

Storage op instellingsniveau

Aanbevelingen lange termijn

Auteur(s): SURFnet

Versie: 1.01

Datum: oktober 2012

Inhoudsopgave

Managementsamenvatting	4
1 Inleiding	5
2 Huidige storage behoefte	6
2.1 Algemene eisen.....	6
2.2 Twee verschillende categorieën gebruik.....	6
2.3 Technische eisen aan storage	7
3 Dilemma's, keuzes, oplossingsrichtingen	10
3.1 Alles op één oplossing of niet	10
3.2 Zelf doen of uitbesteden?.....	10
3.3 Financiële/ contractuele aspecten/ afwegingen	11
4 Vormen van storage.....	13
4.1 Drie klassen van storage.....	13
4.2 Block vs. file vs. object (SAN vs. NAS)	13
4.3 Replicatie.....	14
4.4 Auto-tiering	14
4.5 Disks.....	15
4.6 Toegang tot storage: iSCSI vs Fiber Channel	16
4.7 Tape Backup	17
4.8 Hype Cycle	17
5 De rol van de cloud	19
5.1 Cloud storage diensten	19
5.2 SURF Cloud strategie	20
5.3 Aanbod kant van de cloud.....	21
5.4 Privacy en veiligheid	22
5.5 Voor welke storage klassen is de cloud bruikbaar?	23
5.6 Exit strategie en migratie bij de cloud	23
6 Aanbevelingen.....	24
6.1 Stel de huidige eisen en wensen op functioneel niveau vast	24



6.2	Ontwikkel een visie op storage	24
6.3	Creëer overzicht in de IT architectuur	25
6.4	Samenwerking gebruikers, informatiebeleid en IT diensten	26
6.5	Managen interne kennis voor Storage omgeving	26
6.6	Samenwerking met andere instellingen	27
6.7	Inkoopproces.....	28
6.8	Structureel test- en acceptatie inplannen.....	29
7	Appendix: storagedefinities.....	30

Managementsamenvatting

Voor de instellingen die aangesloten zijn bij SURF is storage een steeds belangrijker onderwerp. De behoefte aan meer storage vanuit instellingen stijgt snel, en de kosten van de storage en het beheer daarvan nemen ondanks de prijsdalingen van de componenten nog altijd toe.

De afgelopen vijf jaar is veel gebeurd op het gebied van storage. Storage hardware kan steeds meer data aan, en de snelheid stijgt. Daarnaast zijn cloud oplossingen, mede door de toenemende beschikbaarheid van breedband verbindingen, steeds populairder aan het worden. Deze ontwikkelingen volgen elkaar in hoog tempo op, waardoor het up to date houden van de interne kennis op dit gebied tijdrovend en daardoor ook kostbaar is.

Instellingen moeten als gevolg van deze ontwikkelingen andere keuzes maken omtrent storage dan voorheen. Waar het een aantal jaar geleden helder was dat een IT afdeling ook zelf de storage moet beheren, is dit nu niet meer vanzelfsprekend en ook soms zelfs niet meer haalbaar. Het uitbesteden van storage aan een andere partij brengt echter nieuwe issues met zich mee. Instellingen moeten nu keuzes maken in hoeverre de storage intern of extern geregeld moet worden.

Bij het zelf doen (ook bij het opstellen van een Programma van Eisen voor de inkoop) moeten instellingen rekening houden met verschillende technische eisen voor verschillende toepassingen. In het algemeen worden hier drie “tiers” onderscheiden: tier-1 (primaire) storage vereist een zeer lage latency, een zeer hoge beschikbaarheid, en een hoge “throughput”, omdat het direct gebruikt wordt voor de primaire processen. Tier-2 en tier-3 (bulk storage) stellen lagere eisen.

Naast de storage niveaus moeten instellingen kiezen voor block-based, file-based en object-based storage en alle daarbij behorende consequenties voor de verdere architectuur.

De cloud biedt drie algemene aanbiedingsvormen van storage aan, namelijk de publieke cloud (toegankelijk voor iedereen), de community cloud (toegankelijk voor een specifieke groep), en de private cloud (enkel toegankelijk voor één instelling). Alle vormen van cloud diensten brengen beveiligings- en privacy risico's met zich mee.

Om te profiteren van de verschillende technologische ontwikkelingen die zich op het gebied van storage voordoen, dienen instellingen in eerste instantie hun eigen wensen en eisen goed in kaart te brengen. Het is hierbij belangrijk de algehele IT architectuur inzichtelijk te maken, de kennis in de verschillende interne afdelingen te managen en de behoeften van de verschillende groepen gebruikers afstemmen met de IT afdeling.

Voordat instellingen commerciële partijen in overweging nemen, is het raadzaam om met andere bij SURF aangesloten instellingen samen te werken om samen oplossingen op te zetten, gezamenlijk in te kopen, of om kennis te delen. Via SURFmarket kan gedacht worden aan gezamenlijk inkoop of aan het gebruik maken van ongebruikte storage ruimte bij andere instellingen.

Bij het inkopen (ook bij cloud oplossingen) moeten instellingen bedacht zijn op prijsdalingen van hardware in de toekomst, een goede planning maken voor het testen en accepteren van de oplossing, en rekening houden met de migratie van data.

1 Inleiding

De behoefte aan “storage” (opslagruimte voor data) voor onderwijs en onderzoek blijft stijgen. Tegelijkertijd zijn er steeds meer soorten en maten van storage mogelijk. Met de komst van snelle breedbandige verbindingen is het bovendien mogelijk geworden om storage uit te besteden naar de cloud.

De meeste instellingen hebben de afgelopen jaren geïnvesteerd in enterprise storage oplossingen in het eigen datacenter. Dergelijke storage systemen kenmerken zich door een hoge prijs per datavolume. Nu de storage behoefte blijft stijgen, worstelen instellingen met de vraag of ze moeten blijven investeren in het eigen storagestelsel, hoe ze dat systeem in moeten richten, en of storage wellicht ook uitbesteed kan worden naar de cloud.

Dit rapport biedt antwoorden op vragen rondom storage en geeft een lange termijn visie op het vlak van storage voor instellingen in het hoger onderwijs en onderzoek. Deze visie is gebaseerd op diverse onderzoeken op dit gebied en op de ervaringen van storage experts bij SURFnet en Stratix.

In dit document is alleen gekeken naar de ontwikkelingen op enterprise storage vlak, en niet naar de verdere ontwikkelingen binnen de ICT infrastructuur bij de instellingen.

ICT afdelingen van de instellingen kunnen dit document gebruiken om een eigen visie te vormen over storage oplossingen voor de specifieke situatie van de instelling.

Dit rapport is uitgegeven in opdracht van de Universiteit Twente.

2 Huidige storage behoefte

Dit hoofdstuk bespreekt de storage behoefte van instellingen binnen het hoger onderwijs en onderzoeksdomein. Het gaat hier om toepassingen en applicaties die studenten, docenten, onderzoekers en ICT afdelingen nodig hebben, die invloed hebben op de algehele storage eisen van de instelling. Instellingen hebben zowel zogenaamde enterprise-storage als onderzoeks-storage nodig. Uit de praktijk blijkt dat dit veelal gescheiden storage omgevingen zijn.

2.1 Algemene eisen

Steeds meer onderwijs en onderzoeksmateriaal wordt primair digitaal aangeboden. Daarnaast wordt er in onderwijs en onderzoekstrajecten steeds meer data gegenereerd. Dit leidt in combinatie met een grotere inzet van multimedia tot een stijgende storage behoefte binnen instellingen. De afgelopen jaren hebben veel instellingen aanzienlijk geïnvesteerd in het inkopen van additionele storage.

De eisen die gesteld worden aan de beschikbaarheid van diensten binnen instellingen zijn de laatste jaren flink gestegen. Een enterprise-storage omgeving is een essentiële bouwsteen voor kritische applicaties, zoals e-mail en bestand opslag. Dergelijke applicaties vergen een hoge beschikbaarheid en lage kans op data verlies. Dit vertaalt zich naar de eisen die aan de storage laag worden gesteld.

Gebruikers van applicaties en storage systemen verwachten door de lage kosten van portable harddisks en door de grote hoeveelheid gratis storage van cloud leveranciers dat storage op instellingsniveau ook onder vergelijkbare condities, zoals lage kosten en oneindig schaalbaar kan worden aangeboden.

Onderzoekers, studenten en docenten werken steeds meer samen met nationale en internationale partners. Hierdoor is er een stijgende vraag om data eenvoudig te kunnen delen met deze externen.

2.2 Twee verschillende categorieën gebruik

In een instelling zijn er twee soorten gebruik van storage te onderscheiden, namelijk de onderzoek gerelateerd gebruik en de algemene kantoorautomatisering, waaronder ook onderwijstoepassingen vallen.

2.2.1 Wetenschappelijke toepassingen

Onderzoeksgroepen hebben grote hoeveelheden opslagcapaciteit nodig op korte termijn en voor beperkte tijd. De precieze eisen hangen af van het type onderzoek, maar over het algemeen zijn er twee categorieën, namelijk de grote storage gebruikers en de kleinere storage gebruikers. De eerste categorie heeft meestal een specifiek experiment met speciale storage eisen. De storage is essentieel voor het volbrengen van het onderzoek en deze categorie heeft veelal budget en faciliteiten om bij SARA specifieke storage diensten af te nemen. De kleinere onderzoeksgroepen hebben zeer diverse behoeften, geen speciale voorzieningen, beperkt budget en zoeken vaak zelf naar oplossingen buiten de ICT-afdeling om. Voor beide categorieën geldt dat de storage aan de onderzoekseisen moet voldoen en schaalbaar moet zijn.

ICT-afdelingen worden regelmatig geconfronteerd met onderzoeksafdelingen die binnen een relatief korte termijn een grote hoeveelheid storage willen afnemen. Vaak is het lastig om aan deze vraag te voldoen, doordat er specifiek eisen worden gesteld aan de storage. ICT afdelingen hebben een beperkte keuze binnen de reeds aanwezige storage systemen en daarvan zijn de kosten vaak proportioneel hoger dan bij commerciële (cloud) providers. Onderzoekers hebben daardoor de neiging

om rechtstreeks zaken te doen met commerciële storage providers. Hiermee omzeilen ze het centrale storage aanbod. Dit kan extra risico's met zich mee brengen in termen van veiligheid en privacy, en ervoor zorgen dat de centrale ICT-afdeling het zicht verliest op de aard van de diensten die worden gebruikt.

Onderzoekers in de tweede categorie zoeken hun heil in consumenten storage oplossingen. Externe harde schijven kosten niet veel en kunnen eenvoudig aangesloten worden. Voor de mobiele onderzoeker wordt vaak gebruik gemaakt van een eigen server, Dropbox of simpelweg een USB-stick. Deze eigen oplossingen brengen veel privacy en veiligheid issues met zich mee, maar ook technische risico's in verband met redundantie en back-ups. Omdat de storage afname buiten de IT-afdeling gebeurt, is er geen garantie dat onderzoeksdata op een goede manier gebackupt wordt.

Centrale ICT-afdelingen moeten in staat zijn om flexibele IT-diensten aan onderzoekers te faciliteren, met behoud van controle (of in ieder geval inzicht) over het bestede bedrag. Zij geven de voorkeur aan voorspelbare kosten voor de generieke IT-functies, waarop zij hun begrotingen kunnen baseren. Pay per use modellen zijn voor IT-afdelingen acceptabel, zolang de totale kosten toe te schrijven zijn aan de betrokken onderzoeksgroepen.

ICT-afdelingen gebruiken virtualisatie technieken zoals VMware binnen hun eigen datacenter, om flexibel en efficiënt met servers te kunnen omgaan (IaaS). In de meeste gevallen wordt IaaS niet als dienst aangeboden naar onderzoekers en docenten, maar is er enkel een proces ingericht waar (virtuele) computers worden aangeboden. Er is binnen het hoger onderwijs en onderzoeksdomein behoefte aan flexibiliteit, bijvoorbeeld om tijdelijke virtuele servers beschikbaar ten behoeve van een experiment. Het heeft dan ook toegevoegde waarde als IaaS diensten rechtstreeks toegankelijk komen voor docenten en onderzoekers.

2.2.2 Kantoorautomatisering

Met kantoorautomatisering wordt bedoeld de HR systemen, financiële systemen, (virtuele) werkplekken en e-mail systemen. Deze toepassingen hebben snelle storage nodig met hoge beschikbaarheid. Deze toepassingen zijn essentieel voor de bedrijfsvoering en betrouwbaarheid en het hebben van een goede backup is zeer belangrijk. Ook voor deze toepassingen worden virtualisatie technieken ingezet om efficiënter met de beschikbare hardware om te gaan.

Voor IT-afdelingen is deze categorie makkelijker te bedienen in vergelijking met de wetenschappelijke toepassingen. De diensten zijn beter te managen qua omvang en kosten. Wel is het een vereiste om zeer betrouwbaar en stabiele systemen te kunnen aanbieden.

Binnen het hoger onderwijs is er een trend van Bring-Your-Own-Device (BYOD), waarbij medewerkers en studenten hun prive devices zoals o.a. Ipad's, iPhones en Android devices ook binnen de instelling inzetten. Het beschikbaar stellen van reguliere IT-diensten op deze devices is verre van triviaal. IT-afdelingen zijn vaak gewend om een standaard werkplek te moeten ondersteunen en BYOD brengt daar een andere dynamiek in. Daarnaast zijn er uitdagingen om de data die op dergelijke devices wordt opgeslagen veilig te backuppen.

2.3 Technische eisen aan storage

2.3.1 capaciteit

De behoefte aan storage is de afgelopen jaren sterk toegenomen. Hierdoor hebben veel instellingen hun storage omgeving moeten uitbreiden en/of vervangen. Ook de komende jaren wordt een sterke

toename verwacht aan opslagcapaciteit, maar nu steeds meer instellingen diensten in de cloud afnemen zal een deel van de storagegroei worden afgewenteld op de cloud leverancier.

2.3.2 Performance

Er zijn verschillende performance en prijs niveaus van storage. Het is belangrijk om te weten welk type storage gewenst is voor een bepaalde toepassing. Er zijn vijf kenmerken te onderscheiden:

- **Random access:** De data wordt willekeurig opgevraagd en/of weggeschreven. Dit gebeurt niet gestructureerd en is daarmee niet voorspelbaar. Binnen een bepaalde tijdsspanne zullen er hot-zones zijn op de fysieke storage, waarin veel lees en schrijfacties plaatsvinden.
- **Sequential access:** Het lees en/of schrijfproces gebeurt primair sequentieel. Daarmee is het gedrag goed voorspelbaar.
- **Latency:** De latency wordt vaak uitgedrukt in een tijdseenheid. Latency geeft aan hoeveel milliseconde er gewacht moet worden op een reactie bij een verbinding. Hiermee wordt aangeduid hoe snel de storage benaderbaar is.
- **IOPS:** Deze performance indicator stelt het aantal input/output operaties per seconde voor (IOPS). Hiermee wordt aangeduid hoeveel blokjes (block-based storage) de storage systeem per seconde kan verwerken.
- **Bandwidth:** De hoeveelheid data die binnen een 1 seconde kan worden verstuurd. Meestal uitgedrukt in megabytes of gigabytes per seconde.

Van belang is om per platform of applicatie een goed beeld te krijgen van het soort gebruik, en de storage waar nodig te optimaliseren voor dat type gebruik.

2.3.3 Snelle vs. langzame storage

Een belangrijke overweging is de mix tussen snelle en langzame storage. Deze vormen van storage kunnen gemengd zijn, maar er zijn ook een aantal goede argumenten om deze vormen op aparte systemen onder te brengen¹. Op die manier kunnen instellingen de verschillende systemen optimaal configureren voor de eisen die de toepassing stelt.

Indien instellingen besluiten om snelle en langzame storage gecombineerd te houden, is het in elk geval nuttig om per platform of applicatie te inventariseren welke performance benodigd is, en de configuratie van het storage systeem daarvoor te optimaliseren.

2.3.4 'High' vs. 'regular' beschikbaarheid

Voor verschillende applicaties is een verschillende mate van beschikbaarheid en betrouwbaarheid nodig. In het dataopslagbeleid zou per applicatie vastgesteld moeten worden wat de *Recovery Point Objective* (RPO)² en *Recovery Time Objective* (RTO)³ zouden moeten zijn, maar doordat elk platform verschillende applicaties ondersteunt gelden in feite per platform steeds de strengst mogelijke eisen aan RPO en RTO.

Hoge eisen aan RPO en RTO leveren niet alleen hoge kosten op, maar vereisen ook technisch complexe oplossingen waarmee weer nieuwe risico's geïntroduceerd kunnen worden. De hoogste

¹ Zo is er bij de meeste storage systemen geen effectieve manier om pieken van minder belangrijke applicaties uit te smeren om daarmee vertraging op belangrijke data/applicaties te voorkomen.

² RPO: *Recovery Point Objective*, een maatstaf (in tijd uitgedrukt) voor de hoeveelheid recente data die bij een calamiteit verloren mag gaan.

³ RTO: *Recovery Time Objective*, de tijd die het mag duren tot systemen na een calamiteit weer beschikbaar zijn.

RPO en RTO kunnen bijvoorbeeld gehaald worden door middel van synchroon gerepliceerde storage, maar dit gaat ten koste van de performance en levert hogere risico's voor de data-integriteit (doordat gegevens corruptie ook onmiddellijk gerepliceerd wordt).

De meeste platformen en/of applicaties kunnen zelf voor redundantie zorgen; in het algemeen verdient het de aanbeveling om de replicatie zo veel mogelijk uit te voeren op de applicatie- of platformlaag. Dat betekent dat je voor die applicaties ook geen synchrone of asynchrone replicatie meer nodig hebt vanuit de storage-oplossing, en dat daarvoor een reguliere snapshot en/of backup voldoet.

3 Dilemma's, keuzes, oplossingsrichtingen

Dit hoofdstuk bespreekt de keuzes die door de instellingen gemaakt moeten worden ten aanzien van het afnemen van enterprise storage. Storage voor kleine onderzoeksgroepen kan qua wensen en eisen vaak gebruik maken van enterprise storage, al is de beschikbaarheid en daarmee de prijs hoger dan gewenst. Grotere onderzoeksgroepen hebben specifieke eisen die verder onderzocht moeten worden en daarmee buiten de scope van dit rapport vallen.

3.1 Alles op één oplossing of niet

De eerste keuze waar een instelling mee geconfronteerd wordt, is de keuze tussen één storage oplossing die aan alle storage behoeftes en eisen voldoet, of verschillende storage systemen voor verschillende behoeftes, toepassingen en eisen.

Een storage systeem dat aan alle eisen moet voldoen is een complex systeem. Het moet een systeem zijn dat voldoet aan de hoogste eisen die gesteld worden door iedere storage toepassing. Dus bijvoorbeeld een hoge capaciteit gecombineerd met snelle toegang, hoge beschikbaarheid en de mogelijkheid om zeer grote maar ook veel kleine databestanden op te slaan. Dit resulteert in een dure en omvangrijke oplossing.

3.2 Zelf doen of uitbesteden?

Up-to-date blijven met de juiste hoeveelheid en type hardware voor storage is moeilijk, omdat de behoeften van de eindgebruikers snel veranderen en nieuwe hardware-oplossingen veelal pas met een interval van vier a vijf jaar worden aangeschaft. Outsourcen van delen van de storage laag ontlast de instellingen van deze schaalbaarheid kwesties, en verschuift het probleem naar de provider, die (als gevolg van schaalvoordelen en specifieke expertise) in een betere positie is de toenemende eisen bij te houden.

Instellingen verwachten dat outsourcing financieel aantrekkelijk is, omdat partijen die zijn gespecialiseerd in het verstrekken van deze diensten in staat moeten zijn om hun systemen efficiënter te beheren en tegen lagere kosten vanwege schaalvoordelen. Ze zijn in staat om efficiënt te werken met grote platforms en de voorziene capaciteit tussen de verschillende eindgebruikers te verdelen.

3.2.1 Beschikbaarheid

Beschikbaarheid is van belang bij het uitbesteden van onderdelen van een vitale infrastructuur. Dit is met name in het IT domein van belang, omdat bij uitval van een onderdeel uit de keten, direct de gevolgen zichtbaar zijn op dienst niveau. In het bijzonder binnen het domein van de kantoorautomatisering is het van belang dat applicaties altijd beschikbaar zijn. Voor onderzoeksdoeleinden daarentegen is de beschikbaarheid veelal een minder kritieke factor en telt de rekenkracht en de hoeveelheid opslagruimte zwaarder.

3.2.2 Vendor lock-in en migratie

Bij het overstappen naar een extern storage provider is overeenstemming ten aanzien van een exit-plan van belang. Sommige providers geven systematisch contractuele garanties dat ze zullen meewerken aan het verplaatsen van de gegevens naar een andere provider aan het einde van een contract, en sommige hebben zelfs escrow contracten om ervoor te zorgen dat hun klanten het eigendom van hun gegevens behouden indien de aanbieder zou failliet gaan. Klanten zullen moeten nadenken over dergelijke contractuele regelingen, alsmede de technische inregeling, om flexibel te blijven en het voorkomen van vendor lock-in.

3.3 Financiële/ contractuele aspecten/ afwegingen

Bij het inkopen of uitbesteden van storage komen er naast technische eisen ook contractuele en financiële zaken kijken. Contracten via aanbestedingen hebben doorlooptijden van vier jaar en daarom moet men rekening houden met technische- en prijs ontwikkelingen op het gebied van storage over de looptijd van het contract.

3.3.1 Total cost of ownership

Bij het inkopen van een storage systeem komen naast de inkoopkosten, andere, soms onvoorziene kosten kijken. Hieronder een aantal belangrijke elementen om rekening mee te houden:

- Hardware
- Onderhoud
- Energie
- Slijtage disks
- Huur ruimte (racks)
- (Storage) netwerk
- Investering in kennis (product training)
- Inkoopproces (tender)
- Migratie kosten

Vooraf energie is één van de kostenposten waar vaak niet bij stilgestaan wordt. De prijs van energie stijgt, en daarmee stijgen de totale kosten van het eigendom van een storage systeem. Energie is te verdelen in grijze of groene energie en ook deze keuzes bepalen hoe de total cost of ownership uitvalt.

Migratie kan leiden tot dubbele kosten, omdat het oude systeem gehouden moet worden terwijl de nieuwe ook er al is. Migratie is tevens een onderschatte kostenpost. Met de steeds groeiende hoeveelheid data wordt migratie steeds complexer, heeft een lange doorlooptijd met potentieel langdurige onderbrekingen. Nieuwe storage architecturen worden steeds meer ontworpen om dit probleem op te lossen, bijvoorbeeld door data online te migreren tussen de oude en de nieuwe fysieke storage. Dit is een belangrijk aspect bij tender requirements.

Verder zijn er verborgen kosten van storage. Een voorbeeld hiervan is wanneer medewerkers ervoor kiezen om individueel storage diensten zoals Dropbox in de cloud af te nemen, en deze vervolgens te declareren. De kosten hiervoor komen dan misschien niet in het budget van de ICT afdeling, maar het zijn additionele kosten en zorgen ervoor dat besparingen op algemene instelling storage teniet gedaan worden.

3.3.2 Prijzen vastzetten of niet

Bij het inkopen van storage wordt vaak geselecteerd op de laagste prijs. Dit is niet altijd verstandig. De laagste prijs wordt namelijk soms bereikt doordat een storage oplossing basic is en niet is voorzien op eventuele uitbereidingen vanwege toenemende storage behoefte. Hierdoor lopen de kosten sterk op indien er additionele storage bijgekocht wordt.

In verband met potentiële toekomstige uitbreiding van de storage moet bij de initiële aanschaf geregeld worden dat toekomstige uitbreidingen conform de toekomstige marktprijs geschieden. Normaal gesproken dalen deze kosten per kwartaal. Daarnaast gebeurt het vaak dat na het aflopen van het contract, de support verlengt moet worden. Deze verlenging is vaak zeer kostbaar en het is aan te bevelen om voor de aanschaf hier al prijsafspraken over te maken.

Sommige aanbieders van storage hanteren een storage on-demand model, dit kan handig zijn bij de prijsafspraken. Maar ook hier moet in het contract rekening gehouden worden met prijsdalingen in de toekomst.

4 Vormen van storage

Dit hoofdstuk bespreekt de verschillende type technische oplossingen die beschikbaar zijn voor de storage behoefte van de instellingen.

4.1 Drie klassen van storage

De primaire storage, ook wel tier-1 storage genaamd, is de storage die direct, met zeer lage latency beschikbaar is vanaf een server of vanaf de applicatielaag. Voorbeelden waarbij deze storage ingezet wordt zijn: de storage waarop het operating systeem (Windows, Linux) is geïnstalleerd of de storage die gebruikt wordt voor een database applicatie.

Secundaire storage, ook wel tier-2 storage, is de storage waarop data wordt geplaatst waarbij lage latency minder kritisch is voor de performance van een applicatie of besturingssysteem. Bijvoorbeeld de opslag van media files of documenten.

Ten slotte is er ook tier-3 storage. Kenmerkend voor deze storage is dat de latency nog veel hoger is. Tier-3 wordt met name gebruikt voor grote bulk opslag, zoals bijvoorbeeld het maken van backups van tier-1 en tier-2 storage.

4.2 Block vs. file vs. object (SAN vs. NAS)

Centrale storage systemen bestaan in principe in twee vormen: SAN (Storage Area Network) en NAS (Network Attached Storage). Een SAN presenteert “block level” storage, en emuleert daarmee een opslagmedium dat direct met een server verbonden is. Een NAS presenteert “file level” storage, waar servers en clients direct gebruik van kunnen maken. Daarnaast bestaat de “object level” storage.

Block Level storage biedt de klant toegang op het lage niveau van de opslag van de provider, en is zeer flexibel: de klant kan beslissen hoe de opslag wordt gestructureerd, en kan kiezen voor het optimale file systeem of database systeem per applicatie. Block level storage kan zeer lage latency behalen tussen de server en de opslag.

File Level storage geeft direct toegang tot het bestandsysteem van de opslagdienst. Gebruikers kunnen rechtstreeks gebruik maken van deze vorm van storage of via een server. Deze manier van opslag wordt vaak gebruikt als een makkelijke manier om bestanden op te slaan en te delen. File level storage biedt een specifiek bestandsysteem aan (bijvoorbeeld NTFS of NFS) voor alle gebruikers van het systeem. Het wordt vaak geïmplementeerd als een netwerkschijf die toegankelijk is voor verschillende gebruikers via hun PC's. Ook kunnen verschillende toegangsrechten geregeld worden. File level storage is echter niet geschikt voor database systemen die een ‘raw’ volume vereisen.

Object Level storage wordt met behulp van high-level flexibele opslag objecten ('containers') georganiseerd. Rond 2000 waren de eerste oplossingen gebaseerd op Content-addressable *storage* (CAS). CAS is een tegenhanger van de meer bekende location based storage. Traditionele filesystemen zijn location based. De data wordt namelijk in een bepaalde directory geplaatst en de gebruiker hoeft enkel te weten in welke directory en met welke naam de data is weggeschreven. Bij een CAS systeem wordt door het storage systeem na het wegschrijven van data dmv een hashing algoritme aangegeven hoe bepaalde content kan worden opgevraagd. De gebruiker dient de verkregen string te onthouden, omdat anders de content niet meer benaderd kan worden. Een voordeel van object storage is dat het triviaal is om de storage backend geografisch te distribueren en te repliceren.

Traditionele CAS omgevingen, zoals bijvoorbeeld EMC's Centera, zijn niet populair geworden, maar sinds dergelijke omgevingen meer als een cloud omgeving worden aangeboden stijgt het gebruik. Met name in web omgevingen zijn dergelijke systemen populair. Een groot voordeel van dit type storage is de grote mate van schaalbaarheid.

Interfacing gebeurt met behulp van web interfaces, meestal op basis van SOAP of REST. Een voorbeeld is Amazon's S3 storage.

In het verleden liepen server consolidatie trajecten vaak hand in hand met het inrichten van een SAN. Daarmee werd het mogelijk op een homogene manier aan de verschillende storage behoeften van de servers te voldoen. Dit is aan het veranderen doordat enerzijds de virtualisatie zijn intrede heeft gemaakt en anderzijds de verschillen in performance tussen een NAS en SAN verdwenen zijn: de meeste virtualisatieplatforms kunnen gebruik maken van NAS en SAN omgevingen.

4.3 Replicatie

De meeste SAN's ondersteunen zowel synchrone als asynchrone replicatie. Beide benaderingen hebben echter nadelen. Synchrone replicatie kan, afhankelijk van de afstand tussen de locaties, extra vertraging opleveren, terwijl asynchrone replicatie nooit uit kan sluiten dat de data bij een calamiteit in een inconsistente toestand achterblijft.

Waar mogelijk wordt replicatie van data daarom meestal over gelaten aan de applicatie of het platform zelf. Veelal is de replicatie binnen een toepassing optimaal ingericht voor die toepassing. De applicatie is op de hoogte van de toestand van de verschillende kopieën, en bij uitval kan de applicatie of het platform zelf zorgdragen voor continuering en herstel. Producten als Oracle Data Guard leveren daardoor vaak een veel efficiëntere wijze van replicatie dan de storage omgeving. Replicatie op storageniveau ligt vooral voor de hand als replicatie niet mogelijk is op applicatie of platform niveau. Belangrijk is om te realiseren dat een replicatie op storage niveau geen vervanger is voor het maken van backups.

Bij het maken van backups speelt een vergelijkbare afweging. In veel gevallen zullen applicaties of platforms beschikken over de mogelijkheid om backups te maken; dit is meestal efficiënter dan een backup op storage niveau (bijvoorbeeld door snapshots vanuit de storage omgeving).

Een interessante tegenhanger van replicatie zijn de Information Dispersal Algorithms. Met deze techniek wordt een data block opgesplitst in meerdere kleinere stukjes en verspreid naar meerdere opslagsystemen. De totale som van kleine stukjes data is altijd groter dan het originele data block en hiermee wordt redundantie gecreëerd. Hierdoor kan de originele data gereconstrueerd worden indien niet alle kleine stukjes data beschikbaar zijn.

Information Dispersal Algorithms zijn het bekendst in peer-to-peer netwerken, maar worden als techniek ook steeds meer ingezet in cloud storage oplossingen vanwege de robuustheid. Bijvoorbeeld de cloud storage omgeving van Scality en Cleversafe maken gebruik van deze technieken.

4.4 Auto-tiering

Een alternatief voor het verwerven van verschillende storage oplossingen is een "tiering" model, waarbij één storage systeem verschillende vormen van storage aanbiedt: hoge capaciteit met lage performance, hoge performance maar tegen hogere kosten, etc. Per soort gegevens kan dan binnen het systeem de optimale vorm gekozen worden. Een ontwikkeling van de laatste jaren is "automated

tiering”, waarbij het systeem zelf (of een apart stuk software) aan de hand van het gebruik van gegevens vaststelt in welke tier de gegevens bewaard moeten worden.

Het is handig dat SAN-storage-vendors tiered storage functionaliteit aanbieden, want bij de meeste bedrijven kan het overgrote merendeel van de data op de relatief goedkope SATA-disk worden geplaatst binnen de SAN-omgeving. Het grote nadeel is dat SAN-leveranciers tiered storage alleen binnen hun eigen product aanbieden en ondersteunen. Als klant ben je dan nog steeds met handen en voeten gebonden aan deze leverancier en uitwijk naar een derde partij voor storage is bijzonder lastig.

Auto-tiering oplossingen beginnen nu pas volwassen te worden en er zijn een paar instellingen die tevreden zijn over de werking van een dergelijke oplossing. Eerdere implementaties hadden vaak veel “tuning” nodig (en dus veel expertise) om de voordelen daadwerkelijk te bereiken. Een niet goed werkend auto-tiering systeem kan leiden tot veel verplaatsingen van gegevens binnen de storage oplossing, waardoor de performance juist verslechtert. Op dit moment zijn in het top-segment goede oplossingen verkrijgbaar, al is er nog steeds veel kennis nodig om de systemen optimaal in te zetten⁴. Auto-tiering is naar schatting over twee tot vijf jaar volwassen.

4.5 Disks

Voor storage zijn er disks nodig, maar afhankelijk van het type storage en de gewenste beschikbaarheid daarvan moet er een optimum gevonden worden tussen verschillende kosten en performance eigenschappen van disks.

Er zijn vele typen disks met ieder specifieke eigenschappen. SSD's zijn sneller dan harddisks, maar ook een stuk duurder, en worden dan ook veelal voor caching gebruikt of voor kleine hoeveelheden gegevens waar men snel toegang toe wil hebben, maar niet voor 'bulk storage'. Ook tussen types harddisks zijn er grote verschillen, en afhankelijk van de wens voor performance of capaciteit kunnen hier andere disks gekozen worden. Voor een gegeven type storagegebruik wordt getracht binnen randvoorwaarden van de applicaties het optimum te vinden in de verhoudingen tussen de verschillende types disks in het storagesysteem.

Een vooraanstaand onderzoeksbureau heeft inzichtelijk gemaakt welke ratio's tussen verschillende typen disks nodig zijn bij gangbare storage omgevingen om enerzijds een goede performance te garanderen en anderzijds de kosten zoveel mogelijk te beperken.

Voorbeeld van verschillende verhoudingen tussen disks voor verschillende type storage:

	SSD	High performance HDD	Medium to low HDD
General purpose	2%	30%	68%
Performance oriented	3%	62%	35%
Capacity based	0.5 %	0%	99,5%

⁴ Zie bijvoorbeeld: http://wikibon.org/wiki/v/Automated_Tiered_Storage_Ready_for_Prime_Time

Deze verdeling geldt voor een “gemiddelde” organisatie en een “klassieke” storage omgeving. Per organisatie zal gekeken moeten worden naar de verdeling van storage gebruik, en zal geanalyseerd moeten worden welke verhouding tussen SSD's, High Performance HDD's en Medium/Low performance HDD's voor dat specifieke geval, en binnen de randvoorwaarden, optimaal is.

Aangezien de technologie van disks en geheugen sterk in beweging is, zullen de genoemde verhoudingen niet constant zijn. Zo is het de vraag of de high-performance (15k SAS) HDD's in de toekomst nog een relevante oplossing vormen nu SSD's deze functie grotendeels overnemen⁵. De ontwikkeling naar nog sneller Harddisks (20k) lijkt in de praktijk niet (efficiënt) haalbaar vanwege praktische bezwaren zoals stroomverbruik en warmteontwikkeling. SSD's zullen naar verwachting de taken van de “high-performance HDD's” overnemen, terwijl de ‘reguliere’ 10k SAS en 7,2k SAS/SATA drives blijven voor ‘bulk’. Hierdoor zal dan ook de optimale verhouding tussen de typen verschuiven.

In toekomstige storage omgevingen zullen disks, die op bepaalde momenten niet in gebruik zijn, steeds vaker tijdelijk op non-actief geplaatst worden door de storage appliance om stroom te kunnen besparen.

4.6 Toegang tot storage: iSCSI vs Fiber Channel

Bij een keuze voor block level storage zal er ook gekozen moeten worden hoe deze storage ontsloten wordt. Fiber channel is vanuit oudsher de de-facto standaard om een SAN te ontsluiten. Fiber channel is begin jaren '90 ontworpen als storage netwerk technologie en heeft daardoor jarenlang grote voordelen gehad zoals een hoge throughput en lage latency. Een nadeel van Fiber Channel is dat het een apart netwerk is met eigen controllers (HBA's). De verwachting is dat Fiber Channel de komende 5 tot 10 jaar nog wel regulier zal worden ingezet in SAN omgevingen. Fiber Channel is relatief duur doordat vendors nauwelijks concurreren op prijs en doordat de techniek minder profiteert van de massamarkt zoals bij Ethernet. Inmiddels is er een techniek genaamd: Fiber Channel over Ethernet die enerzijds profiteert van de standaardisatie van Ethernet op WAN netwerken en die anderzijds probeert om de voordelen van Fiber Channel te benutten. Deze techniek is naar schatting over drie jaar volwassen om ingezet te worden in een omgeving waar hoge eisen aan support en beschikbaarheid gesteld worden.

iSCSI is een storage network protocol dat over een reguliere Ethernet infrastructuur kan worden ingezet. Dit protocol is ongeveer 10 jaar geleden bedacht. Groot voordeel van iSCSI is dat relatief goedkope hardware kan worden ingezet. Wel moet er expliciet op netwerk niveau rekening gehouden worden met het storage transport, zoals beschikbare capaciteit en redundante routes tussen servers en storage. iSCSI kan zeer slecht tegen “packet loss”, en heeft daarom in principe een “lossless” netwerk nodig; dat stelt hoge eisen aan de switches en vooral aan de verbindingen tussen de switches en daardoor is het zinvol om een afzonderlijk storage netwerk te gebruiken. iSCSI werd in het verleden geassocieerd met een lage performance, maar is inmiddels volwassen en doet anno 2012 onder de juiste condities niet meer onder voor Fiber Channel. De verwachting is dan ook dat iSCSI de komende jaren steeds meer de de-facto standaard zal worden.

Naast de opties⁶ Fiber Channel en iSCSI zijn er nog enkele alternatieven, die echter veel minder gebruikt worden: Infiniband, Fiber Channel over IP.

⁵ Zie bijvoorbeeld http://wikibon.org/wiki/v/Will_Faster_Hard_Disk_Drives_Compete_with_SSDs_and_Flash-based_Caches_like_VFCache%3F

⁶ <http://www.infoworld.com/d/data-explosion/Fiber-channel-vs-iscsi-the-war-continues-806>

4.7 Tape Backup

Tape backup is was vroeger voor veel organisaties de primaire manier om data te backullen. Inmiddels is tape backup sterk afgenomen als primaire backupvorm, doordat de opslagprijs van data naar disks steeds goedkoper is geworden en er software beschikbaar is gekomen die backups naar disks ondersteunt. Een paar voordelen van backup naar disk ten opzichte van tape zijn bijvoorbeeld dat disks sneller benaderd kunnen worden, een disk-setup in een remote locatie opgezet kan worden zodat er geen fysiek transport van tapes nodig is en er bij disks redundancy gecreëerd kan worden door middel van RAID. Een andere reden waardoor tape minder populair is als primaire opslag komt doordat veel storage omgevingen snapshot en replicatie functionaliteit ondersteunen in combinatie met deduplicatie technieken. Door middel van een snapshot kunnen eenvoudig veranderingen ongedaan worden gemaakt en kan “oude” data worden teruggezet. De deduplicatie zorgt er vervolgens voor dat de data ontdubbelt wordt.

Uit onderzoek bij marktpartijen blijkt dat tape backup als primair backup in 2005 nog bij twee derde van de bedrijven in gebruik was. Inmiddels is dat gedaald tot ongeveer 10 procent. Dit wil zeker niet zeggen dat tape backup dood is! Tape wordt namelijk nog wel volop gebruikt als archivering, om bijvoorbeeld data die nauwelijks gebruikt wordt te offloaden naar tape. Daarnaast blijven de technologische ontwikkelingen op tape vlak gestaag doorgaan, waardoor bijvoorbeeld de opslagcapaciteit van een enkele tape al gegroeid is naar 5TB. Tegenwoordig is “groen” ook een belangrijk argument en “Tape is groen”, omdat de data beschikbaar blijft zonder dat er stroom verbruikt wordt.

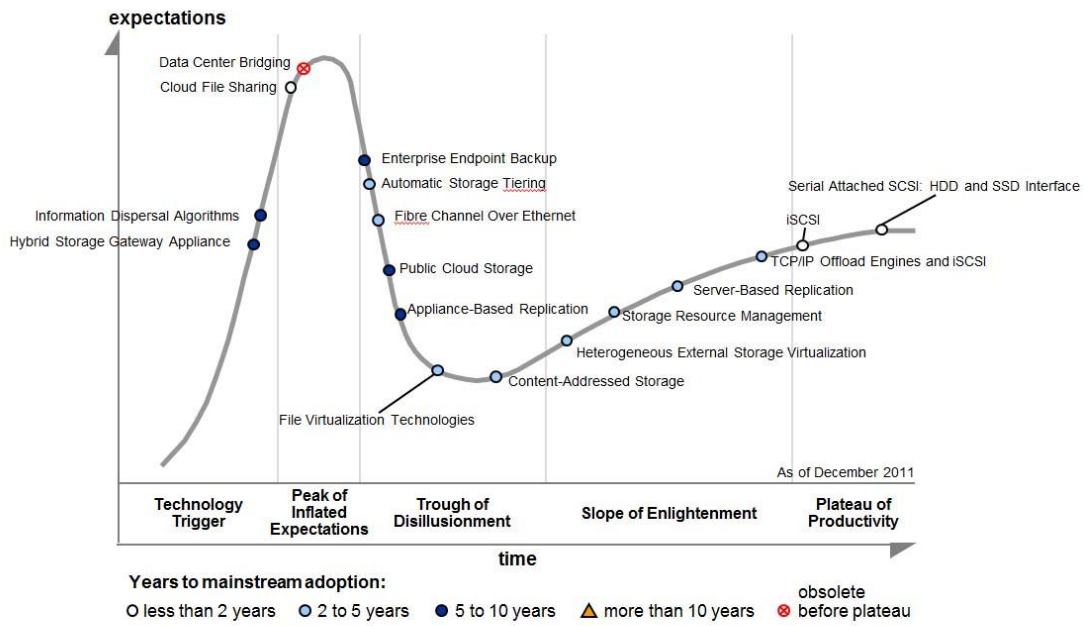
Het wel of niet inzetten van tape backup moet onderdeel zijn van een data life cycle management proces. Het is van belang dat instellingen goed bekijken wat voor type data binnen de instelling onderscheiden kan worden. Vervolgens kan bepaald worden welke mix van technieken zoals replicatie, snapshotting, backup naar disk of tape en tape archivering optimaal is.

4.8 Hype Cycle

In het figuur hieronder zijn verschillende cloudoplossingen en technieken te vinden uitgetekent door SURFnet in een Gartner Hype Cycle template. De Hype cycle laat zien hoe volwassen een technologie is. Aan het begin van de curve zijn technologieën nog niet volwassen, en aan het eind van de curve zijn de technologieën bijna mainstream, maar hebben ze slechts een kwart van de potentiële markt geopenetreerd. SURFnet heeft belangrijke storage technologieën, gebruik makende van het Gartner Hype Cycle template, gewogen en vervolgens afgebeeld als Hype Cycle.

In de Hype cycle is de cloud gesplitst in verschillende technologieën, zoals *public cloud storage* en *cloud file sharing*⁷. Beide technologieën zijn nog aan het begin van de curve en nog niet mainstream. Dat wil zeggen dat instellingen zich ervan bewust moeten zijn dat de techniek nog niet volwassen is en dat er (te) hoge verwachtingen zijn van de betreffende techniek. Bij het invoeren van een dergelijke techniek moet men zich bewust zijn van het feit dat dit additionele inspanningen vergt in vergelijking met meer volwassen technieken. Het toepassen van technieken die tegen de hype aanzitten kan bijdragen aan het innovatieve vermogen van een organisatie. Daarnaast straalt het qua PR positief uit. Het toepassen van mainstream technieken wil niet zeggen dat er geen problemen kunnen optreden, maar voor deze technieken is er volop expertise beschikbaar om eventuele problemen effectief te kunnen oplossen.

⁷ De specifieke definities van deze termen en andere termen op de hype cycle zijn in de appendix te vinden.



*This chart is a composite, derived from Gartner published Hype Cycles.
 The particular combination and comparison of items made here has not been reviewed by Gartner.*

Figuur 1: Storage Hype Cycle (gebaseerd op Gartner Hype Cycle Template)

5 De rol van de cloud

Er is de laatste jaren veel te doen over opslag in “de cloud”. De cloud is echter niet één oplossing, maar kent verschillende vormen die nauwelijks vergelijkbaar zijn; het is dan ook beter deze vormen als geheel verschillende oplossingen te bezien.

In het algemeen wordt voor Cloud diensten onderscheid gemaakt tussen de volgende diensten⁸:

- SaaS (Software as a Service);
- PaaS (Platform as a Service);
- IaaS (Infrastructure as a Service).

Bij het afnemen van een SaaS of PaaS dienst wordt de applicatie c.q. het platform als dienst afgenomen, inclusief de storage. Daarmee wordt er geen gebruik gemaakt van lokaal aanwezig storage, maar wordt er qua server hardware en storage volledig gebruik gemaakt van de capaciteit die de SaaS of PaaS aanbieder beschikbaar heeft. Bij PaaS en SaaS diensten is de storage functionaliteit niet zichtbaar voor de instelling, en heeft de instelling alleen met de geleverde dienst te maken. Dat kan overigens ook een nadeel zijn: de instelling heeft niet altijd goed inzicht in de risico's die de aanbieder neemt in relatie tot security, continuïteit, en privacy⁹.

IaaS is een vorm van cloud computing waarbij fysieke storage, storage netwerk en server hardware zijn gevirtualiseerd en bij een aanbieder ondergebracht. Hiermee wordt het mogelijk om virtuele computers en/of storage beschikbaar te stellen als cloud dienst. Een IaaS computing cloud zoals Amazon EC2¹⁰ is uitermate geschikt wanneer er tijdelijk behoefte is aan virtuele computers, bijvoorbeeld in het kader van een experiment of onderwijs situatie. Deze virtuele computers zijn niet altijd geschikt om daarop traditionele applicaties te installeren, onder andere omdat publieke IaaS leveranciers momenteel nog een slechte SLA bieden voor wat betreft de beschikbaarheid. Ook is de netwerkvertraging (latency) tussen gebruiker en IaaS cloud vaak te groot voor intensieve client/server applicaties. Een IaaS cloud is wel uitermate geschikt om cloud-aware applicaties op te draaien. Dat wil zeggen dat de applicaties er rekening mee houden dat individuele cloud-resources kunnen wegvallen en dan direct overgaan op andere, nog beschikbare resources. Cloud-aware applicaties volgen meestal een web services architectuur, die ook beter bestand is tegen latency tussen de verschillende componenten.

5.1 Cloud storage diensten

Ook IaaS storage diensten, zonder cloud computing, zijn door de hoge latency vooral geschikt voor cloud-aware architecturen. Dergelijke diensten (bijvoorbeeld de Amazon S3 dienst¹¹) bieden daarom meestal een web services interface.

Een bijzondere vorm van een IaaS storage dienst is een backup dienst vanuit de cloud. Backup in de cloud is een steeds gangbaardere manier om ervoor te zorgen dat er voldoende opslagruimte

⁸ Zie voor meer detail: *Met SURF in de Wolken*, SURF (2011), <http://www.surf.nl/nl/publicaties/Pages/MetSURFindewolken.aspx>

⁹ Enkele van deze elementen worden verder uitgewerkt in SURF publicaties, zie <http://www.surfsites.nl/cloud/interessante-links/>

¹⁰ Elastic Computing Cloud, zie <http://aws.amazon.com/ec2/>

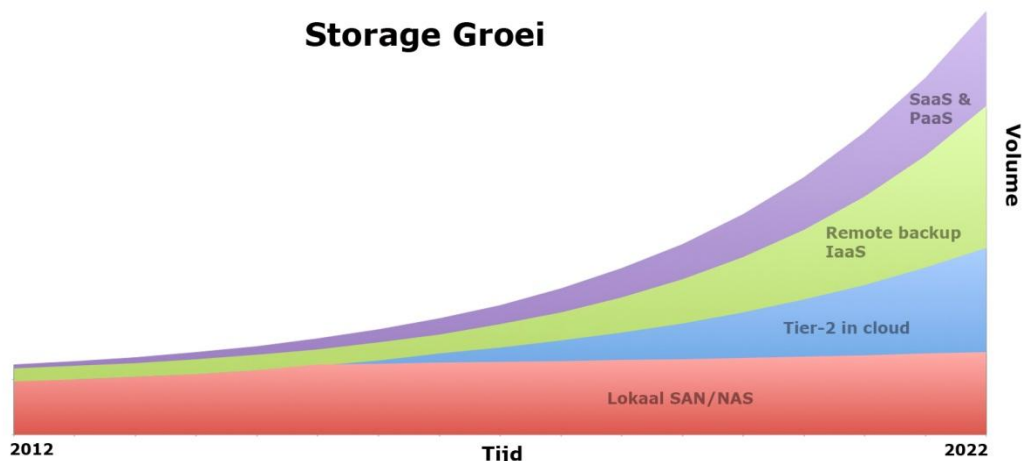
¹¹ Amazon Simple Storage Service, zie <http://aws.amazon.com/s3/>

beschikbaar is voor het reguliere backup proces. Om een backup naar een publieke cloudomgeving van bijvoorbeeld Amazon of Rackspace te maken is het primair van belang om een snelle internetverbinding beschikbaar te hebben. Daarnaast is de tooling van belang en is het belangrijk om de backups met afdoende encryptie in de cloud te plaatsen.

Cloud storage aanbieders bieden op grote schaal storage aan en zijn wat dat betreft zeer schaalbaar richting hun klanten. Deze storage is traag toegankelijk, maar wel geschikt voor backup processen. Qua kosten zijn met name de kosten voor bandbreedte en de hoeveelheid gebruikte storage van belang.

Naast de “pure” storage IaaS diensten zijn er combinaties beschikbaar, zoals bij Hybrid Storage Appliances. Bij Hybrid Storage Appliances biedt een “appliance” op locatie de storage dienst aan, terwijl deze appliance op de achtergrond de data naar een cloud provider repliceert. De appliance hoeft dan alleen nog maar de recent gebruikte gegevens vast te houden, en haalt de rest op verzoek op, terwijl recente gegevens wel snel beschikbaar zijn. Ook deze oplossing wordt op dit moment vooral voor backup doeleinden gebruikt.

In onderstaande figuur is een prognose geschetst van de storage groei in de komende jaren. Voor instellingen die de komende jaren een eigen datacenter behouden blijft het noodzakelijk om ook een lokale storage omgeving aan te houden, aangezien de meeste applicaties nog niet “cloud-aware” zijn. Uit de figuur wordt duidelijk dat de grootste groei de komende 10 jaar niet meer op deze lokale storage zal plaatsvinden, maar richting de cloud gaat. Het gebruik van remote backup naar de cloud is op dit moment al een beproefde methode, die breder ingezet zal worden. Het offloaden van data naar een tier-2 cloud omgeving, zoals bij Hybrid Storage Appliances, is op dit moment nog niet volwassen maar naar schatting over drie jaar wel en zal dan de lokale storage substantieel ontlasten.



Figuur 2: Prognose Storage ontwikkeling 2012-2022

5.2 SURF Cloud strategie

Binnen SURF verband is er in 2011, gezamenlijk met instellingen, gewerkt aan een cloud strategie. Deze strategie heeft als motto “cloud first”. Dat houdt in dat voor het afnemen van diensten er allereerst gekeken wordt of de publieke cloud een oplossing kan leveren voor de desbetreffende dienst. Indien de publieke cloud niet voldoet, bijvoorbeeld qua eisen aan functionaliteit, performance,

privacy of security, dan is het mogelijk om te onderzoeken of een community of private cloud een oplossing biedt.

Voor het hoger onderwijs en onderzoek zijn cloud-services onvermijdelijk. En hoewel sommige instellingen al op individueel niveau de vruchten plukken van het gebruik van dergelijke diensten, kunnen de voordelen op veel grotere schaal worden benut als instellingen in SURF-verband samen optrekken. Met instellingsoverstijgende samenwerking kan schaalgrootte worden gerealiseerd en inkoopkracht gebundeld. Tegelijkertijd is samenwerken dé manier om kennis en ervaring te delen, en met kracht de specifieke eisen en behoeften van het (Nederlandse) hoger onderwijs en onderzoek bij leveranciers neer te leggen.

Met andere woorden: een gemeenschappelijke SURF cloud-strategie biedt de mogelijkheid om gezamenlijk en kostenefficiënt activiteiten te ontplooiën in de community en de public cloud en op deze manier daadwerkelijk en op grote schaal te profiteren van de voordelen die de cloud te bieden heeft.

Tegelijkertijd is de cloud niet zaligmakend. SURF realiseert zich dat cloud niet dé oplossing voor alle problemen kan zijn, en waar het dat wel is werkt het alleen als we dit onderwerp gezamenlijk en op de juiste manier oppakken.

SURFmarket sluit contracten af met public cloud aanbieders zodat studenten, onderzoekers, docenten en medewerkers tegen zeer gunstige voorwaarden kunnen beschikken over relevante cloud-diensten. Met SURFconext kunnen gebruikers vervolgens met één identiteit – die van hun instelling – inloggen bij alle cloud-aanbieders die op SURFconext zijn aangesloten. En in samenwerking met de instellingen en internationale partners ontwikkelt SURFnet koppelingen tussen SURFconext en andere federatieve systemen, om optimale samenwerking ook over landsgrenzen mogelijk te maken. SURFnet voert ook diverse pilots uit op het terrein van cloud computing.

5.3 Aanbod kant van de cloud

De publieke cloud aanbieders zijn vaak internationaal, zoals Amazon, en zijn gericht op massaal gebruik. Een private of community cloud aanbieder biedt een meer op een specifieke groep gerichte storage oplossing aan, en is vaak lokaal gevestigd.

De meeste providers in Nederland bieden standaard en maatwerk storage diensten. Ze leveren primaire en secundaire storage tot tientallen terabytes (TBs), gebaseerd op standard componenten. Dit biedt schaalvoordelen voor de providers. Desondanks is maatwerk in de hardware ook vaak mogelijk, bijvoorbeeld voor grotere storage behoefte die op kunnen lopen tot meerdere Petabytes.

Naast storage diensten bieden de meeste providers ook additionele diensten zoals het ontwerpen van een integrale storage oplossing. Implementatie en datamigratie zijn ook onderdeel van het additionele dienstenpakket.

De connectiviteit van en naar de storage oplossing speelt een extra belangrijke rol bij cloudoplossingen. Sommige aanbieders regelen dan ook een verbinding tot aan het datacenter waar de data van de klant opgeslagen is, terwijl andere aanbieders zelfs hun eigen dedicated glasvezelring hebben gebouwd om verschillende datacentra in Nederland te verbinden met hun klanten.

Een groot aantal aanbieders biedt in Nederland grootschalige en maatwerk cloudoplossingen aan. Enkele voorbeelden:

- Proact¹² (NL),
- StoreData¹³ (NL),
- Vancis¹⁴ (NL),
- Dutch Cloud¹⁵ (NL),
- GreenQloud¹⁶ (IJsland),
- Rackspace¹⁷ (Gevestigd in de VS, maar vertegenwoordigd in GB),
- Amazon¹⁸ (Gevestigd in de VS, maar vertegenwoordigd in Ierland).

Naast deze bedrijven bestaan er ook storage aanbieders die alleen remote backup aanbieden en niet een gehele storage omgeving.

5.4 Privacy en veiligheid

Risico's met cloud computing¹⁹ zijn voor een deel vergelijkbaar met andere extern geleverde ICT-diensten. Belangrijke elementen zijn gegevenssegregatie, privacy, toegang, stabiliteit van de leverancier, beschikbaarheid en herstel, en data migratie. De locatie-onafhankelijkheid en de mogelijkheid dat een cloud provider zelf weer gebruik maakt van diensten van derden brengen voor cloud computing specifieke risico's met zich mee, vooral als het gaat om privacy. Om die reden kan de locatie waar gegevens worden opgeslagen van belang zijn. De Europese privacy regels (en de Nederlandse privacy wetgeving, die gebaseerd is op diezelfde regels) vereisen dat gevoelige gegevens worden opgeslagen:

Binnen de Europese Unie, in welk geval de Europese privacy-regels automatisch gelden;
Binnen een land dat dezelfde mate van bescherming biedt als de Europese privacy-regels (zoals vastgesteld bij een Europees besluit van de Commissie); dit betreft in elk geval de landen binnen het Europese Economische Gebied (EEA = European Economic Area), maar ook, Zwitserland, Canada, Argentinië, Guernsey en het eiland Man, en daarnaast de Verenigde Staten zolang de betreffende aanbieder de Safe Harbor-principes heeft ondertekend, of,

Binnen elk ander land, zolang er voldoende contractuele regelingen zijn getroffen om dezelfde mate van privacy als in de Europese Unie te garanderen.

Toch zouden veel van de bij SURFnet aangesloten instellingen bij voorkeur restrictiever willen zijn, door ervoor te zorgen dat hun gegevens worden opgeslagen binnen het Europese Economische Gebied of zelfs alleen binnen Nederland. In het bijzonder maken instellingen zich zorgen over de Amerikaanse Patriot Act, aangezien veel dienstverleners Amerikaanse bedrijven zijn. Deze wet geeft de Amerikaanse overheid het recht om toegang te eisen tot alle gegevens die zijn opgeslagen binnen de VS, maar ook alle gegevens die zijn opgeslagen in andere plaatsen door aanbieders die gevestigd zijn in de VS.

¹² <http://www.proact.nl/>

¹³ <http://www.storedata.nl/index.php/solutions>

¹⁴ Vancis (<http://www.vancis.nl/>) is eigendom van stichting SARA, maar biedt commerciële diensten aan.

¹⁵ <http://www.dutchcloud.com/private-cloud>

¹⁶ <http://www.greenqloud.com/>

¹⁷ <http://www.rackspace.nl/>

¹⁸ <http://aws.amazon.com/products/>

¹⁹ Zie ook de SURFnet publicatie "Cloud Security Checklist"
http://www.surfnet.nl/Documents/rapport_201012_Cloud_Security_checklist_v1.0.pdf

Nationale overheden binnen Europa hebben ook bepaalde rechten om gegevens in hun land op te eisen. De regels van deze landen worden echter geacht voldoende waarborgen te bieden voor de privacy, en in het algemeen hebben de bij SURFnet aangesloten instellingen meer vertrouwen in de Europese privacy waarborgen, en met name in de Nederlandse, in vergelijking met die elders in de wereld.

Naast de risico's kent cloud computing ook beveiligingsvoordelen. Door de schaalvoordelen zijn beveiligingsmaatregelen goedkoper te implementeren. Ook kunnen cloud providers beveiligingsmiddelen sneller en dynamischer inzetten. Tot slot beschikken cloud providers vaak over diepe inhoudelijke beveiligingskennis die aan de afnemerszijde soms ontbreekt.

Cloud providers leveren cloud diensten op basis van zeer complexe, gedistribueerde en gevirtualiseerde infrastructuren, waar meerdere klantgroepen gebruik van maken. Dit resulteert in een aantal uitdagingen op het gebied van risicobeheersing waar bestaande evaluatiemethodieken niet direct mee om weten te gaan.

5.5 Voor welke storage klassen is de cloud bruikbaar?

De komende jaren zullen instellingen die een eigen datacenter hebben altijd behoefte hebben aan tier-1 storage. Tier-2 en tier-3 storage zal daarentegen steeds vaker uit de cloud afgenomen worden. Voor tier-2 storage geldt dat de fysieke afstand tot de cloud wel belangrijk is voor de latency en dus voor de performance, met name bij de gebruikelijke block- en file-based storage vormen. Deze tier-2 storage zou daarom fysiek in Nederland of vlak over de landsgrens aangeboden moeten worden. Voor tier-3 is latency nauwelijks nog van belang, en kan de fysieke cloud storage zich op grotere afstand kunnen bevinden.

Voor de afname van cloud storage is het wel van belang dat er voldoende breedte end-to-end beschikbaar is.

De eerder genoemde Hybrid Storage Appliances kunnen het latency probleem oplossen, en zijn daarom in principe geschikt voor alle storage tiers. Ze zijn echter momenteel nog niet volwassen genoeg voor tier-1 storage behoeften; op dit moment worden ze daarom vooral voor back-up toepassingen gebruikt.

5.6 Exit strategie en migratie bij de cloud

Voor het uitbesteden van een essentieel onderdeel van de infrastructuur als storage naar een externe partij is het essentieel dat bij beëindiging van het contract, de data in een gangbaar formaat van de provider verkregen kan worden en/of dat de provider actief meewerkt aan een migratie naar een andere provider (of terug naar de instelling zelf). Dit vereist niet alleen technische functionaliteit, maar ook contractuele voorwaarden die ervoor zorgen dat de instellingen de controle over de gegevens houden. Een escrow regeling kan nodig zijn om dit te bereiken in het geval de dienstverlener failliet gaat.

Contractueel is het belangrijk dat er, bij het afnemen van een dergelijke dienst, een exit-strategie is. Bijvoorbeeld doordat de betreffende aanbieder alle data uit de IaaS cloud beschikbaar stelt in een open format. Ook de snelheid waarmee data gemigreerd kan worden is hierbij een belangrijk aspect, dat contractueel geborgd moet zijn.

6 Aanbevelingen

6.1 Stel de huidige eisen en wensen op functioneel niveau vast

Voordat een storage traject ingezet kan worden zullen de huidige wensen en eisen eerst in kaart gebracht moeten worden, zowel in termen van functionele eisen als in termen van hoeveelheden en performance. Dit bevat onder andere de netto hoeveelheid data en de geprognoseerde groei, de benodigde performance, en eisen aan redundantie en beschikbaarheid. Vervolgens zal per applicatie en per platform in kaart gebracht moeten welke randvoorwaarden deze aan de storage stellen.

Deze uitgangspunten dienen vervolgens als input voor de technische eisen en wensen die aan de verschillende storage systemen gesteld zouden moeten worden. Let wel op dat wensen vanuit de applicaties en platformen niet zonder meer tot eisen worden verheven – vaak zijn er alternatieven mogelijk. Een Windows server omgeving die block-based storage verwacht, kan bijvoorbeeld gevirtualiseerd worden, waardoor de server alsnog op file-based storage aangesloten kan worden. Dit soort overwegingen raakt de hele keten van gebruiker tot storage, en zal dus in samenhang gezien moeten worden.

Advies 1: *Breng eerst de wensen en eisen per applicatie en platform in kaart, en destilleer hieruit de technische wensen en eisen van het storage systeem.*

6.2 Ontwikkel een visie op storage

In de praktijk stellen verschillende toepassingen sterk verschillende eisen aan de bijbehorende storage. Dit kan te maken hebben met technische aspecten van de toepassing: vraagt deze om block-based of om file-based storage, met welke latency moet de storage toegankelijk zijn, gaat het vooral om throughput of om capaciteit, om meer lezen of meer schrijven, om sequentiële of willekeurige toegang, etc. Een database systeem heeft bijvoorbeeld andere eisen dan een file server, en zelfs voor een database systeem kunnen de eisen nog verschillen afhankelijk van de toepassing (bijvoorbeeld omdat sommige toepassingen regelmatig full table scans uitvoeren op grote tabellen, waardoor de toegang een heel ander karakter krijgt dan de meer gebruikelijke random access). Ook spelen bedrijfsaspecten een rol: kritische data kan een veel hogere beschikbaarheid nodig hebben dan andere data.

Indien één enkele storage oplossing alle toepassingen dient te kunnen bedienen, zal voor alle aspecten de meest ‘zware’ eis aangehouden moeten worden. Zonder verdere differentiatie kan dit leiden tot hoge kosten. Uiteraard is het mogelijk binnen een oplossing nog differentiatie aan te brengen (bijvoorbeeld door een deel van de data vaker te repliceren dan de rest, of door snelle en langzame opslag te combineren), maar de oplossing moet functioneel wel aan alle eisen voldoen, en veel configuratiekeuzes gelden voor de gehele installatie. Ook dan kan het voorkomen dat er – doordat verschillende toepassingen anders met storage omgaan – tegenstrijdige wensen ontstaan, waardoor er een compromis gesloten moet worden.

De afgelopen jaren is men dan ook enigszins teruggelopen van de eerdere trends om alle storage in één systeem te consolideren. ‘Best practice’ is veelal om meerdere storageoplossingen naast elkaar te gebruiken, ieder ingericht op de specifieke eisen van een toepassing of categorie van toepassingen. Soms betreft het meerdere installaties van hetzelfde type, maar verschillend geconfigureerd; vaak zijn het ook geheel verschillende oplossingen, bijvoorbeeld een SAN en een NAS omgeving afzonderlijk van elkaar.

Een combinatie van meerdere storage oplossingen heeft bovendien het voordeel dat ze niet gelijktijdig hoeven te worden aangeschaft, waardoor het mogelijk wordt om een nieuwe oplossing langzaam in te voeren terwijl kritieke data nog enige tijd op de eerder aangeschafte oplossing blijft staan. Hierdoor kunnen risico's beter gespreid worden. Nadeel is uiteraard dat er kennis in huis moet zijn van de verschillende systemen.

Vaak wordt er weinig rekening gehouden met het opstellen van een data life cycle beleid bij het ontwikkelen van een storage visie, terwijl het wel een belangrijk aspect is van storage. Afhankelijk in welke fase van de life cycle bepaalde data is hoort deze tot een bepaalde storage tier. Het inzichtelijk maken van de life cycles en het beleid geeft inzicht in hoe data gedistribueerd kan worden over de verschillende tiers.

Advies 2: Kies, op basis van een duidelijke visie, één integrale oplossing voor alle storage behoeften, versus verschillende oplossingen voor verschillende storage behoeftes.

6.3 Creëer overzicht in de IT architectuur

Van groot belang is dat er een goed overzicht is over de huidige situatie, niet alleen voor storage maar voor de gehele IT infrastructuur, en dat deze situatieschets ook als uitgangspunt kan dienen voor verdere discussies over aanpassingen en verbeteringen. De storage omgeving kan niet als een losstaande bouwsteen gezien worden, maar moet in samenhang met de rest van de omgeving bekeken worden.

De huidige situatie dient in kaart gebracht te worden aan de hand van een duidelijk *Architectuur-document*, waarbij een *Architectuur-plaat* het overzicht vormt van de verschillende systemen, en waar verder de verschillende onderdelen, en hun samenhang, in beschreven staan. Deze plaat zou alle onderdelen in hun context moeten bevatten.

Voor elk onderdeel dient additionele informatie aanwezig te zijn, bijvoorbeeld in de vorm van een bijlage bij de architectuurplaat. Hierin komen verdere details te staan als software en firmware levels, protocollen, etc., maar in het geval van SAN's ook de aantallen en types disks, de indeling van de spindles in pools / disk groups, allocatie van LUN's, WWN's en IP adressen. Veel van deze informatie kan met automatische tools gegenereerd worden, hetgeen het eenvoudiger maakt om de documentatie up to date te houden.

Vanwege het centrale belang van een goede storageomgeving voor de verschillende platformen, toepassingen en gebruikers is het van belang dat de architectuurrol omtrent de samenhang tussen storage en overige systemen goed wordt ingevuld.

Om dit te bereiken dient duidelijk te zijn waar de "architecten-rol" ligt. Deze persoon of personen dienen het overzicht te hebben over de samenhang tussen de diverse systemen. Daarbij dienen deze person(en) gemachtigd te zijn om, op basis van technische aspecten en gebruikerswensen architectuurkeuzes en beslissingen te maken. Het helpt hierbij als iemand de rol van Enterprise Infrastructure Architect toebedeeld krijgt.

Een goed overzicht van de bestaande configuratie is van belang om in een complexe omgeving waarin diverse systemen met elkaar samenhangen problemen te kunnen oplossen, en nieuwe oplossingen optimaal in te passen.

Hierbij dienen de architectuurkeuzes steeds expliciet gemaakt te worden. Een keuze voor block-based vs. file-based storage kan bijvoorbeeld alleen zinnig gemaakt worden aan de hand van de consequenties voor de gehele keten, en niet alleen voor de storage.

Advies 3: *Maak de IT-architectuur en de samenhang met storage inzichtelijk; beleg de rol van de Enterprise Infrastructure Architect.*

6.4 Samenwerking gebruikers, informatiebeleid en IT diensten

Er ligt een natuurlijk spanningsveld tussen de wensen en eisen van eindgebruikers, het informatiebeleid van een onderwijsinstelling, en het dienstenaanbod van de IT afdeling. Het is van belang dat er een balans ontstaat tussen dienstenaanbod, beleid en uitvoering. Hoewel de gebruikers als 'interne klant' natuurlijk hoge eisen zullen stellen, is het zaak om vanuit het IT beleid en in overleg met de IT afdeling te komen tot een realistische set eisen.

Het is van belang al in een vroeg stadium (bij een aanbesteding dus al tijdens het voortraject, waarin bepaling van het PvE door de diverse belanghebbenden plaatsvindt) de eisen duidelijk te krijgen, en waar mogelijk te komen tot een goed compromis tussen de eisen van gebruikers enerzijds en haalbaarheid en complexiteit anderzijds. Indien men uit de conflicterende belangen tussen de verschillende interne partijen niet een gezamenlijk uitgangspunt kan destilleren dan dient de eindverantwoordelijke hier een beslissing over nemen. Ook hierbij zal een architect een belangrijke rol moeten spelen.

Als blijkt dat de verschillende platforms niet-verenigbare eisen stellen, terwijl de gestelde eisen wel redelijk zijn, dan zal moeten worden gekeken naar een alternatieve oplossing, zoals het onderbrengen van de storage op verschillende platforms (die ieder dan voldoen aan een andere set van eisen).

Advies 4: *Zorg voor een balans tussen de wensen en eisen van eindgebruikers en het storage aanbod van de IT afdeling, gebaseerd op een expliciet IT beleid.*

6.5 Managen interne kennis voor Storage omgeving

Alhoewel binnen onderwijsinstellingen zeker kennis aanwezig is van de gebruikte storage omgevingen, is dit niet altijd voldoende om problemen te onderkennen en op te lossen. In veel gevallen zal een beroep op de leverancier gedaan moeten worden, maar ook om diens voorstellen te kunnen beoordelen is gedetailleerde kennis nodig.

Het is niet in alle gevallen nodig om dergelijke gedetailleerde kennis in huis te hebben, en dat is waarschijnlijk ook niet haalbaar. Belangrijk is wel dat onderwijsinstellingen toegang hebben tot onafhankelijke expertise, los van de leverancier, en er ook voor zorgen dat deze kennis op de juiste momenten beschikbaar is. De mate waarin externe kennis nodig zal zijn zal verschillen per organisatie en per platform.

Vanuit de organisatie dient tenminste voldoende mankracht en kennis beschikbaar te zijn om vast te kunnen stellen wanneer hulp of extra kennis nodig is, en of deze kennis binnen of buiten de onderwijsinstelling beschikbaar is, en om deze vervolgens aan te kunnen sturen.

Advies 5: *Bepaal welke strategische storage kennis intern bijgehouden en/of opgebouwd moet worden en bepaal wat uitbesteed kan worden.*

6.6 Samenwerking met andere instellingen

Storage is in de afgelopen jaren uitgegroeid tot een specialisatie. Het is kostbaar om binnen de IT afdelingen van elke instelling een brede storage expertise te onderhouden. Veelal hebben instellingen alleen expertise opgebouwd ten aanzien hun bestaande storage omgeving en ontbreekt er expertise over andere oplossingen; hierdoor wordt de aanschaf en ingebruikname van een nieuwe oplossing bemoeilijkt. Het is dan zinvol indien de bij SURF aangesloten instellingen expertise op dit vlak uitwisselen, en vooral indien deze expertise gebundeld beschikbaar kan worden gesteld.

Instellingen kunnen ook samenwerken door ongebruikte storage ruimte in hun datacentra aan te bieden als cloud storage, bijvoorbeeld via SURF. Sommige universiteiten stellen hun faciliteiten al beschikbaar; een voorbeeld daarvan is de Rijksuniversiteit Groningen (RUG). De RUG heeft in het kader van een aantal grote onderzoeksprojecten, waaronder Target, veel ervaring opgedaan met verschillende typen storage omgevingen. Met deze ervaring bouwt de RUG op dit moment een state-of-the-art storage omgeving die enerzijds de nieuwste technieken, zoals auto-tiering, ondersteunt en anderzijds geïntegreerd kan worden met de reeds aanwezige SAN/NAS storage en tape backup faciliteiten. Deze nieuwe storage omgeving is bijzonder schaalbaar. Dit maakt het zelfs mogelijk om deze omgeving te integreren met de storage omgeving van andere instellingen. De RUG staat er voor open om samenwerkingen met andere instellingen binnen de SURF doelgroep op storage vlak te verkennen.

Voor onderzoeksgelateerde storage kunnen instellingen bij SARA terecht. Als nationaal centrum voor High Performance Computing-diensten ondersteunt SARA onderzoekers in Nederland en werkt zij intensief samen met de academische gemeenschap en het bedrijfsleven. SARA heeft ten behoeve van onderzoeksprojecten een aantal storage omgevingen opgezet, waaronder een Grid based storage omgeving en een HPC Cloud storage omgeving. Met name voor onderzoeksprojecten waar veel storage benodigd is en/of gebruik wordt gemaakt van High Performance Computing maken onderzoekers gebruik van de storage diensten van SARA.

De aanschaf van storage verloopt veelal via een aanbesteding. Iedere instelling doorloopt hiervoor een vergelijkbaar traject. Het is interessant om nader te onderzoeken of een dergelijke aanbesteding onder SURFmarket vlag gezamenlijk getrokken kan worden. Hierbij zou initieel gedefinieerd moeten worden wat voor type storage de instellingen nodig hebben. Vervolgens zou SURFmarket in een dergelijk traject raamovereenkomsten met storage leveranciers, die voldoen aan de gebruikscriteria, af kunnen sluiten.

Voordeel van dit traject is dat instellingen niet meer zelf een storage aanbestedingstraject hoeven te doorlopen en dat de door SURFmarket gebundelde storage expertise meegenomen kan worden in de aanbesteding.

Op dit moment hebben de meeste instellingen één of meerdere eigen datacenters in gebruik. De verwachting is dat het totale aantal datacenters zal afnemen door samenwerking tussen instellingen, door een toenemend gebruik van commerciële SaaS diensten (bijvoorbeeld Blackboard als dienst, in plaats van een eigen Blackboard omgeving), en door toenemend gebruik van externe datacentra, met name voor redundantie (waardoor er geen tweede datacenter meer nodig is). Ondanks deze afname

is de verwachting dat de grotere instellingen de komende vijf à tien jaar nog wel minstens één eigen datacenter zullen behouden.

In een datacenter is er, bij de huidige stand van de techniek, altijd een noodzaak om tier-1 storage lokaal te houden. Dit komt met name door de gevoeligheid voor latency, die bij enkele milliseconden al voor ernstige performance problemen zorgt. Voor tier-2 is het echter qua performance acceptabel dat de storage nodes ergens anders in het land geplaatst zijn. Instellingen zouden dan ook kunnen onderzoeken hoe zij gezamenlijk een dergelijke tier-2 storage kunnen opzetten, dan wel afnemen bij een marktpartij. Er zijn al technieken beschikbaar die een dergelijke samenwerking kunnen faciliteren, en het is aan te bevelen om een pilot op te starten om enerzijds de requirements helder te krijgen voor een dergelijk tier-2 gebruik en om anderzijds de techniek te beproeven.

Ten aanzien van tier-3 storage geldt dat het maken van (encrypted) backups naar een (publieke) cloud omgeving al een beproefde methode is, waardoor het qua techniek niet meer noodzakelijk is om hiervoor lokale voorzieningen aan te houden. Wel zijn er soms nog contractuele of prijstechnische redenen om dit voor de korte termijn nog lokaal te doen. De verwachting is dat binnen vijf jaar de meeste instellingen gebruik zullen maken van cloud technieken voor het maken van backups.

Advies 6: *Onderzoek mogelijkheden tot samen te werken binnen de SURF community..*

6.7 Inkoopproces

Ongeacht welke storage systemen een instelling uiteindelijk kiest is het belangrijk om altijd op bepaalde zaken te letten bij het inkopen en/of het uitbesteden van storage. Deze adviezen gelden ook voor de keuze voor cloudoplossing.

Belangrijke elementen zijn daarbij:

- *Migratie:* een goede storage oplossing op zichzelf is niet voldoende, de oplossing moet ook nog geïmplementeerd worden en de data moet er naar toe gemigreerd worden. In een later stadium zal de data er ook weer vandaan gemigreerd moeten worden naar een toekomstige nieuwe oplossing. Belangrijk is dan ook dat de oplossing goede migratietools biedt, maar ook dat de aanbieder bereid is hier een actieve rol in te spelen (tenzij de instelling zelf voldoende mankracht en expertise heeft, hetgeen vaak niet het geval zal zijn). Datamigratie is bijzonder relevant bij cloud diensten: bij veel van deze diensten is de reguliere interface niet geschikt (of bijzonder duur) voor grote migraties, en zal er dus een apart mechanisme ingericht moeten worden voor de migratie (bijvoorbeeld door het verzenden van grote aantallen disks of tapes). Daarbij is het met name belangrijk om contractueel vast te leggen hoe na beëindiging van het contract de migratie naar een nieuwe oplossing ondersteund zal worden.
- *Acceptatietesten:* hoe goed de oplossing op papier ook klinkt, er is altijd een uitgebreid testtraject nodig om zeker te stellen dat de oplossing ook in de praktijk, in de specifieke situatie bij de instelling, goed werkt. In de praktijk treden hier vaak complicaties op, waarbij bepaalde combinaties van hard- en software niet werken, ingewikkelde work-arounds nodig maken, of een veel lagere performance leveren dan in theorie haalbaar zou moeten zijn. Een testtraject voor een nieuwe storage oplossing kan dan ook al snel enkele maanden duren, met meerdere firmware en software updates gedurende het traject om alles werkend te krijgen. Goede tools voor het simuleren van zwaar gebruik en het monitoren van de performance over de hele keten zijn essentieel voor een goed testtraject.

- *Contracten en prijzen:* de prijzen op de markt voor een gegeven hoeveelheid storage capaciteit dalen snel. In het algemeen is het dan ook niet efficiënt om bij de aanschaf van een nieuwe omgeving al meteen de capaciteit in te richten die over enkele jaren nodig zal zijn; beter is het om met een kleinere configuratie te beginnen en later uit te breiden. Dan is het wel belangrijk dat de prijsstelling in het contract ook waarborgt dat de toekomstige uitbreidingen tegen de dan geldende marktprijzen aangeschaft kunnen worden; een contract met vaste prijzen is dus al snel ongunstig. Beter is het om een contract uit te onderhandelen met een prijspeil dat mee daalt met de markt.
- Naast deze specifieke aandachtspunten gelden natuurlijk ook hier de algemene aandachtspunten voor het verwervingsproces, die te vinden zijn in de SURF Sourcing Toolbox²⁰.

6.8 Structureel test- en acceptatie inplannen

Ook na acceptatie van de storageomgeving zal een instelling regelmatig nieuwe combinaties van hard- en software moeten testen. De meeste instellingen zullen niet voor elk platform een volledige OTAP straat (Ontwikkeling, Test, Acceptatie en Productie) willen aanschaffen, maar het is wel belangrijk om een representatieve testomgeving te onderhouden waarin nieuwe configuraties getest kunnen worden. Daarbij moet er in de testomgeving een vergelijkbare storage monitoring plaats vinden als in de productieomgeving. Eventuele verbeteringen of verslechtingen kunnen dan snel gedetecteerd worden.

Tevens dienen instellingen een goede policy voor test en acceptatie van nieuwe hard- en software te hanteren. Hierin dienen de procedures vastgelegd te worden over hoe om wordt gegaan met updates, hoe deze getest worden, en onder welke omstandigheden een update goed dan wel af wordt gekeurd.

Een dergelijke structurele manier van testen biedt de mogelijkheid specifieke configuratieproblemen vroegtijdig op te sporen, en de daadwerkelijke updates zo in te richten dat deze met minimale risico's voor de dienstverlening kunnen worden doorgevoerd.

²⁰ Zie <http://www.surfgroepen.nl/sites/toolboxsourcing>.

7 Appendix: storagedefinities

Term	Definitie
Appliance-Based Replication	Synchrone of asynchrone replicatie die vanuit de storage appliance plaatsvindt.
Automatic Storage Tiering	Auto tiering is het proces waarbij de storage omgeving een algoritme gebruikt om te bepalen op welke tier de data geplaatst moet worden. De data kan vervolgens, afhankelijk van het toepassing, dynamisch worden verplaatst tussen de verschillende tiers.
Cloud File Sharing	Dit zijn de cloud diensten die het mogelijk maken om bestanden te delen tussen verschillende typen devices. Dropbox is hiervan een bekend voorbeeld.
Content-Addressed Storage	Content-addressed storage is een object gebaseerde scale-out storage omgeving die primair in een LAN of campus omgeving wordt ingezet.
Data Center Bridging	Dit betreft een collectie van standaarden die tot doel hebben om een efficiënt en lossless Ethernet netwerk te creëren.
Enterprise Endpoint Backup	Enterprise Endpoint Backup is de backup oplossing voor alle devices, zoals laptops, desktops, tablets en smartphones.
Fibre Channel Over Ethernet	Is een nieuwere Fibre Channel techniek waarbij gebruik wordt gemaakt van Ethernet technology.
File Virtualization Technologies	File virtualization technologies biedt een virtueel filesysteem bovenop onderliggende heterogene storage omgevingen.
Heterogeneous External Storage Virtualization	Een virtualisatielaag die een homogene storage omgeving bouwt bovenop heterogene block based storage appliances.
Hybrid Storage Gateway Appliance	Een appliance die enerzijds primaire storage lokaal aanbiedt en anderzijds bepaalde data offload naar een storage cloud.
Information Dispersal Algorithms	Information dispersal algorithms is een techniek waarmee data uitgesplitst kan worden naar diverse nodes. Niet alle nodes hoeven beschikbaar te zijn om de oorspronkelijke data te kunnen reconstrueren. Veelal in gebruik bij P2P.
iSCSI	Internet Small Computer System Interface (iSCSI) is een block based protocol
Metro-Area SANs	Metro-area storage area networks (SANs) is een SAN omgeving waarbij de afstand tussen de datacenters tussen de 10 en 100 kilometer is.
Public Cloud Storage	Public Cloud storage is een storage omgeving met cloud kenmerken (pay per use, scalability etc).

Term	Definitie
SAN Management	Storage area network (SAN) management software biedt real-time monitoring, alerting en historische informatie.
Serial Attached SCSI: HDD and SSD Interface	Serial Attached Small Computer System Interface (SCSI) — SAS — is een enterprise-class interface voor hard-disk drives.
Server-Based Replication	Server-based replication software repliceert op server niveau de data via een block-based of filebased manier.
Storage Cluster File Systems	Storage cluster file systems zijn producten waarmee een single namespace filesystem bovenop verschillende storage nodes gecreerd wordt.
Storage Resource Management	Storage resource management (SRM) software biedt informatie over het gebruik van de storage omgeving.
TCP/IP Offload Engines and iSCSI	Een TCP/IP offload engine (TOE) is een NIC of een HBA die ervoor zorgt dat de CPU ontlast wordt en de performance omhoog gaat.