DE IT-DIEPTE IN 60 JAAR DISK OPSLAG

Dhr. J. (John) Stroonsnijder, Innovatie Manager, DMO/JIVC/KIXS

Gaat de harddisk met pensioen?

Misschien heeft u het gemist, maar de *hard disk drive* (HDD) heeft vorig jaar de respectabele leeftijd van 60 jaar bereikt[1]. Al die jaren heeft de harddisk ons enorm geholpen met opslagvraagstukken en daarbij heeft de ontwikkeling in de afgelopen 60 jaar zeker niet stilgestaan. Met zicht op de pensioengerechtigde leeftijd, begint de harddisk terrein te verliezen ten opzichte van nieuwere technieken gebaseerd op flash-geheugen. ➔

Is het einde van de magnetische opslag in zicht? En zo ja, wat kunnen we de komende jaren op het gebied van opslag verwachten? Om hier antwoord op te kunnen geven is het goed om eerst eens terug te kijken naar de geschiedenis en ontwikkeling van de harddisk, om daarna te kijken welke kant de techniek op gaat.

De ontwikkeling van de harddisk

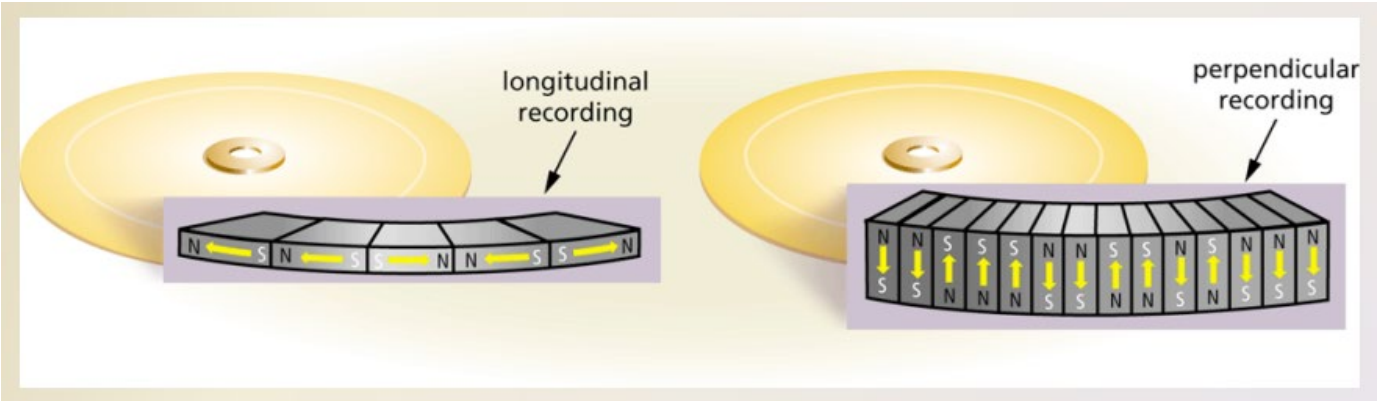
Op 14 september 1956 publiceerde IBM een persbericht met de aankondiging van de RAMAC 350 *(Random Access Method of Accounting and Control)*. Deze eerste commerciële harddisk was een enorm kamervullend apparaat van 1000 kg, uitgevoerd met een gigantische elektromotor die verbonden was met 50 magnetische dubbelzijdige schijven in een grote trommel. Dit apparaat had bij de introductie een opslagcapaciteit van 'slechts' 5 miljoen tekens (≈5 MB). Het opzoeken van een sector op deze disk duurde ongeveer 600 ms.

Het apparaat was alleen via lease beschikbaar en omgerekend naar de huidige prijzen kostte dit zo'n $ 250.000 per jaar. Een aardig bedrag voor een schijf, waarop (ter vergelijking met de huidige behoefte aan opslagcapaciteit) nauwelijks een paar muziekbestandjes pasten. Vervoer werd gedaan door middel van vrachtvliegtuigen en bij verplaatsing op locatie kwam er een vorkheftruck aan te pas. De harddisk is door de jaren heen verder ontwikkeld. De snelheden zijn omhoog gegaan en de capaciteit is enorm toegenomen.

Capaciteit van de harddisk

De magnetische opnametechniek is door de jaren heen sterk verbeterd waardoor veel meer data geschreven kan worden per vierkante centimeter disk. Initieel werd gebruikt gemaakt van de *longitudinal recording* technologie, waarbij de data (bitjes) achter elkaar op de disk werden geschreven.

Later werd een nieuwere opnametechniek geïntroduceerd genaamd *perpendicular recording*, waarbij de bitjes naast elkaar worden geschreven. Inmiddels worden ook nieuwe technieken beproefd, zoals *shingled recording* en *heat assisted magnetic recording* (HAMR)[2], waarmee de dichtheid verder toe kan nemen[3]. Dit zijn technische hoogstandjes. Bijvoorbeeld bij de HAMR technologie wordt het diskoppervlak met een la-



ser verwarmd, zodat de opname van de magnetische lading beter mogelijk is en de dichtheid toe kan nemen. Maar ondanks al deze ontwikkelingen lijkt er een einde te komen aan de rekbaarheid van de techniek. Er komt een punt dat er gewoonweg niet meer capaciteit op een disk kan[4]. Dit punt is nog niet bereikt, maar het einde is wel in zicht.
De groei is nu nog maar 10 tot 20% per jaar, waarbij de ontwikkeling dus niet meer gelijk oploopt met *Moore's law*[5], waarbij iedere 18 tot 24 maanden een verdubbeling in beeld is. De snelheid van de HDD's neemt daarbij nauwelijks toe, waardoor de harddisk een beperkende factor gaat worden voor snellere servers die wel in lijn met de wet van Moore doorgroeien.

Snelheid van de harddisk

De snelheid van een harddisk is van diverse factoren afhankelijk, maar het belangrijkste is wel de rotatiesnelheid van de disk. Des te sneller de disk draait, des te eerder de lees-/schrijfkop bij de juiste sector is en de bits gelezen of geschreven kunnen worden. De rotatiesnelheid is door de jaren heen toegenomen tot 15.000 toeren. Het voordeel van deze hoge snelheid is de lage *rotational latency*, oftewel de tijd die nodig is om de kop van de harddisk op de juiste plek te krijgen. Er is veel gebeurd sinds de eerste harddisk met 600ms latency, zoals de onderstaande tabel al aangeef.

Al met al is er vanaf de eerste disk (600ms) tot de snelste disks van nu (2ms) zo'n factor 300 snelheidswinst geboekt! En dat is goed, want des te minder wachttijd tussen de lees- en schrijfsessies, des te meer data verwerkt kan worden door het systeem.
Aan de rotatiesnelheid zit echter ook een grens. De eerder genoemde 15.000 toeren disks, zijn het hoogst haalbare. Om deze snelheid te halen moet al een compromis gemaakt worden. De disks mogen namelijk niet groter zijn dan 2,5 inch, omdat de snelheid aan de randen van een 3,5 inch disk gewoonweg te hoog zou worden. Sterker nog, deze zou de geluidsbarrière doorbreken (voor de liefhebbers: ruim 400 m/s) als ze op 15.000 toeren draaien. Een interessant weetje voor de mensen die weleens een een 15.000 toeren disk hebben gekocht: Er werden wel 3,5 inch disks als 15.000 toeren-disk verkocht, maar deze waren intern toch echt maar 2,5 inch. Deze disks waren weliswaar sneller, maar er kon minder data op. De snelheid gaat dus ten koste van capaciteit.

HDD's: huidige stand van zaken

De nieuwe opnametechnieken zijn echt een klein wonder van de techniek, maar de ontwikkeling en productie van dit type disk is te duur en complex geworden. Uiteindelijk doet de opkomst van *solid state disks* (SSD) de HDD de das om. De eerste tekenen zijn er al: de grote harddiskproducenten ontwikkelen de 15.000 toeren disks al niet meer verder sinds 2016. Andere disk types wachten op termijn waarschijnlijk hetzelfde lot.

De opvolgers van de harddisk

Al een paar jaar heeft de harddisk concurrentie van de *solid state drives* (SSD). SSD's zijn gegevensdragers die data elektrisch in plaats van magnetisch opslaan op een medium waarbij iedere sector direct is te adresseren. Dit heet *bit addressable memory*. Omdat het geheugen niet vluchtig is (zoals DRAM) wordt dit type opslag ook wel *non-volatile memory* genoemd. De huidige generatie SSD's maakt voornamelijk gebruik van NAND-flash[6] geheugen.
Er is dus geen sprake meer van draaiende delen. De latency is daarom zeer laag en doet zelfs de duurste draaiende disk verbleken.

Rotatie snelheid	Gemiddelde latency	Toepassing
15.000 toeren	2 ms	High-end enterprise disks
10.000 toeren	3 ms	Enterprise disks
7.200 toeren	4,2 ms	Consumenten disks
5.400 toeren	5,6 ms	Instap model consumenten disks

Toch is de introductie van de SSD's op basis van flashgeheugen niet zonder slag of stoot gegaan. De techniek kende zeker in het begin wat kinderziektes. Veel leveranciers van traditionele disks en de leveranciers van de grote disk-array's speelden hier maar al te graag op in door veel Fear, Uncertainty en Doubt (FUD) te verspreiden. Echter, die tijd is voorbij. SSD's zijn inmiddels heel volwassen. Sterker nog, de opvolgers komen er ook al aan en die zijn nog sneller!

Solid state drives op basis van NAND-flash geheugen

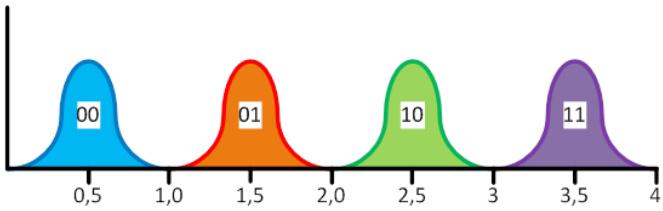
SSD's op basis van flashgeheugen zijn er in vele soorten en maten. De voordelen liggen enerzijds in de geweldige performance ten opzichte van de traditionele mechanische disks en anderzijds in de opslagcapaciteit die potentieel ver voorbij gaat aan de traditionele mechanische disks. 60TB SSD's zijn al aangekondigd en deze techniek blijft voorlopig wel in lijn met Moore's law doorgroeien, in tegenstelling tot de mechanische disks die tegen hun einde aan zitten.

Daarnaast is het energieverbruik veel lager en kunnen SSD's letterlijk tegen een stootje. Bijkomend voordeel is dat de storagebeheerders veel minder tijd kwijt zijn aan performance management. Het performancemanagement is met traditi-

onele draaiende disks veelal een hoofdpijndossier, maar wordt met SSD's vele malen eenvoudiger omdat er gewoonweg een overvloed aan performance is.
Er kleven ook nadelen aan het gebruik van flashgeheugen. Het is gebleken dat de cellen na verloop van tijd en bij veelvuldig gebruik minder goed gaan werken en langzaam elektrische lading verliezen. Ondanks dat er geen bewegende delen zijn, is er toch sprake van slijtage. Digitale slijtage wel te verstaan. Verder bleek het management van de geheugencellen in relatie tot de digitale slijtage een groot aandachtspunt. De eerste SSD's gingen hier slecht mee om en daarbij stortte de performance na verloop van tijd in. De disk was dan meer bezig met het intern reorganiseren van de cellen dan met het afhandelen van lees- en schrijfopdrachten. Dit is door betere samenwerking met de besturingssystemen en aanpassingen in filesystemen (o.a. met behulp van TRIM en SCSI UNMAP[7]) sterk verbeterd. Het probleem met de digitale slijtage blijft echter wel op de achtergrond spelen, echter de levensduur van een SSD is dermate lang en de slijtage is daarbij heel voorspelbaar (en meetbaar), zodat dit geen probleem meer hoeft te zijn.

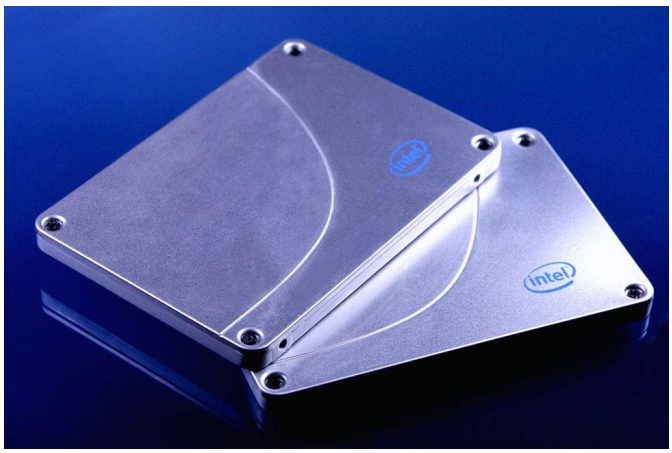
SLC en MLC

Op hoofdlijnen zijn er twee varianten SSD's te koop: *Single Level Cell* (SLC) en *Multi Level Cell* (MLC). SLC disks zijn de snelste en meest betrouwbare variant. Bij SLC wordt altijd maar één waarde (een bit) in een cell geschreven. Een cell is dus een 'nul' of een 'één'. Dit wordt weergegeven met twee voltages: laag of hoog. Bij MLC wordt niet één waarde in de cell opgeslagen, maar meerdere waardes. Het verschil in de waardes wordt gemaakt door het voltage. Dit komt de opslagcapaciteit natuurlijk te goede. De eerste MLC's sloegen twee bits per cell op, maar inmiddels zijn er ook varianten met drie (*Triple Level Cell* (TLC)) en zelfs vier waardes (*Quad Level Cell* (QLC)). Iedere combinatie van bitjes heeft een eigen voltage.



SSD met 2 bits per cell, waar 4 voltages bij horen

Bijgaande illustratie geeft dit weer voor een SSD met twee bits per cell, waar vier voltages bij horen. Er is één voltage per unieke combinatie van bits: de impact van de digitale slijtage van de cellen is verschillend per type disk. Als een MLC cell zijn lading verliest, is de afstand tot een andere waarde relatief klein. Een waarde '10' kan dan al snel gezien worden als '01', met alle gevolgen (data corruptie!) van dien.
Bij SLC drives is dit probleem het kleinst, omdat hier maar twee voltages worden gebruikt: laag en hoog. Bij MLC drives, zeker bij TLC en QLC SSD's wordt dit probleem natuurlijk alleen maar groter naarmate er meer bits en dus meer voltages



in één cell kunnen voorkomen. Gelukkig is de kwaliteit van SSD's tegenwoordig dermate goed, dat zelfs een consumenten-SSD bij normaal gebruik heel lang mee gaat[8]. Voor professioneel gebruik zijn de SLC's nog steeds de beste keus als het gaat om kwaliteit en performance, maar de enterprise-MLC's (eMLC's) doen het ook zeer goed. Daarbij is de prijs van eMLC veel gunstiger dan die van SLC's. Belangrijk bij eMLC's is om te kijken naar de *Drive Writes Per Day* (DWPD) waarde binnen de garantietermijn. Leveranciers geven bijvoorbeeld aan dat een MLC drive 3 keer per dag volledig overschreven (DWPD = 3) kan worden gedurende 5 jaar zonder problemen. Dat is zeer veel, in de praktijk doet niemand dat!

SSD doorontwikkeling

De ontwikkeling van SSD's staat niet stil. Er wordt gewerkt aan snellere *interfaces* vanwege de hoge snelheid van de SSD's en daarnaast wordt gekeken naar nog snellere geheugentechnieken.

IO-interfaces

SSD's zijn initieel als alternatief voor de traditionele mechanische harddisk gepositioneerd. Om die reden werden ze uitgerust met een SATA of SAS interface, zodat ze in iedere bestaande computer, laptop of server gebruikt konden worden zonder aanpassingen. Omdat de SSD's potentieel meer performance konden leveren dan de SATA en SAS interfaces, zijn alternatieven gezocht in een directe koppeling op de PCI-express bus, dichter bij de CPU's.

Om dit probleem voor te blijven, kwamen er op PCIe gebaseerde IO-kaarten op de markt met fenomenale performance cijfers, zoals de kaarten van FusionIO. Omdat deze kaarten direct in de PCIe-sloten van de servers werden gestoken, was de performance optimaal. Deze kaarten waren geen complete vervanger voor een HDD of SSD. Een systeem kan bijvoorbeeld niet opstarten (boot proces) van een dergelijke PCIe kaart, omdat er eerst een driver geladen moet worden. De performance was echter wel heel goed: zeker 10 keer sneller dan een normale SSD. Daar waar performance de doorslag geeft, komen dergelijke devices dus in beeld. Er zijn inmiddels kleinere standaarden ontwikkeld, initieel in de vorm van

de mSATA disken, die nog wel een SATA controller vereisten. mSATA wordt nu opgevolgd door M.2 (spreek uit als M-dot-two). De M.2 standaard ondersteunt ook de nieuwe *Non-Volatile Memory express* (NVMe) standaard, die van belang is voor verdere ontwikkeling van SSD technologie.

De prijs van SSD

Capaciteit en kwaliteit zijn geen issue meer voor SSD's. Maar hoe zit het met de prijs ten opzichte van mechanische disks? Zeker in de begintijd van de SSD's, moest je voor een kwalitatief goede SSD diep in de buidel tasten. Dit is lang een argument geweest om niet naar *full-flash* te gaan maar hooguit SSD te combineren met traditionele draaiende disks in een hybride model. In 2016 is echter een kantelpunt ontstaan, waarbij de prijsverschillen tussen SSD en de enterprise traditionele disks vrijwel zijn verdwenen. Samsung verwacht dat dit in 2020 ook op zal gaan voor consumenten-disks.
SSD's gaan de plek van mechanische disks de komende jaren dus innemen, zeker nu naast de capaciteit en performance ook de prijs in het voordeel begint te werken.

Andere vormen van niet-vluchtig geheugen

Naast SSD's op basis van NAND-flash, die gekoppeld worden op een SATA of SAS interface wordt ook gewerkt aan opslag op basis van geheugentechnologie die lijkt op het RAM geheugen van computers en servers. De snelheid van dergelijke hardware is zo hoog dat al niet meer naar SATA of SAS wordt gekeken, maar voornamelijk naar op PCIe gebaseerde (NVMe) devices of zelf directe interfaces met de CPU's. RAM geheugen is sneller dan NAND-flash-geheugen en heeft vrijwel geen last van de digitale slijtage die bij NAND-flash na verloop van tijd optreedt.

RAM heeft echter één groot nadeel: het is vluchtig. Als de stroom wegvalt, is het geheugen leeg. Dit is natuurlijk niet acceptabel voor opslag van data. De

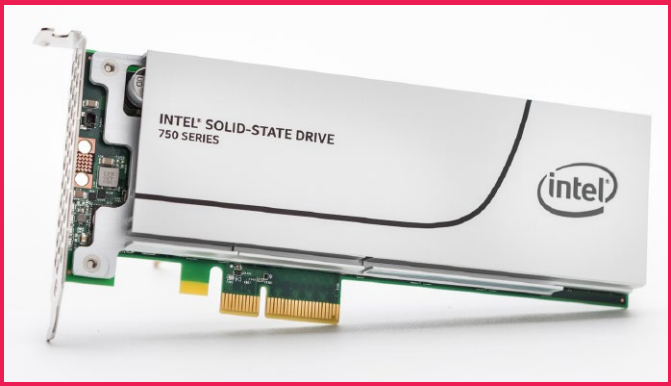
markt werkt daarom aan niet-vluchtig RAM geheugen om de afstand tussen de snelheid van RAM en SSD's kleiner te maken. Hierbij zijn verschillende technieken in beeld, zoals *Non-Volatile RAM* (NVRAM), *Storage Class Memory* (SCM) en 3D XPoint. Het gebruik van dergelijk niet-vluchtige RAM geheugen is weer een grote stap vooruit. De latency van dergelijk geheugen is weer vele malen lager dan de latency op SSD's. De kloof tussen de snelheid van de processoren en de opslag wordt zo gedicht, zoals onderstaande tabel aangeeft:

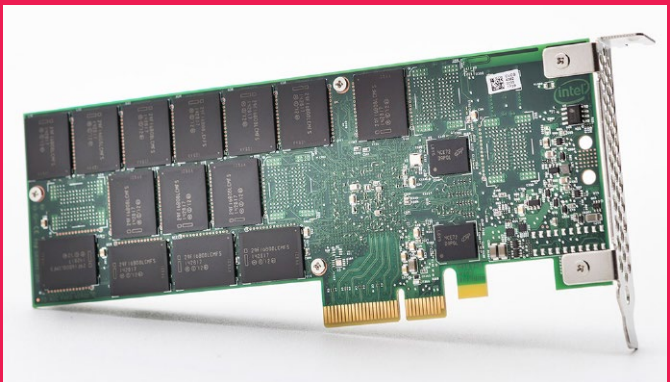
Type geheugen/opslag	Snelheid gemeten in	Relatieve snelheid
CPU cache	Enkele nano secondes (1 ns)	1
DRAM (vluchtig geheugen)	nano secondes (≈ 100 ns)	x 100 trager
SSD op basis van Non volatile RAM	Microseconden (≈20 µs)	X 20.000 trager
SSD op basis van Flash geheugen	Microseconden (≈200 µs)	X 200.000 trager
Mechanische harddisk (15.000 rpm)	Milliseconden (≈2 ms)	X 2.000.000 trager

2 ms lijkt niet veel, maar voor een computer lijkt het wel een eeuwigheid. Bij gebruik van mechanische disks staat de CPU daarom het overgrote deel van de tijd te wachten op reactie van de harddisks. De toegangstijd van disks is immers miljoenen keren trager dan de toegangstijd van de CPU cache. Met SSD's wordt dit zeker factor 10 beter. Met Non-volatile RAM technieken wordt een nog veel grotere stap gemaakt. Er komt dus nog een hoop moois aan.

Samengevat

De harddisk is inmiddels meer dan 60 jaar oud. De techniek is in de afgelopen 60 jaar zeer verfijnd en kwalitatief enorm ontwikkeld. Echter, na 60 jaar is de meeste rek er wel uit. SSD's op basis van NAND-Flash gaan de plek van harddisks innemen. Ze zijn veel sneller, de capaciteit groeit voorlopig in gelijke tred met de wet van Moore door en ze gebruiken minder energie. Storagebeheerders kunnen straks voor het eerst rustig achterover leunen en met vertrouwen garanties op performance afgeven. Daarnaast is de betrouwbaarheid nu





voldoende en is de prijs door het toenemende gebruik sterk gedaald. Naar verwachting zal rond 2020 een SSD net zo duur zijn als een mechanische harddisk. Maar op het gebied van SSD's liggen er al nieuwe ontwikkelingen klaar.

Diverse niet-vluchtige geheugentechnieken gaan de komende periode hun intrede doen en zullen mogelijk de huidige generatie SSD's weer vervangen. Onze oude vertrouwde harddisk zal echter niet zonder slag of stoot weggaan.
Wel is het aannemelijk dat de mechanische harddisk anders wordt ingezet. Zeker op het gebied van backup- en archiefopslag kan de oude vertrouwde harddisk nog wel nuttig gebruikt worden (voor zolang dat stand houdt dan).

Er zal een verschuiving komen in de verschillende opslagmedia, waarbij ook een andere oude techniek (tape) langzaam een stapje terug zal moeten doen:

Nu	2020
15.000 toeren hard disk	Non-volatile memory op basis van NVMe
10.000 toeren harddisk	SLC of eMLC SSD's
7.2000 toeren harddisk	MLC (inclusief TLC/QLC) SSD's
Tape voor backup (on-site)	7.200 toeren harddisk (>8 TB per disk)
Tape voor backup (offsite)	Cloud object storage (encrypted)

En what about cloud?

Maar waarom zou je je druk maken over disktechnieken, als je alles uit de cloud kan halen? Opslag is net zo makkelijk te consumeren als water uit de kraan, toch? Mee eens, daar gaan we wel naar toe.

Maar hoe denk je dat de grote *cloud providers* hun diensten maken? Dat komt echt niet uit de kraan, maar staat gewoon ergens in een datacenter te draaien. Inclusief harddisks of SSD's. De grote *cloud providers* houden de balans tussen prijs en kwaliteit scherp in de gaten.
Sommige diensten zijn al over op SSD en bij andere kun je kiezen of je traditionele of SSD opslag wilt. Amazon als voorbeeld heeft al een paar jaar de keuze uit meerdere op SSD gebaseerde storage diensten en biedt nu ook storage op basis van NVMe aan. Voor *bulk storage*, zoals object (of blob) storage zijn de traditionele harddisks nog goedkoper op dit moment. Maar reken er maar op dat de *cloud providers* hard aan het rekenen zijn aan hun *Total Cost of Ownership* (TCO) en dat ze op termijn ook dit soort massa-opslag over gaan zetten op SSD.

Als één schaap over de dam is, volgen er natuurlijk meer. En als dat gebeurt, zal de markt voor SSD's nog meer groeien en gaan de prijzen nog verder dalen. Dus kom maar op Google, Amazon en Microsoft!

1 Zie http://www-03.ibm.com/ibm/history/exhibits/storage/storage_350.html
2 Zie http://www.ewh.ieee.org/r6/scv/mag/MtgSum/Meeting2010_10_Presentation.pdf
3 Zie http://spectrum.ieee.org/computing/hardware/laser-heated-hard-drives-could-break-data-density-barrier
4 Uit onderzoek door Seagate in 2007 kwam naar voren dat er maximaal 37,5 TB op een disk past
5 De Wet van Moore stelt dat het aantal transistors in een geïntegreerde schakeling door de technologische vooruitgang elke 2 jaar verdubbelt
6 Zie https://en.wikipedia.org/wiki/Flash_memory
7 Zie https://en.wikipedia.org/wiki/Trim_(computing)
8 Zie http://techreport.com/review/27909/the-ssd-endurance-experiment-theyre-all-dead