Machines

Bij het gebruik van virtuele machines komt er ook storage kijken. Het beheren van je de data van virtuele machines kan een moeilijk process zijn en is enorm belangrijk om ervoor te zorgen dat je VM's altijd up and running zijn. Het programma die virtuele machines beheert op een server is gekend als een **Hypervisor**.

Een **Hypervisor** dient om meerdere besturingssystemen tegelijk op een computer te laten draaien. Hypervisors zijn opgedeeld in 2 types. Type 1 (Native) en Type 2 (Hosted).

Type 1

Een type 1 hypervisor draait rechtstreeks op de computer hardware en daarom wordt deze ook **Bare Metal** genoemd. Er is geen tussenkomst van het Besturingssysteem van de host. Dit wil zeggen dat Type 1 hypervisors efficiënt zijn om resources te gaan uitdelen aan de virtuele machines.

Enkele voorbeelden van type 1 hypervisors zijn: **VMware ESXi**, **Citrix Xen**, **KMV** en **Microsoft Hyper-V**.



Type 2

Een type 2 hypervisor zal niet rechtstreeks werken op de computer hardware. Er zit nog een Besturingssysteem tussen. Een voordeel van een type 2 hypervisor is dat het gemakkelijk te gebruiken is omdat het kan geïnstalleerd worden als een programma. Het nadeel hiervan is dat het minder efficiënt is aangezien er nog een besturingssysteem tussen zit.

Voorbeelden van type 2 hypervisors zijn: **Oracle VirtualBox**, **VMware Workstation**, **Parallels Desktop**.



Dataline maakt gebruik van 2 verschillende hypervisors om hun virtuele machines te laten draaien. Ze maken gebruik van KVM en ESXi, beide type 1 hypervisors. KVM is een gratis open source linux gebaseerde hypervisor terwijl ESXi een licentie nodig heeft om beschikking te hebben over de volledige functionaliteit.

ESXi wordt enkel gebruikt voor de telefonie servers. De reden hiervoor is omdat het bedrijf die de telefonie servers aanbied enkel werkt met VMWare ESXi. Dataline gebruikt een gratis licentie die beperkte functionaliteit heeft. Aangezien enkel de telefonie servers moeten werken met ESXi is een gratis licentie voldoende.

Storage van Virtuele Machines

Een probleem dat Dataline nu nog heeft op het vlak van storage is met de telefonie servers. Deze draaien allemaal op virtuele machines. Elke virtuele machine bestaat uit een aantal bestanden die de status van de machine voorstelt. Als beveiliging worden er back ups genomen van die bestanden. Deze aanpak heeft echter enkele nadelen:

- Er is een single point of failure in de telefonie servers. Als 1 iets kapot gaat kunnen mensen niet meer telefoneren.
- Back ups nemen is lastig
- Er zijn maar een paar mensen die weten hoe een virtuele machine herstelt kan worden van een back up.

Er moet een manier zijn om de bestanden van de VM te beveiligen tegen wanneer er iets misloopt. Dit moet een process zijn dat automatisch gebeurt.

SAN

Een SAN is een storage area network. Het is een apart netwerk speciaal gemaakt om de storage op een centrale plek op te kunnen slaan. Servers communiceren dan via dit netwerk om gebruik te maken van storage.

Dit is de ideale oplossing om de virtuele machines bestendig te maken tegen fouten. De reden hiervoor is omdat in een SAN data 2 maal kunnen opslaan. Wanneer er iets misloopt met de storage is er altijd nog een kopie van de data. Dit process kan automatisch gebeuren en er hoeft niemand manueel tussen te komen.

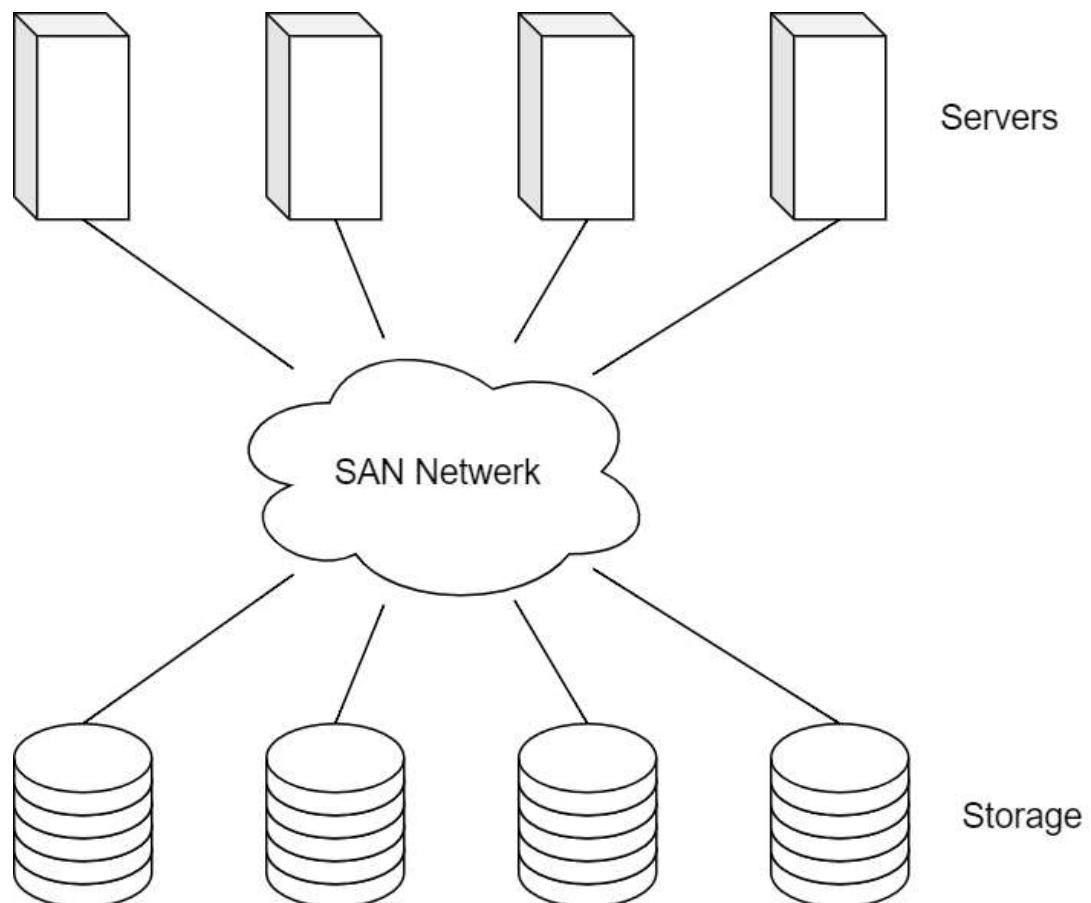
Deze oplossing wordt verder uitgediept in het volgende hoofdstuk.

SAN vs vSAN

Om de telefonie servers van Dataline fout tolerant te maken moet er een oplossing zijn die automatisch of zeer makkelijk virtuele machines kan beveiligen. Een SAN of vSAN wordt door veel bedrijven gebruikt om hun storage fout tolerant te gaan maken. In dit hoofdstuk worden beide oplossingen vergeleken en wordt de werking ervan uitgelegd.

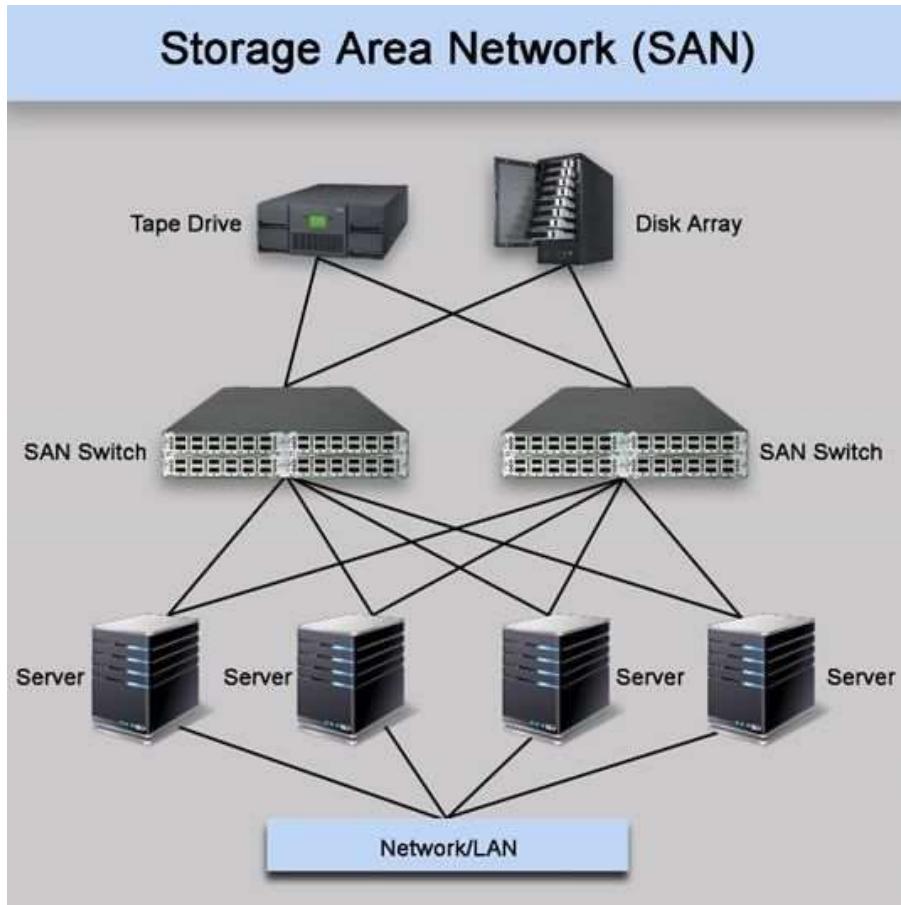
SAN

Een Storage Area Network (SAN) is een netwerk dat specifiek gebruikt wordt om grote hoeveelheden data te gaan opslaan. Een SAN bestaat uit 3 grote delen. De **servers**, het **netwerk** en de **storage**. De servers zijn allemaal verbonden via het netwerk. Het netwerk zelf bestaat uit een aantal switches die allemaal verbonden zijn met hoge snelheid verbindingen, dit zodat de data zo snel mogelijk kan doorgegeven worden naar de servers. De storage zal de data meerdere malen gaan opslaan voor redundancy.



SAN Netwerk

Het SAN netwerk uit hoge snelheid verbindingen en switches die worden verbonden met alles in het netwerk. Dit zorgt ervoor dat er voor elke verbinding een vervanging is, zo kan gelijk welke verbinding in de figuur verwijderd worden en alles zal nog bereikbaar zijn. SAN is enorm schaalbaar want er kan zoveel storage en servers gebruikt worden als nodig is. Deze kunnen makkelijk achteraf nog toegevoegd worden.



SAN Fiber

Een SAN maakt storage devices beschikbaar voor de servers over het netwerk. Er worden speciale protocollen gebruikt om IO operaties over een netwerk te sturen. Een SAN maakt altijd gebruik van Block storage, dit wil zeggen dat data geschreven en gelezen wordt in blokken. Er zijn 2 frequente protocollen die gebruikt worden voor een SAN.

Protocol	Beschrijving
Fibre Channel Protocol (FCP)	Dit protocol is de snelste optie en geeft verbindingen van 2 GBit/s tot 128 GBit/s. Werkt enkel op fiber verbindingen. Special netwerk kaarten zijn nodig om dit te ondersteunen.

Protocol	Beschrijving
Internet Small Computer System Interface (iSCSI)	Dit is een goedkoper alternatief dan FCP. Maar het is ook trager. Werkt op gewone ethernet verbindingen.

SAN Use Cases

Om beter te begrijpen waarom een SAN nodig zou zijn, worden hier enkele use cases overlopen van een SAN.

Use Case	Uitleg
Oracle databases	Worden veel gebruikt en hebben hoge performance + availability nodig
Microsoft SQL Server databases	Bevatten vaak kritische data dus hoge performance + availability nodig
Large virtualization deployments using VMware, KVM, or Microsoft Hyper-V	Meestal bevatten deze omgevingen vele verschillende virtuele machines. Vele verschillende besturingssystemen en applicaties draaien op deze virtuele servers. Daarom dat het belangrijk is dat de infrastructuur betrouwbaar is en fout tolerant. Want een enkele failure kan meerdere applicaties beïnvloeden.
Large virtual desktop infrastructures (VDIs)	Vele virtuele desktops kunnen lastig zijn om te beheren. Een SAN zorgt dat alle data gecentraliseerd is en makkelijker te beheren.

Voordelen

- SAN's zijn fout tolerant en verwijderen de single point of failure. Zelf als er een server, een switch of een storage unit zou weg vallen zijn er nog andere apparaten in het netwerk die de taak kunnen overnemen.
- Deze manier is ook enorm schaalbaar omdat er gemakkelijk meer storage en servers kunnen toegevoegd worden.

Nadelen

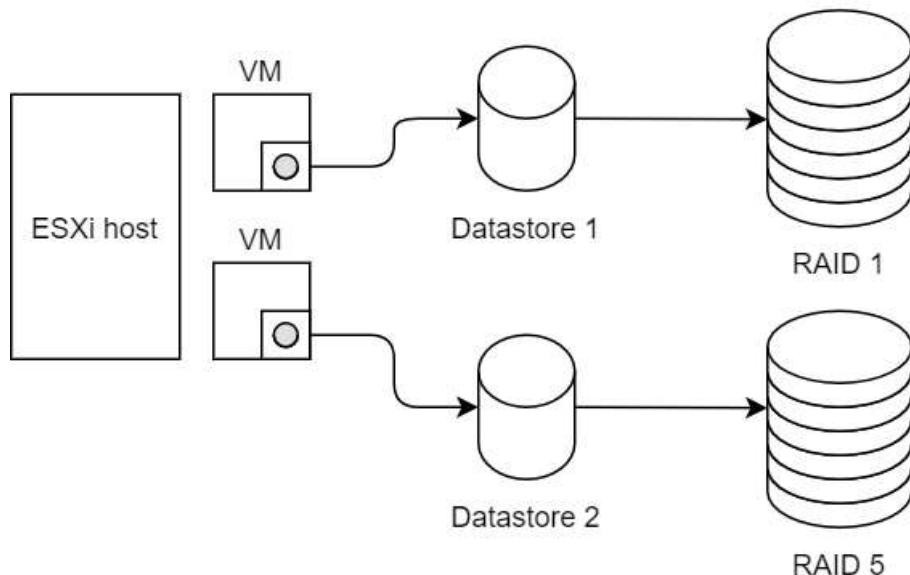
- Een fysieke SAN aanmaken kan enorm duur zijn.
- Het is complex om te realiseren

VSAN

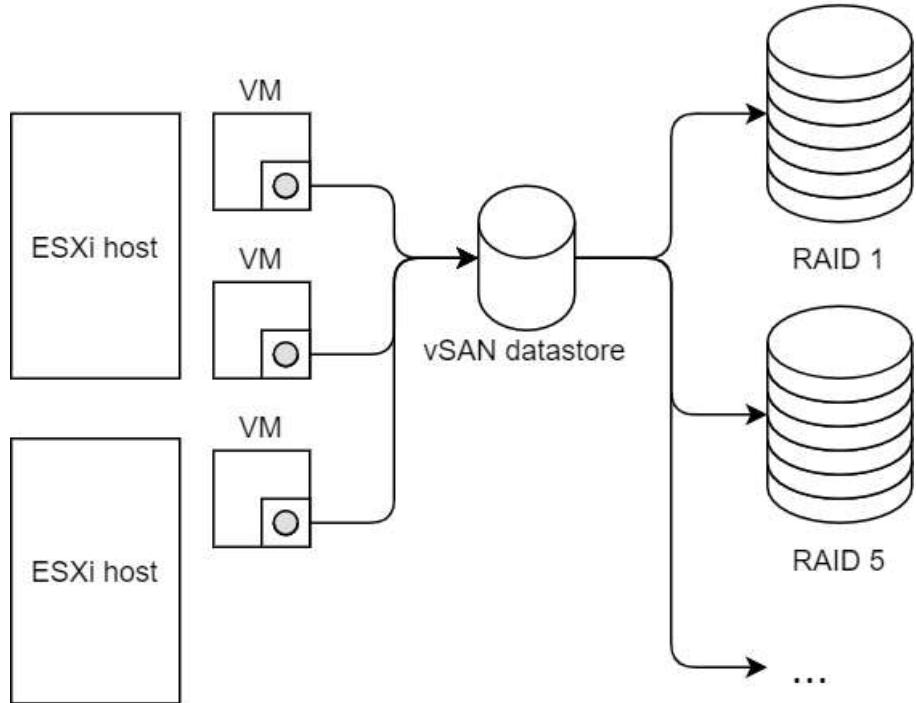
Een vSAN is een alternatief voor een gewone SAN. Het is een virtuele versie van een SAN. Een vSAN abstraheert de opslag en zorgt dat applicaties en virtuele machines toegang hebben tot een virtuele datastore. Het combineert de storage van verschillende virtuele servers tot **1 algemene datastore**.

Een vSAN is dus een soort **virtueel netwerk** die gebruikt wordt om de storage af te handelen. Het voordeel hiervan is dat de storage door software zal worden beheert en dat er meer vrijheid en opties zijn. Zo kan data bijvoorbeeld geabstraheerd en verdeeld worden over verschillende harde schijven. Zo kan fout tolerantie gerealiseerd worden met RAID 1, 5 of 6.

Soms wil een bedrijf verschillende data anders gaan behandelen. Bijvoorbeeld als er 2 virtuele machines zijn. Stel op VM1 is er simpele RAID 1 bescherming, terwijl de data op VM2 RAID 5 bescherming nodig heeft. Op de traditionele manier kan niet met 1 datastore gewerkt worden en zouden er 2 nodig zijn aangezien ze verschillende RAID methoden gebruiken.



Maar met vSAN kan de storage worden geabstraheerd zodat er een enkele centrale datastore is. In die datastore wordt er dan gezorgd dat de RAID 1 en 5 bescherming geïmplementeerd worden. Dit gebeurt automatisch en kan ingesteld worden naar de noden van de gebruiker.



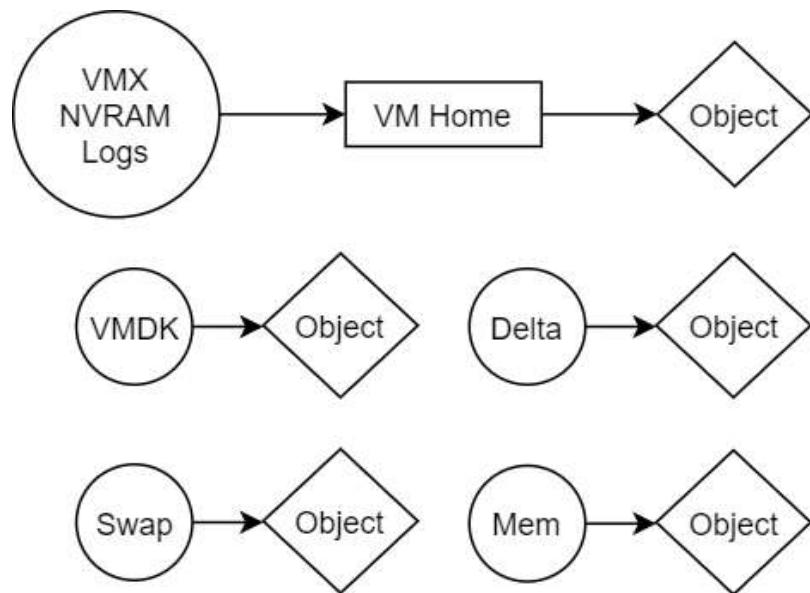
Er zijn een heleboel opties waaruit kan gekozen worden om vSAN te implementeren. Veel gebruikte opties zijn bijvoorbeeld: **VMWare vSAN**, **Starwind vSAN** en **Microsoft Storage Spaces Direct**. De voorkeur gaat naar VMWare aangezien deze gemakkelijk te gebruiken is, maar Starwind vSAN werkt ook met ESXi en is goedkoper. Microsoft Storage Spaces Direct (S2D) werkt dan weer met Hyper-V en Windows datacenter dus dit is geen oplossing.

VMWare vSAN

Alhoewel VMWare enorm duur is, zijn de oplossingen die ze bieden de beste en gemakkelijkste. Zelf nieuwe gebruikers leren zeer snel werken met VMWare ESXi. In tegenstelling tot KVM is VMWare de meest gebruiksvriendelijke oplossing. Daarom wordt hier in detail bekeken hoe VMWare vSAN in elkaar zit.

Objecten

VMWare vSAN zet files in de datastore als objecten die opgeslagen worden over de verschillende disk groups. Meestal wordt er voor elke file een object gemaakt. Kleine files worden soms ook samen gevoegd tot 1 enkel object. In onderstaande figuur worden de VMX, NVRAM en logs samen gevoegd tot 1 object. VMWare noemt deze files samen het VM Home object. Deze bevat allerlei meta data van de VM.



Disk Groups

Disk groups (DG) is een verzamel naam voor verschillende harde schijven. Elke ESXi host kan een maximum van **5 disk groups** hebben. Elke disk group bestaat dan weer uit maximum 8 storage devices. Er zal 1 storage device gebruikt worden als cache, de andere 7 kunnen gebruikt worden als **capacity**. De capacity bevat de eigenlijke data en de cache wordt gebruikt om reads en writes te versnellen. Er zijn 2 opstellingen voor disk groups.

Hybrid

Bij een hybride oplossing word er gebruikt gemaakt van SSD's en HDD's. Als cache word er een enkele SSD gebruikt, voor capacity worden enkel HDD's gebruikt. De cache wordt 70% gebruikt voor reads en 30% voor writes.

All Flash

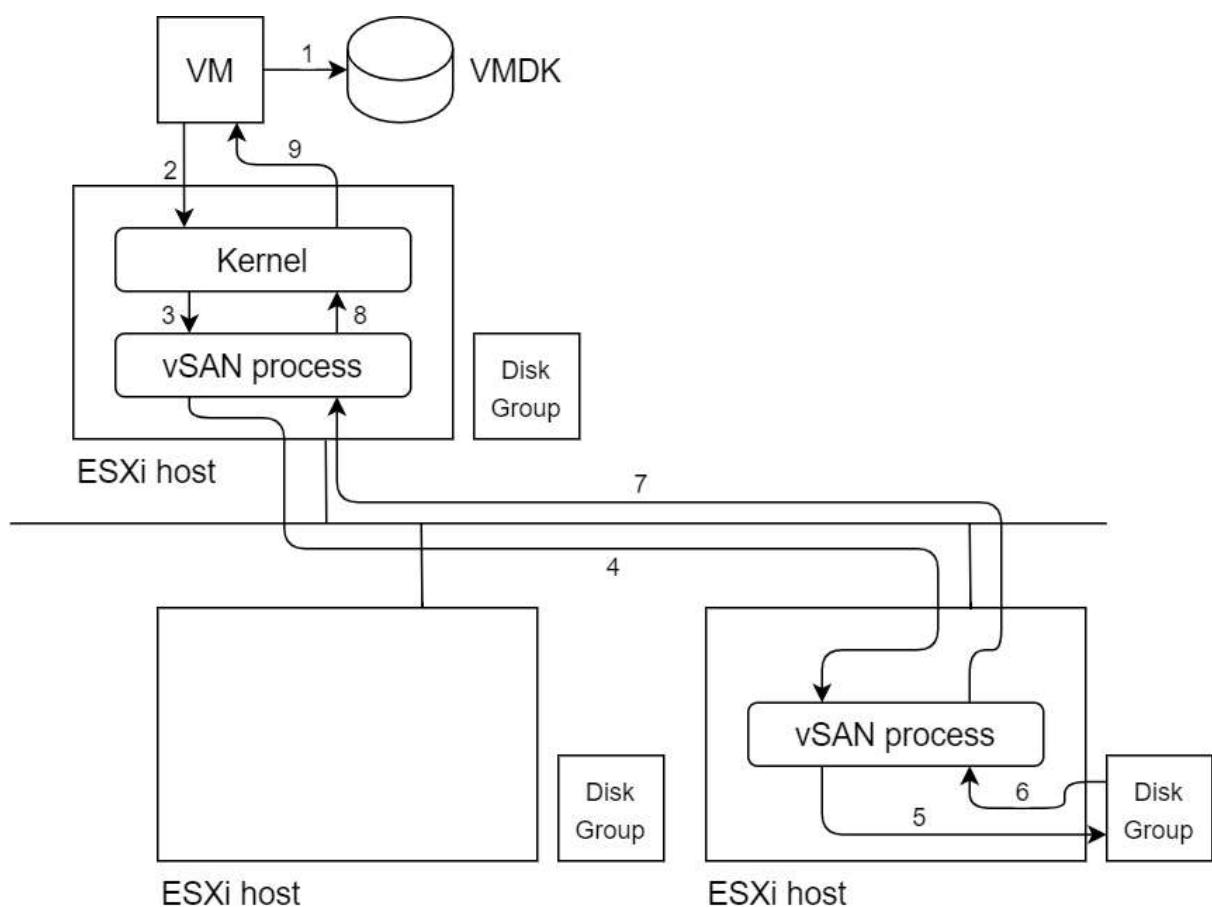
Bij een all flash oplossing word er enkel gebruikt gemaakt van SSD's. Dezelfde verdeling geldt als bij een hybride disk group alleen word de cache nu enkel nog maar gebruikt om te schrijven naar de capacity. De cache wordt dus enkel gebruikt voor writes in een all flash set up.

Data placement

Het is belangrijk om te weten hoe de data wordt verdeeld over verschillende ESXi hosts. Hier worden de verschillende stappen overlopen die gebeuren wanneer er data ingelezen of geschreven wordt.

Stappen:

1. VM stuurt een write naar de VMDK file
2. Die write word opgevangen door de Kernel
3. De kernel roept dan het vSAN process op om de write af te handelen
4. Het vSAN process weet waar in het netwerk de write naar toe moet en stuurt het naar de juiste ESXi host
5. Het vSAN process van de andere host stuurt de write dan naar de Disk Group van de host zelf en past die toe
6. De disk group reageert met een success of fail acknowledgement
7. Het vSAN process stuurt die acknowledgement door naar de originele host
8. Het process licht de kernel in
9. De VM krijgt de acknowledgement van de kernel wat evenwaardig is als een acknowledgement van de VMDK



Waar precies het vSAN process de reads en writes naar toe stuurt hangt er van af. Het vSAN process zal de blokken data van de VMDK file gaan verspreiden over het netwerk. Het vSAN process onthoudt waar alles bewaard wordt. Zodat wanneer een read of write zich voort doet het process weet naar welke host het moet sturen.

Failures to tolerate

Failures To Tolerate of FTT staat voor het aantal fouten die zich mogen voordoen in een vSAN. Wanneer bijvoorbeeld de FTT gelijk is aan 1 dan wil dat zeggen dat zelf als er 1 fysieke component weg valt de virtuele machine nog steeds toegang heeft tot die component.

Hoe zou de data verspreid worden over het netwerk als er een FTT gelijk aan 1 zou zijn? Een RAID 1 methode zou kunnen gebruikt worden. Dit wil zeggen dat de data 2 maal wordt opgeslagen op verschillende locaties. Dit wordt ook een **2 node setup** genoemd aangezien er op 2 plekken data wordt opgeslagen. Voor een 2 node setup hebben ook nog een derde host nodig die de Witness zal zijn.

Witness

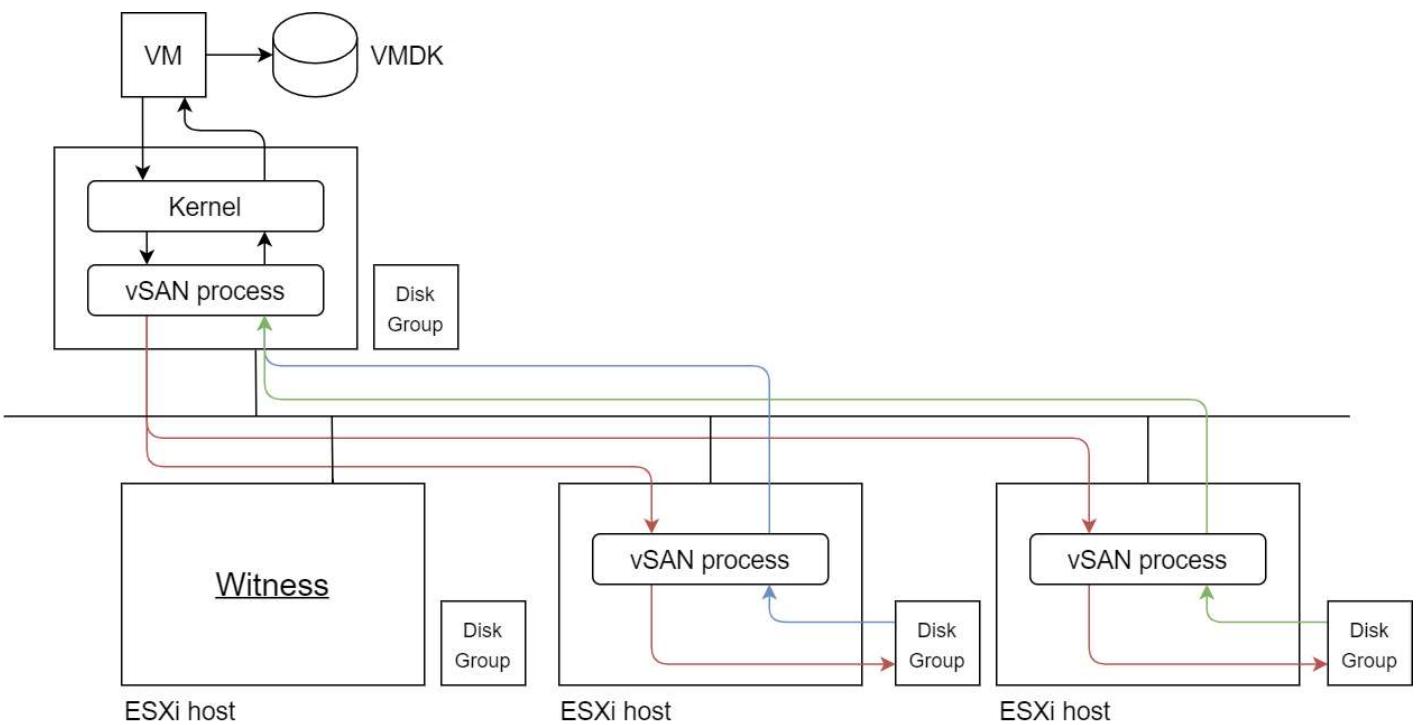
Waarom is de witness nodig? Omdat anders kan het zijn dat de 2 hosts niet gesynchroniseerd zijn en de verschillen bij elkaar niet zien. Dan is er een **split brain** situatie. In zo een situatie weet het vSAN process niet meer welke van de 2 hosts nu de correcte data bevat. Daarom moet er een derde host zijn die de verschillen bijhoudt van beide om zo de beste beslissing te kunnen maken.

De witness zelf slaat geen data op van de virtuele machine maar zal enkel metadata gaan opslaan. Deze data wordt ook de **Witness Component** genoemd in een vSAN en is ongeveer 4 MB groot.

Schema

Op de afbeelding kun worden verschillende hosts voorgesteld om te zien en wat er precies gebeurt als de VM wil schrijven of lezen naar de VMDK file wanneer deze een FTT hebben van 1. Het is gelijkaardig met de afbeelding van data placement maar er zijn nog een paar kleine verschillen.

1. De stappen van de VM tot het vSAN process zijn identiek
2. Het vSAN process stuurt de write of read door naar **alle hosts** die de data bevatten.
3. Elke host stuurt dan een acknowledgement terug naar de host die de VM bevat
4. Pas wanneer elke host terug stuurt zal het vSAN process de acknowledgement doorgeven aan de kernel



Pricing

Nu één van de belangrijkste aspecten is de prijs van de oplossing. Als er gekeken wordt naar de prijs van VMWare vSAN dan valt het op dat de prijzen enorm hoog zijn. VMWare is een bedrijf die zich meer zal focussen op grotere bedrijven die honderden virtuele machines moeten beheren. VMWare vSAN wordt voorlopig uitgesloten omdat de prijs gewoon veel te hoog voor zo'n klein bedrijf als Dataline.

Starwind vSAN aan de andere kant is meer gericht op kleinere bedrijven die niet te veel willen betalen voor een vSAN oplossing. Daarom zou Starwind vSAN een ideaal alternatief zijn. Maar er moet niet alleen gekeken worden naar prijs, kwaliteit is ook belangrijk. VMWare is een zeer gekende speler op het vlak van vSAN, maar Starwind is minder bekend. Daarom is het belangrijk dat de Starwind vSAN eerst getest wordt op een omgeving om te kijken hoe goed het nu eigenlijk echt werkt.

In volgend hoofdstuk wordt dieper ingegaan op de werking van Starwind vSAN en wordt een test omgeving opgesteld.

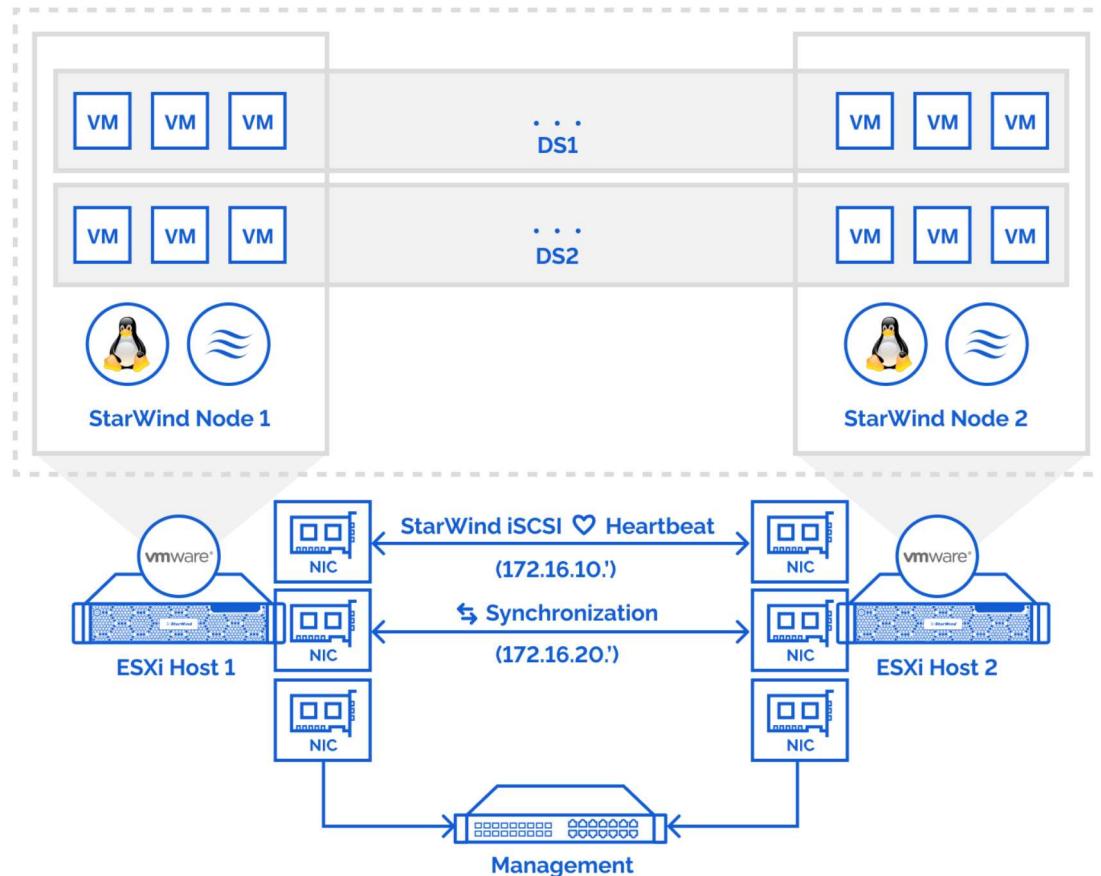
Test Starwind vSAN

Starwind vSAN is een bekende speler op de markt van Hyper Converged Infrastructure. Hun oplossing voor vSAN zou ideaal zijn voor Dataline aangezien er geen dure VMWare licenties nodig zijn. Maar de vraag is natuurlijk hoe goed is Starwind vSAN? Als eens gekeken wordt naar de reviews dan zijn er veel bedrijven tevreden met Starwind vSAN. Om te zijn van ons stuk wordt een test opstelling opgezet, hier wordt de performance, efficiëntie en kwaliteit van Starwind vSAN onder de loep genomen.

Opstelling

Als opstelling worden 2 gebruikte desktop computers die Dataline nog liggen had gebruikt. Die breiden worden dan uitgebreid met een gloed nieuwe snelle netwerkkaart (NIC) met 2 poorten (25 Gbps) en een 350 GB SSD.

Starwind heeft een 30 day free trial waarmee er een 2 node setup kan getest worden. Twee nodes is het absolute minimum dat nodig is om een vSAN te gebruiken. Met een node wordt hier een server bedoeld die de Starwind virtuele machine aan het draaien is. De setup die zal gevuld worden is de volgende.



In totaal zullen er 3 verbindingen nodig zijn op de nodes:

- iSCSI/Heartbeat
- Synchronisatie
- Management

Het iSCSI/heartbeat en synchronisatie kanaal wordt gebruikt om data en commando's door te geven tussen de hosts. Deze verbindingen gebeuren best met de nieuwe netwerkkaart zodat de data zo snel mogelijk kan gesynchroniseerd worden. Management is bedoeld als aanspreekpunt voor het beheren van de ESXi host en Virtuele machines.

Failover strategie

Bij het geval dat de verbinding tussen 2 nodes zou weg vallen door een netwerk problemen dan hebben is er een zogenaamde netwerk partitie, dit zorgt dat de nodes niet meer kunnen communiceren met elkaar en kan leiden tot een **split brain** scenario. Split brain kan zeer erge gevolgen hebben zoals het verliezen van data en het moet absoluut vermeden worden.

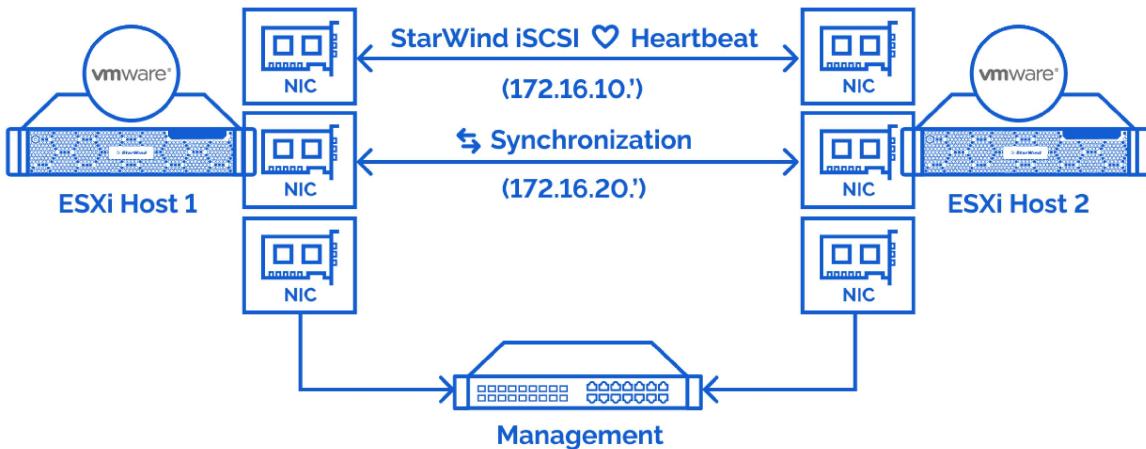


In deze situatie zal elke node naar zijn eigen storage schrijven. Zo ontstaan er data inconsistenties tussen de 2 nodes waardoor ze steeds meer van elkaar gaan verschillen. Uiteindelijk gaan ze zoveel met elkaar verschillen dat de data niet meer gesynchroniseerd kan worden.

Als moet hersteld worden dan zal een node moeten gekozen worden om verder op te werken. Alle wijzigingen van de andere node worden daardoor ongedaan gemaakt. Het risico van een split brain scenario moet daarom zo laag mogelijk zijn. Starwind vSAN geeft ons 2 manieren om zo'n scenario te vermijden.

Heartbeat

Er wordt een zogenaamde heartbeat verbinding opgesteld tussen nodes. Wanneer een node merkt dat de synchronisatie verbinding niet meer werkt tussen een partner node dan zal er een ping gestuurd worden via het heartbeat kanaal. Als de partner node antwoord dan zal Starwind de node blokkeren met een lagere prioriteit tot de synchronisatie verbinding terugkomt. Als de partner node niet antwoord dan gaat Starwind vSAN er vanuit dat die offline is. Starwind vSAN markeert de partner node dan als niet gesynchroniseerd.



Stel bijvoorbeeld dat er 2 nodes zijn zoals in de figuur en de data kan niet verzonden worden via de synchronisatie verbinding. Dan kunnen de 2 nodes niet gaan synchroniseren met elkaar. Via het heartbeat kanaal checkt de primary node de status van de 2 systemen. Het ziet dat beide systemen nog in sync zijn en zal de secundaire node blokkeren zodat het niet meer reageert op requests. De secundaire node check regelmatig de verbinding van het synchronisatie kanaal. Vanaf dat er terug verbinding is, zal het synchronisatie process terug starten en wordt de secundaire node weer actief.

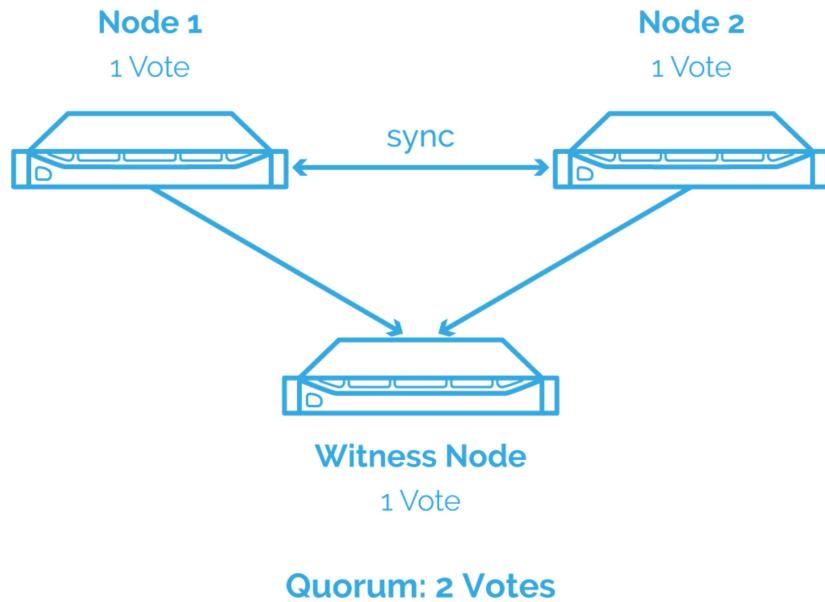
Nadeel:

Indien alle verbindingen (heartbeat + synchronisatie) verbroken worden tussen de 2 hosts dan zal er een split brain situatie ontstaan. Starwind vSAN probeert dit probleem te minimaliseren door toe te laten dat de heartbeat verbinding over het management netwerk gaat. Starwind raad deze aanpak enkel aan wanneer er meerdere heartbeat verbindingen zijn.

To summarize, this kind of strategy is mostly applicable to the systems where you have enough network links that can be used as the additional heartbeat channels and are physically separated from the primary ones. **Starwind** ↗

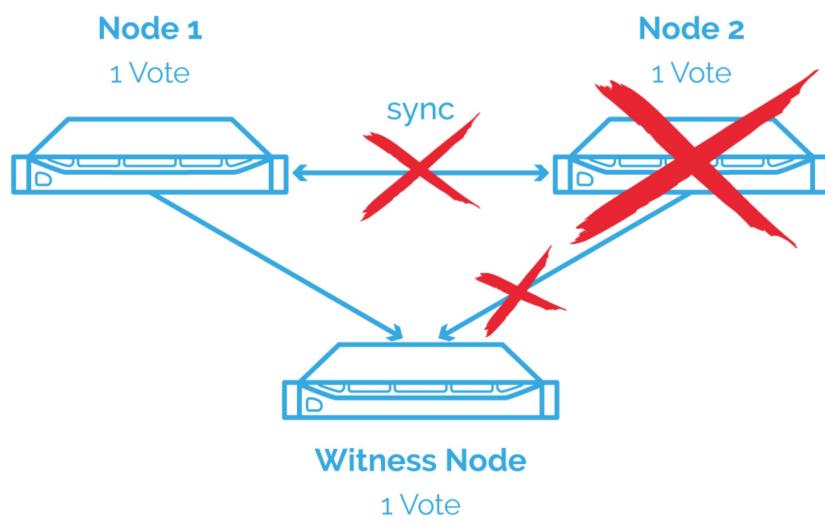
Node majority

Een andere manier om split brain te gaan vermijden is door een **Witness node** te gaan toevoegen. Het is een gekende strategie en wordt ook door VMWare gebruikt. De witness houd informatie bij van beide nodes en wanneer de synchronisatie zou weg vallen dan zal de witness mee beslissen over welke node de primaire node is.



Om te beslissen wie de primaire node wordt zal er een stemming gedaan worden. De node met de meeste stemmen wordt dan de primaire node. Elke node zal voor zichzelf stemmen waardoor een meerderheid van 2 stemmen nodig is. Daarom moet er een derde node om de knoop door te hakken. De witness zelf houd gewoon meta data bij en heeft dus zelf niet veel storage en resources nodig.

Cluster still working



Stel bovenstaande configuratie wordt gebruikt en de synchronisatie valt weg. Dan zal de witness node die verbonden is met beide nodes, een meerderheid vormen met de node die de meest relevante data heeft (in dit voorbeeld node 1). Daardoor zal node 2 gemarkerd worden als niet gesynchroniseerd en zal die niet meer luisteren naar requests. Het grootste voordeel van deze aanpak is dat split brain zich niet meer kan voordoen, maar er is wel een extra node nodig.

Voordelen

- Het split brain scenario is volledig uitgesloten
- Een extra heartbeat verbinding is niet nodig

Nadelen

- Bij 2 node setup is een derde witness node nodig
- met 3 nodes mag er maar 1 failure voorkomen

Synchronisatie test

Om te kijken of de synchronisatie tussen de nodes gebeurt moeten er eens gecontroleerd worden als de er degelijk gerepliceerd wordt over de nodes. Met een simpele test wordt dit gecontroleerd.

1. Maak een bestand aan op één node.
2. Sluiten de vm af
3. Start dezelfde VM op de andere node
4. Kijk of dit bestand ook op de andere node zich bevindt

Dit lukt zonder problemen en alles werkt zoals verwacht.

Fio

De manier waarop een workload zal gesimuleerd worden is door gebruik te maken van een Disk IO test tool genaamd **fio**. Het is een command line tool die zeer veel parameters heeft om te gaan testen. De gebruikte parameters zijn gebaseerd op een artikel van Oracle over het testen van Block storage [Oracle](#).

Met fio worden IO operaties uitgevoerd om statistieken te verzamelen zoals:

- Het aantal IOPS (Input/Output operaties per seconde)
- De bandbreedte van de data stroom (in MB/s of GB/s)
- Hoe snel de storage antwoordt op IO requests (latency)

Het aantal IOPS is een zeer belangrijke statistiek. Databases zullen zeer veel kleine reads en writes gaan uitvoeren op een storage device. Daarom moet een database server best zo hoog mogelijk aantal IO operaties kunnen uitvoeren om efficient vele requests te kunnen vervolledigen.

Om de verschillende scenario's van de storage te controleren gaan er 4 soorten testen uitgevoerd worden:

- Random reads
- File random reads/writes
- Random read/writes
- Sequentiële reads

Alle testen worden rechtstreeks op de storage uitgevoerd behalve de file random reads/writes. Deze test zal IO operaties doen op een file. Het verschil met rechtstreeks werken is dat hier het besturingssysteem nog tussen komt. Er wordt dus verwacht dat deze test iets trager zullen zijn.

Performance testen

Storage is de traagste factor van elke computer, daarom is de performance van storage zeer belangrijk voor virtuele machines. Om te kijken of Starwind vSAN een goede optie zou zijn, moeten er gecontroleerd worden hoe efficient Starwind omgaat met storage. Een heleboel factoren hebben invloed hebben op de performance van vSAN, dus het is belangrijk om verschillende opstellingen te testen en te kijken wat het beste optie is

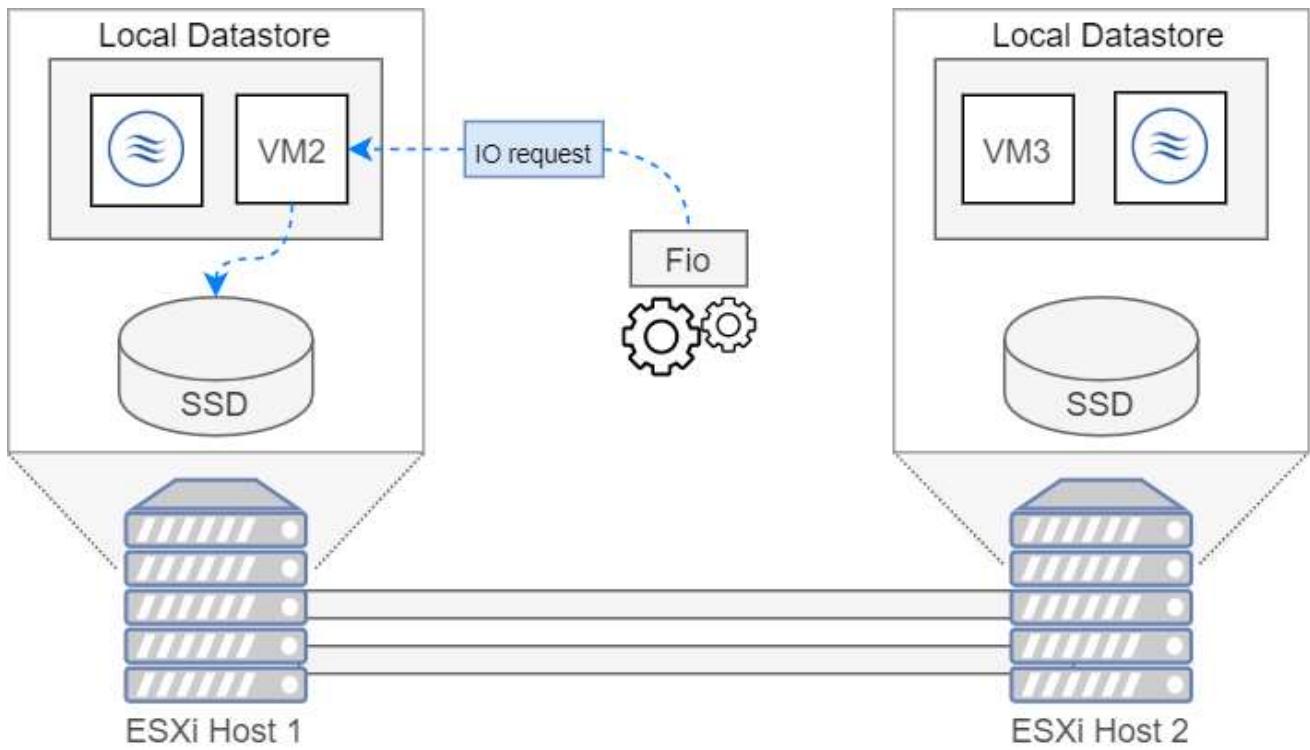
Er zijn 3 scenario's die interessant zijn om te testen voor performance:

- VM op lokale datastore
- VM op vSAN datastore
- VM moet werken via iSCSI omdat lokale host niet gesynchroniseerd is.

De laatste test is belangrijk want deze stelt het scenario voor dat een lokaal storage device niet meer beschikbaar zou zijn. De virtuele machine zal dan over het netwerk IO operaties doen met behulp van iSCSI. Op deze manier zal de virtuele machine nooit down komen te staan.

Lokale Datastore

Er wordt een vm op de lokale datastore opgestart. Met deze opstelling zal de bandbreedte van IO operaties gelijk moeten zijn aan de snelheid van een SATA SSD wat ongeveer 350MB/s zou moeten zijn. In onderstaande figuur wordt het pad dat een IO request volgt eens gevisualiseerd.



Resultaten

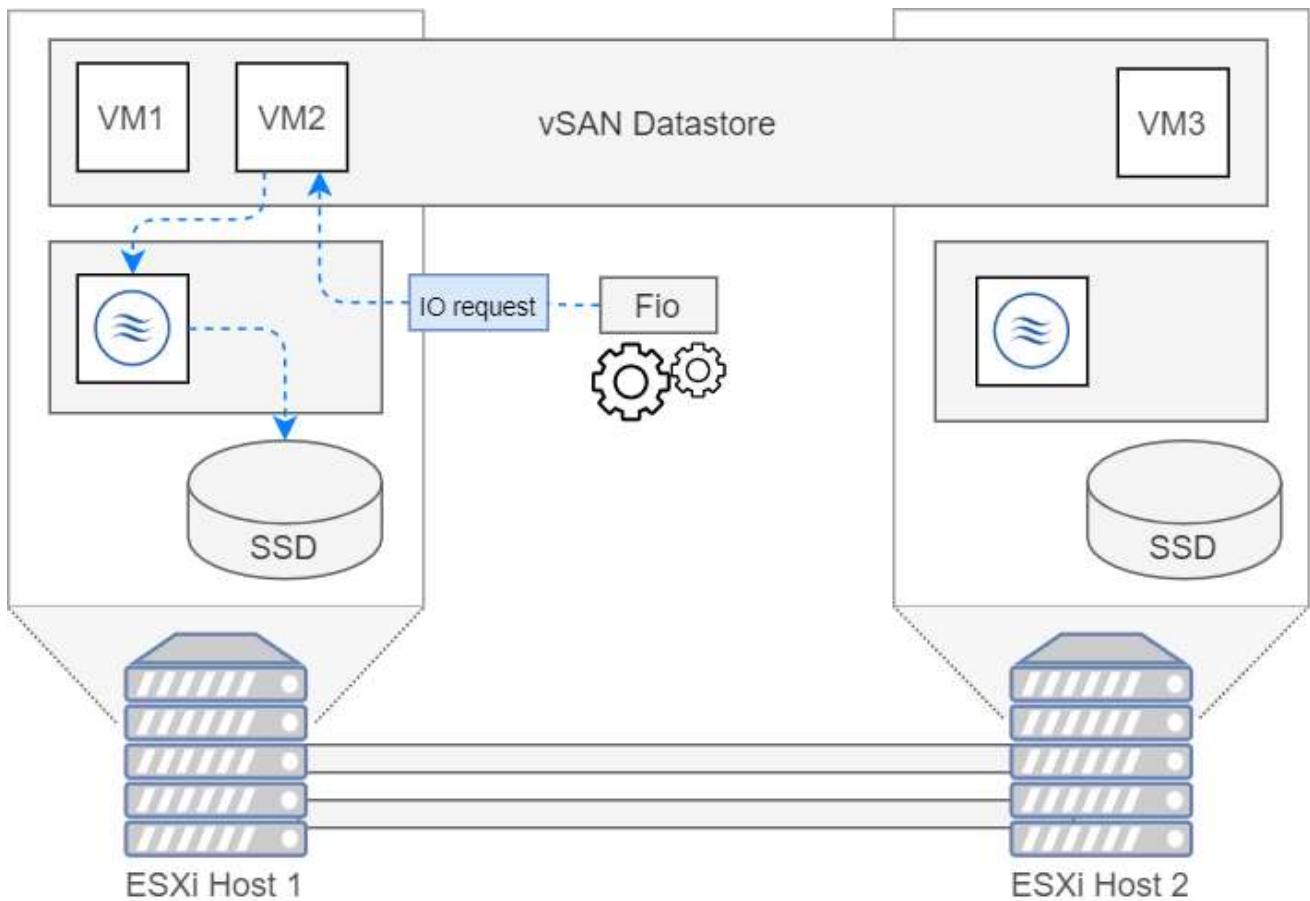
IOPS performance test	IOPS (reads)	IOPS (writes)
Random Reads	285k	/
Random reads/writes	70,9k	70,9k
File random reads/writes	69,4k	69,3k
Sequential reads	292k	/

Throughput Performance Tests	BW (Read)	BW (Write)
Random Reads	2468 MB/s	/
Random reads/writes	347 MB/s	348 MB/s
File random reads/writes	319 MB/s	319 MB/s
Sequential reads	2698 MB/s	/

Latency Performance Tests	Tijd (µs)	
Random Reads	45,6 µs	/
Random reads/writes	109,69 µs	51,20 µs

vSAN Datastore

Er wordt een vm op de vSAN datastore opgestart. Met deze opstelling zal nog steeds gebruik gemaakt worden van de lokale SSD maar dan via de vSAN datastore. Dit zou enige overhead moeten creëren waardoor er wat lagere resultaten zullen behaalt worden. In onderstaande figuur wordt het pad dat een IO request volgt eens gevisualiseerd.



Resultaten

IOPS performance test	IOPS (reads)	IOPS (writes)
Random Reads	202k	/
Random reads/writes	26,3k	26,3k
File random reads/writes	25,7k	25,7k
Sequential reads	251k	/

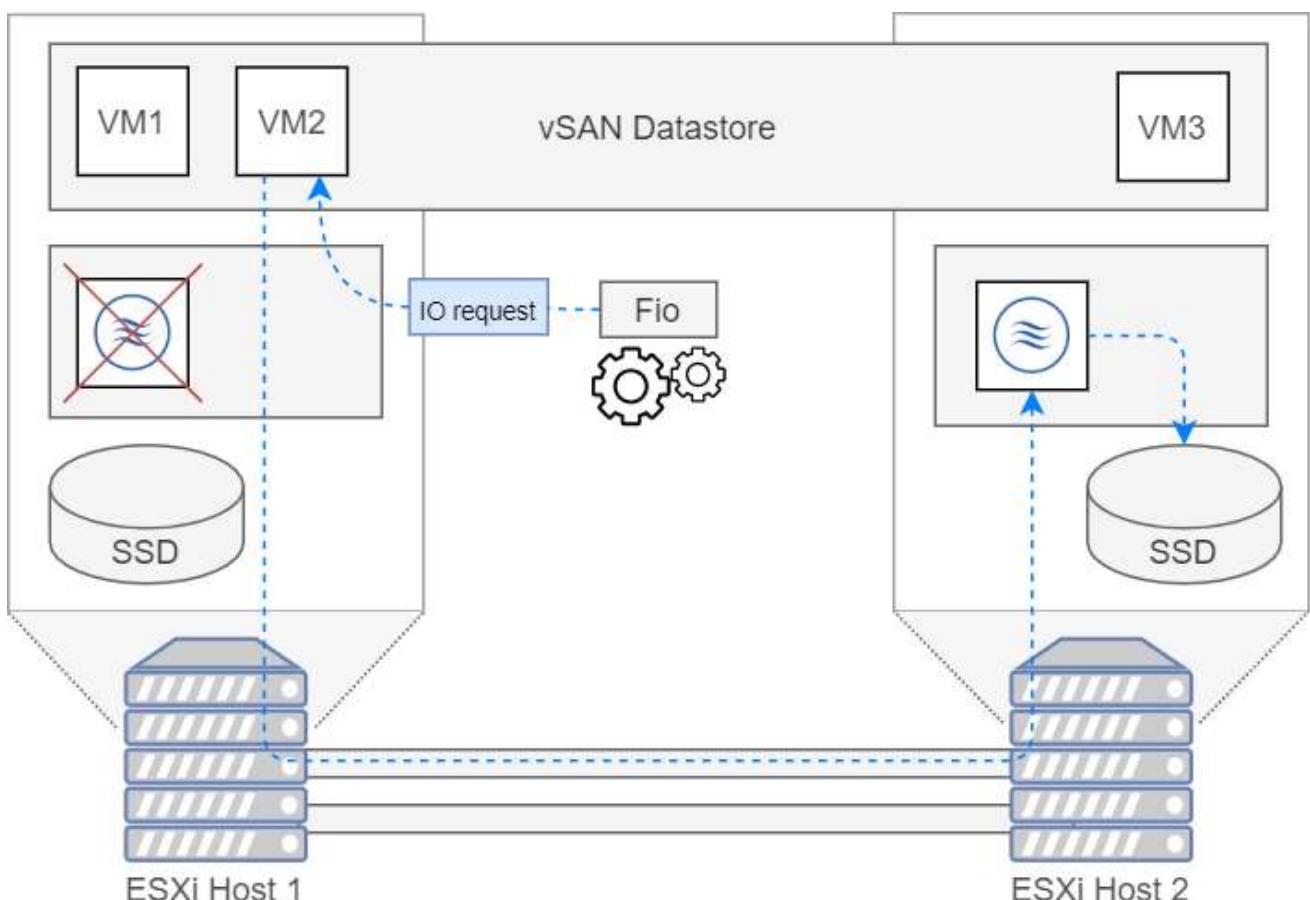
Throughput Performance Tests	BW (Read)	BW (Write)
Random Reads	2622 MB/s	/
Random reads/writes	289 MB/s	290 MB/s

Throughput Performance Tests	BW (Read)	BW (Write)
File random reads/writes	271 MB/s	300 MB/s
Sequential reads	2210 MB/s	/

Latency Performance Tests	Tijd (Read)	Tijd (Write)
Random Reads	109,12 µs	/
Random reads/writes	266,03 µs	406,39 µs

Via iSCSI verbinding

Deze keer wordt de Starwind vSAN node afgesloten op de host. Dit zorgt dat de vm geen toegang meer zal hebben tot de lokale SSD. De vm moet nu via de vSAN datastore IO operaties gaan uitvoeren op de andere Starwind node. Het zal dit doen aan de hand van iSCSI commando's. De iSCSI verbinding gaat over de nieuwe snelle netwerk kaart (25 Gbps), hierdoor zou er niet zo'n groot verschil mogen zijn ten opzichte van de vSAN datastore.



Resultaten

IOPS performance test	IOPS (reads)	IOPS (writes)
Random Reads	222k	/
Random reads/writes	38,8k	38,8k
File random reads/writes	35,2k	35,2k
Sequential reads	265k	/

Throughput Performance Tests	BW (Read)	BW (Write)
Random Reads	3226 MB/s	/
Random reads/writes	276 MB/s	277 MB/s
File random reads/writes	329 MB/s	330 MB/s
Sequential reads	2662 MB/s	/

Latency Performance Tests	Tijd (µs)	
Random Reads	110,83 µs	/
Random reads/writes	267,58 µs	212,2 µs

Conclusie

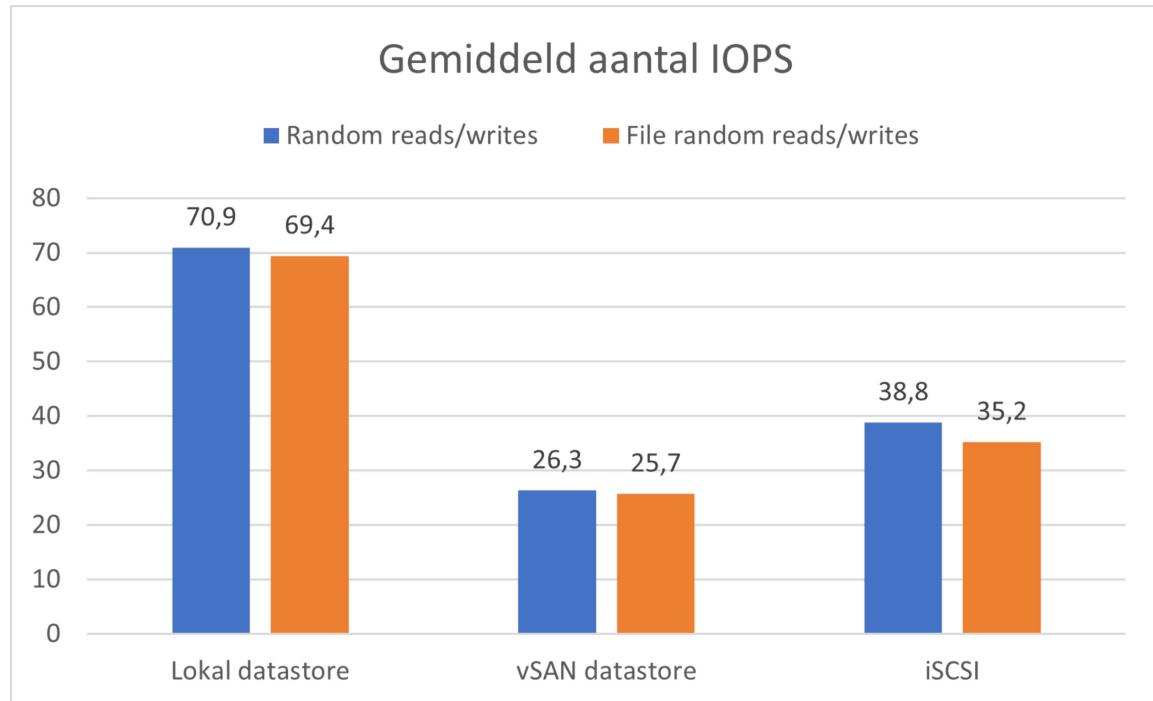
Om uit deze testen een conclusie te halen worden de resultaten eens op een rij gezet.

Opmerking

In de testen hierboven zijn de resultaten van de **random reads** en **sequential reads** normaal groot. Zo groot zelf dat ze sneller zijn dan de SSD. Dit is mogelijk te wijten aan caching van de storage. Deze resultaten zullen om die reden niet meegerekend worden in de conclusie.

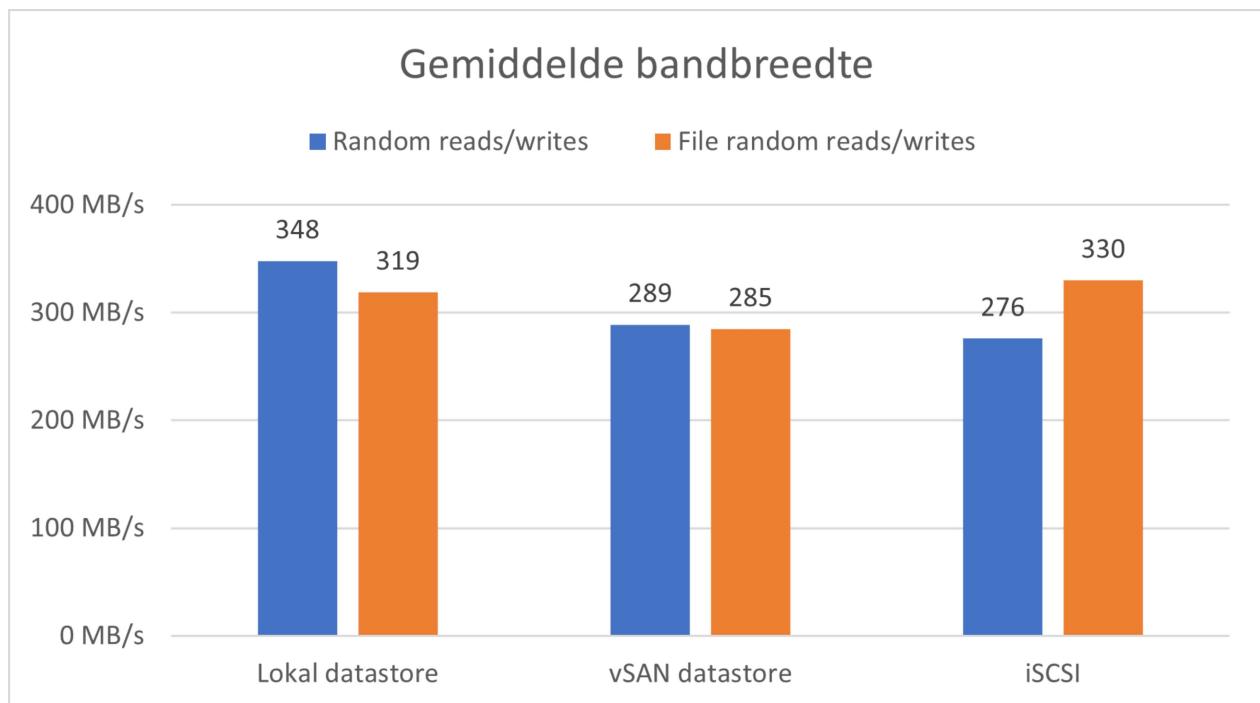
IOPS

In de onderstaande grafiek wordt gekeken naar het gemiddelde aantal IOPS voor elk scenario. Hier kan er gezien worden dat het aantal IOPS duidelijk hoger zal zijn wanneer een lokale datastore wordt gebruikt. Dit is logisch aangezien dat met een lokale datastore de IO requests sneller de SSD zullen bereiken.



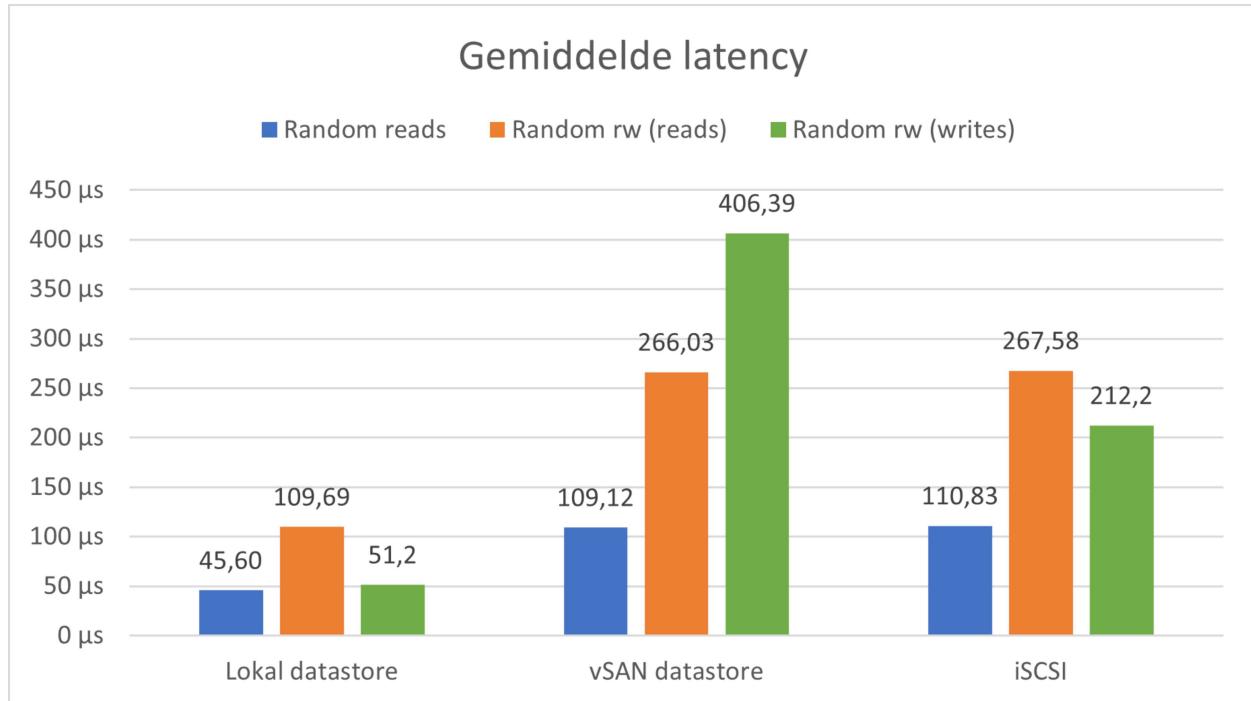
Bandbreedte

Hier wordt gekeken naar de gemiddelde bandbreedte voor elk scenario. Het valt op dat er weinig verschil is tussen de verschillende methoden. Dit is te wijten aan het feit dat voor de bandbreedte test grotere blokken data worden opgevraagd. Dit zorgt dat er minder IO operaties moeten gebeuren waardoor het verschil tussen de 3 minder merkbaar is.



Latency

In de onderstaande grafiek wordt gekeken naar de tijd dat de SSD ertover doet om een IO request te voldoen. Er is een duidelijk verschil namelijk dat requests op de lokale datastore sneller worden beantwoord. Net zoals bij de IOPS zullen requests sneller de SSD bereiken met een lokale datastore.



Conclusie

Voor deze proef werd gekeken naar de huidige IT-omgeving van Dataline Solutions. Applicaties zoals Confluence, Jira, email en de telefonie servers werden onder de loep genomen en er werd van elke applicatie de mogelijke oplossingen opgesomd.

Met mijn onderzoek zijn volgende conclusies behaald:

- Active Directory is klaar gemaakt voor toekomstige synchronisatie met de cloud
- Synchronisatie van Azure werd uitgevoerd
- De permissies en groepen werden opgekuist en klaar gemaakt voor synchronisatie met de cloud
- Confluence moet naar de cloud en een standaard abonnement is genoeg voor Dataline
- De pocketquery plugin van Confluence moet bekijken worden zodat de queries ook werken in de cloud
- Om gebruik te maken van Azure AD users en groepen is een Atlassian Access abonnement nodig
- Indien er budget is voor de mail server wordt er gewerkt met een Outlook mailbox en anders blijft er lokaal gewerkt worden
- Een oplossing voor de telefonie servers is bepaalt namelijk Starwind vSAN
- Getest of Starwind vSAN is een goed alternatief is voor VMWare vSAN
- Gebruik van Starwind vSAN heeft een duidelijke invloed op het aantal IOPS en de latency van IO requests

Het doel van deze proef was om de huidige IT-infrastructuur van Dataline klaar te maken voor vernieuwingen en te kijken welke applicaties naar de cloud moeten. Het doel is gedeeltelijk behaald, er werd gekeken voor een aantal applicaties maar niet voor ze allemaal. Zoals het beheren van de KVM virtuele machines, de file server en de backup server.

Afkortingen

Onderstaande tabel geeft een alfabetische lijst terug met alle afkortingen die in deze verhandeling aan bod komen.

Afkorting	Volledige vorm
AD	Active Directory
CPU	Central Processing Unit
DC	domeincontroller
DG	Disk Group
FCP	Fiber Channel Protocol
FTT	Failures To Tolerate
iSCSI	Internet Small Computer System Interface
NIC	Network Interface Card
OS	Operating System
OU	Organizational Unit
RAID	Redundant Array of Inexpensive Disks
RAM	Random Access Memory
SAN	Storage Area Network
VM	virtuele machine

Bibliografie

- Atlassian. (2021a). Access Pricing. Geraadpleegd op 8 mei 2022, van <https://www.atlassian.com/software/access/pricing> ↗
- Atlassian. (2021). Confluence - Prijzen. Geraadpleegd op 1 mei 2022, van <https://www.atlassian.com/nl/software/confluence/pricing?tab=cloud> ↗
- B. (2021a, december 29). What is hybrid identity with Azure Active Directory? - Microsoft Entra. Microsoft Docs. Geraadpleegd op 25 april 2022, van <https://docs.microsoft.com/en-us/azure/active-directory/hybrid/whatis-hybrid-identity> ↗
- B. (2022, 22 januari). Azure AD Connect: When you already have Azure AD - Microsoft Entra. Microsoft Docs. Geraadpleegd op 27 april 2022, van <https://docs.microsoft.com/en-us/azure/active-directory/hybrid/how-to-connect-install-existing-tenant> ↗
- B. (2022b, mei 5). What is Azure AD Connect cloud sync? - Microsoft Entra. Microsoft Docs. Geraadpleegd op 30 april 2022, van <https://docs.microsoft.com/en-us/azure/active-directory/cloud-sync/what-is-cloud-sync#comparison-between-azure-ad-connect-and-cloud-sync> ↗
- B. (2022c, juni 2). Authentication for Azure AD hybrid identity solutions - Microsoft Entra. Microsoft Docs. Geraadpleegd op 28 april 2022, van <https://docs.microsoft.com/en-us/azure/active-directory/hybrid/choose-ad-authn> ↗
- Bigelow, S. J. (2020, 30 september). What is a SAN? Ultimate storage area network guide. SearchStorage. Geraadpleegd op 5 juni 2022, van <https://www.techtarget.com/searchstorage/definition/storage-area-network-SAN> ↗
- Fisher, C. (2018). Cloud versus On-Premise Computing. American Journal of Industrial and Business Management, 08(09), 1991–2006. <https://doi.org/10.4236/ajibm.2018.89133> ↗
- Hristov, T. (2020). vSAN 2-Node Cluster Guide | VMware. The Cloud Platform Tech Zone. Geraadpleegd op 19 juni 2022, van <https://core.vmware.com/resource/vsan-2-node-cluster-guide#sec7394-sub5> ↗
- J. (2021, 23 december). Gebruikers, servers & workstations analyse in Active Directory. 365tips.be. Geraadpleegd op 25 mei 2022, van <https://365tips.be/gebruikers-servers-workstations-analyse-in-active-directory-voor-een-security-assessment/> ↗

- Kardashevsky, C. (2020, 16 november). Active Directory LDAP Query Examples – TheITBros. Geraadpleegd op 5 maart 2022, van <https://theitbros.com/ldap-query-examples-active-directory/>
- Kerio Connect Migration Service. (2021). Kerio Connect Migration Service. Geraadpleegd op 30 mei 2022, van <https://manuals.gfi.com/en/kerio/connect/content/server-configuration/export-and-migration/kerio-connect-migration-service-1896.html?cshid=1896>
- Microsoft. (2021). Alle 365-abonnementen vergelijken. Geraadpleegd op 1 juni 2022, van <https://www.microsoft.com/nl-be/microsoft-365/business/compare-all-microsoft-365-business-products?&activetab=tab:primaryr2>
- RAID 1. (2018, 28 april). [Illustratie]. <https://linuxscriptshub.com/wp-content/uploads/2017/04/Raid1.jpg>
- Sample FIO commands for block volume performance tests on linux-based instances. (2022). (C) Copyright 2022. Geraadpleegd op 10 juni 2022, van <https://docs.oracle.com/en-us/iaas/Content/Block/References/samplefiocommandslinux.htm>
- StarWind. (2022, 18 januari). StarWind Virtual SAN® for vSphere 2-Node Hyperconverged Scenario with VMware vSphere 7. Resource Library. Geraadpleegd op 10 juni 2022, van <https://www.starwindsoftware.com/resource-library/starwind-virtual-san-for-vsphere-2-node-hyperconverged-scenario-with-vmware-vsphere-7/>
- Sumina, V. (2022, 7 juni). 26 Cloud Computing Statistics, Facts & Trends for 2022. Cloudwards. Geraadpleegd op 13 juni 2022, van <https://www.cloudwards.net/cloud-computing-statistics/>
- Tuxis internet engineering. (2022, 30 maart). Kerio in de cloud. Tuxis B.V. Geraadpleegd op 21 mei 2022, van <https://kerioindecloud.nl/>
- Vboxx. (2022). Email - vBoxx Hosting and Cloud Solutions. Email - vBoxx. Geraadpleegd op 21 mei 2022, van <https://vboxx.eu/email>
- VMWare. (2021). VMware vSAN licensing, pricing and packaging. <https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsan/vmware-vsani-licensing-guide.pdf>
- What's Split Brain and how to avoid it like the plague? (2019, 7 juni). StarWind Blog. Geraadpleegd op 7 juni 2022, van <https://www.starwindsoftware.com/blog/whats-split-brain-and-how-to-avoid-it>
- Wikipedia contributors. (2022, 12 juni). Hypervisor. Wikipedia. Geraadpleegd op 27 mei 2022, van <https://en.wikipedia.org/wiki/Hypervisor>

Campus Brugge
Xaverianenstraat 10
8200 Brugge
T 050 30 51 00

Campus Brugge station
Spoorwegstraat 12
8200 Brugge
T 050 40 59 00

Campus Kortrijk
Doorniksesteenweg 145
8500 Kortrijk
T 056 26 41 60

Campus Oostende station
Lijndraaiersstraat 60
8400 Oostende
T 059 56 90 00

Campus Oostende VLOC
Nieuwpoortsesteenweg 945C
8400 Oostende
T 059 30 81 50

Campus Roeselare
Wilgenstraat 32
8800 Roeselare
T 051 23 23 30

Campus Torhout
Sint Jozefstraat 1
8820 Torhout
T 050 23 10 30

www.vives.be

