

## Lagring: IP og iSCSI overtar

Denne artikkelen er den siste i en serie på 3 om moderne lagringsteknologi og trender i lagringssegmentet. Den første artikkelen sto i Mellvik-Rapporten nr. 115.

*Vår avslutningsvise introduksjon til IP-basert lagring i forrige artikkel, la ikke skjul på at kostnader er den viktigste drivkraften i overgangen til IP-basert lagring. Prosessen er bare såvidt begynt, men det skal store skylapper til for ikke å se hvilken vei det bærer.*

Teknologiske diskusjoner om FibreChannel, Infiniband og deres fortreffeligheter kan føres forlengs og baklengs, men overser det faktum at teknologisk egnethet kun er en liten del av den totale ligningen som avgjør suksess eller fiasko. IP og Ethernet er begge glitrende eksempler i så henseende: De kan kritiseres og slaktes for sin manglende egnethet i utallige sammenhenger, men fremstår likevel som enestående suksesser – ikke på grunn av, men på tross av sine svakheter.

De samme egenskapene vil sørge for at IP og Ethernet i løpet av de neste to årene seiler opp som den foretrukne infrastruktur for lagringssystemer. Deres sameksistens med FibreChannel er allerede relativt problemfri i teknisk forstand, og bidrar til å akselerere markedsakseptansen. Et nytt prisleie for nettverkslagring vil imidlertid føre til at interessen for FC (og Infiniband)<sup>4</sup> raskt avtar. Samtidig vil de lave prisene føre til at nettverksbasert lagring sprer seg til områder som hittil har vært langt utenfor rekkevidde. Eksperten Jacob Farmer hos velrenommete Cambridge Computer i Boston uttrykte seg slik ved siste årsskifte:

*“iSCSI bringer SAN-kostnadene ned på et nivå som gjør det attraktivt med helt nye anvendelser – å ta i bruk SAN over alt hvor det er [teknisk] mulig i stedet for der kostnadene gjør det overkommelig.”*

### iSCSI

**IETF** – Internet Engineering Task Force, som tar seg av Internett- og IP-relaterte standarder.

På mange måter gjentar iSCSI historien om Ethernet: En velkjent, utbredt, godt standardisert og relativt rimelig teknologi som finner veien inn i nye sammenhenger – fordi summen av fordelene er vesentlig større enn ulempene. iSCSI har vaket på horisonten lenge nok til å ha blitt en standard – hvis endelige ratifisering fra IETF riktignok lar vente på seg, men som er tilstrekkelig stabil til at et stort antall leverandører har valgt å slippe produkter ut på markedet.

### SCSI – gammel travet i ny ham

SCSI-standarden har eksistert siden 80-tallet, og består på den ene siden av et sett kommandoer – en høynivå kommunikasjonsprotokoll, og på den andre siden tekniske spesifikasjoner av fysiske grensesnitt, kabler og så videre. Begge deler har vært gjennom en håndfull utviklingstrinn i perioden. Protokollen er relativt sofistikert, godt oppgavetilpasset og – som vi var inne på ovenfor, velkjent og godt

<sup>4</sup> Infiniband har ytelsesmessige fordeler som sannsynligvis vil sørge for at teknologien får fotfeste og overlever i overskuelig fremtid. Det finnes alltid anvendelser og kunder som har spesielle ytelsesmessige behov der kostnadene er underordnet. I dette segmentet har Infiniband-teknologien og produktene en fremtid.

standardisert. Den forutsetter blant annet at det finnes forholdsvis høy grad av intelligens i selve utstyret, slik at enheten (typisk en disk eller en *tapedrive*, eventuelt en scanner, en fotosetter eller annet spesi- alutstyr) selv kan optimalisere datastrømmene. Mens eksistensen av slik intelligens var en bold antagelse midt på 80-tallet, er det en selv- følge i dag, og har gitt SCSI et permanent om ikke alltid like stort ytel- sesfortrinn i forhold til konkurrerende teknologier.<sup>5</sup>

### **Spentig ytelsespotensiale**

iSCSI flytter kort og godt SCSI-protokollen fra de tradisjonelle flatka- blene, til TCP/IP som infrastruktur. Hva som ligger under IP-nivå, er i prinsippet uinteressant, men spiller naturligvis stor rolle for ytelsen. Derfor er Gigabit Ethernet den foretrukne infrastruktur – over kobber- kabel eller fiberkabel. En slik forbindelse kan gi en overføringshastig- het på ca. 125 MBps, hvilket er konkurransedyktig med de beste av dagens SAN-løsninger. Hvorvidt – og hvordan – en slik hastighet kan utnyttes, er et forhold vi skal komme tilbake til nedenfor.

Denne hastigheten representerer ingen maksimal grense. Det er ingen ting i veien for å parallellisere datastrømmer over flere fysiske forbin- delser, og hastigheten kan dermed flerdobles. Bruken av IP som trans- portprotokoll gjør dette teknisk sett trivielt, og utstyr som støtter slik parallellisering har vært demonstrert. Ethernet-porter – selv de mest sofistikerte variantene – er rimelige, lett tilgjengelige, og teknisk sett enkle. Likeledes vaker 10 Gigabit Ethernet i horisonten, og er i disse dager klar til å ta steget over fra *bleeding edge* til *leading edge* i marke- det. FC-tilhengerne legger ned stor innsats i å holde følge i denne ytel- sesmessige kappestriden, og kan muligens klare det, men kommer likevel til kort på kostnadssiden.

I den andre enden av ytelsessegmentet ser vi at også standard Fast Ethernet (100 Mbps) kan spille en rolle i iSCSI- og lagringssammen- heng. Med effektive datahastigheter på ca. 8 MBps fra Windows-systemer og 10-11 Mbps fra Linux-systemer, gir standard Ethernet hastighet nok for mange applikasjoner – og mer enn nok for moderne *backup*-enheter som DLT 8000 og AIT-2. Denne spennvidden i hastig- het er én av mange faktorer som gjør iSCSI både unik og attraktiv i markedet.

### **Et nytt fleksibilitetsnivå**

En annen faktor er fleksibilitet. Siden en iSCSI-pakke ikke er noe annet enn en IP-pakke som kan rutes og omrutes med enkle håndgrep og høy hastighet, blir håndtering av feilsituasjonen en størrelsesorden enklere enn med for eksempel FC. Og siden utstyret som sørger for håndtering av slik omruting allerede finnes, blir kostnadene knyttet til høy tilgjengelighet og pålitelighet dramatisk lavere enn før.

<sup>5</sup> Såkalt MULTITHREADING på utstyrsnivå, det vil si i selve disken eller RAID-boksen, er den viktigste faktoren i dette bildet. Utstyret kan ta imot data fra flere ulike kilder samtidig, sortere datablokkene internt og sørge for at de skrives ut i optimal rekkefølge. Tilsvarende kan gjøres for lesing. Når enheter aksesserer fra flere kilder, kan slik optimalisering kun gjøres i selve lagringsutstyret, ikke i vertsmaskinen, som har vært vanlig tidligere – under enklere forhold.

Slik kan vi fortsette – det ene argumentet etter det andre ramler ned til IP og iSCSIs fordel. Vi har de siste 10 årene sett en kannibalisering av nettverksmarkedet, der tradisjonelle leverandører av proprietære, kostbare løsninger har måttet se nykommere med IP og Ethernet i blodet rase forbi. Tilsvarende er i ferd med å skje i lagringsmarkedet. Nettverksleverandører som Cisco sikrer seg stadig nye segmenter ved å være tidlig ute med løsninger og produkter som ikke bare støtter IP-lagring og iSCSI, men samtidig sørger for kompatibilitet med 'den gamle verden', med kombinerte iSCSI/FC-produkter.

### Transparent for klientsiden

Denne fleksibiliteten kommer også til syne når vi ser hvilke alternativer som finnes med hensyn til bruk av iSCSI. På system- (*host*-) siden har vi tre relativt innlysende veier å gå, som illustrert i figur 1 nedenfor. Det enkleste er naturligvis å plugge inn en ekstra programvaremodul, som finnes tilgjengelig fra flere hold for de fleste populære plattformer, for eksempel Windows og Linux. Dette er samtidig det mest ressurskrevende alternativet. Tjenermaskinen må selv gjøre all prosesseringen, hvilket er akseptabelt kun for lavt belastede tjenere. På en Windows-basert maskin med 2 Xeon-prosessorer i 1,x GHz-klas- sen blir ytelsesdegraderingen merkbar ved overføringshastigheter på over 20 MBps.

ASPI		Programvare i vertsmaskin		Programvare i vertsmaskin		Programvare i vertsmaskin
iSCSI						
TCP - IP				Ethernet NIC med TOE [TCP Offload Engine]		Ethernet NIC med iSCSI Engine og TOE
Ethernet		Trad. Ethernet NIC				

ASPI – Advanced SCSI Programming Interface

A

B

C

**Figur 1** Vertsmaskin-siden av en iSCSI-forbindelse kan håndteres på 3 ulike måter, med svært forskjellige ytelsesmessige karakteristika. Mens alternativ A er et tradisjonelt, enkelt Ethernet-grensesnitt (NIC), som overlater all nettverks- og iSCSI-prosessering til vertsmaskinen, introduserer B en såkalt TCP OFFLOAD ENGINE. Alternativ C går ett trinn videre og representerer nivået som vil være alminnelig om et par års tid. Her er hele vertssiden av SCSI-prosesseringen flyttet ned i grensesnittet, som dermed ser ut som et regulært SCSI-grensesnitt sett fra vertsmaskinens og dens operativsystems side.

Alternativ B flytter prosesseringen av transport-protokollene (TCP og IP) over til nettverkskortet (et såkalt TOE-kort eller TNIC, *TCP Offload Engine*), hvilket bedrer ytelsesforholdet vesentlig. Fortsatt er imidlertid belastningen på vertssystemet større enn om det SCSI-baserte lagringssystemet hadde vært lokalt. For å komme opp på et nivå som tilsvarer lokalt tilkoblet masselager, må vi til alternativ C (figur 1), der også SCSI-håndteringen er flyttet til det intelligente nettverkskortet. Her ser vertssystemet kun et SCSI-grensesnitt, akkurat som om disken var direktekoblet. Mens det ikke hersker tvil om at dette er hva vi kommer til å se i fremtiden,<sup>6</sup> vil alternativ A og B være dominante i de neste 12-18 månedene.

**PXE – Preboot eXecution**

*Environment* er en standard utviklet av Intel for oppstart av maskiner via nettverket, og brukes typisk i forbindelse med oppstart av diskløse klienter. Fremveksten av SAN-løsninger og tilhørende reduksjon i bruk av lokalt masselager, har imidlertid gjort mekanismen tilsvarende attraktiv for tjener-maskiner.

En forutsetning for å gjøre lagringsløsningen fullstendig transparent er at vertssystemene også kan starte (*boote*) de nye diskene – som om diskene var lokale. Dette er allerede tatt vare på, og er først og fremst et spørsmål om resident programvare på de intelligente nettverkskortene. Cisco var først ute med å levere en slik løsning (2003), som kan starte Windows (og andre operativsystemer) via en PXE-implementasjon. Bruken av PXE er teknisk sett unødvendig og representerer en midlertidig snarvei til målet. I løpet av inneværende år blir den erstattet av 'ekte' iSCSI-disk oppstart via programvare i de intelligente grensesnittene.

De intelligente grensesnittene med SCSI-prosesseringen innebygget, realiserer målsettingen som lå til grunn da utviklingen av iSCSI først ble lansert: At det hele skulle være totalt transparent for vertssystemene. Det ene SCSI-grensesnittet byttes ut med det andre, og vi har byttet fra lokal lagring (DAS) til SAN uten inngrep eller komplikasjoner på tjener-siden. At vi etterhvert kan komme til å introdusere programvare som styrer eller administrerer masselageret på en ny måte, er en annen sak.

**Forvirrende produktvalg**

Mens vertsmaskinsiden av ligningen er relativt oversiktlig, er situasjonen på lagringssiden vesentlig verre, mange vil si kaotisk. Forholdet er at mens samtlige etablerte leverandører i segmentet har erkjent at veien peker mot iSCSI, er de tilbakeholdne med å ta steget fullt ut – av årsaker vi har vært inne på tidligere. Her er det mye som skal beskyttes, ikke minst etablerte markedsandeler og utviklingskostnader på FC og Infiniband. Samtidig banker nykommere på døra med nye løsninger og smarte alternativer som slår bena under hele produktserier fra 'de gamle'.

Utover at iSCSI vil overta markedet på sikt, kan vi med sikkerhet spå at en konsolidering og opprydding er underveis. Dagens produktmarked er preget av garderinger – alle forsøker å tilfredsstille flest mulig kombinasjoner og scenarier. Vi kan koble iSCSI til FC-svitsjer, rute FC via IP til andre FC-systemer, NAS-produkter får iSCSI-grensesnitt og så videre. Lagringsrutere (som også kalles lagrings-broer, hvilket er et mer dekkende navn) styrer lagringstrafikk mellom klienter (vertsmaskiner) og SAN-systemer, NAS-systemer og rene iSCSI-produkter.

Mens all denne fleksibiliteten bidrar til å sørge for at investeringene vi allerede har gjort på SAN og NAS-siden kan sameksistere med nyere og mer fremtidsrettet iSCSI-utstyr, har løsningene også en tendens til å bli kompliserte og uoversiktlige. Denne kompleksiteten kan i sin tur føre til redusert stabilitet, for eksempel fordi verktøyene vi har til disposisjon ikke dekker de kombinasjonene vi faktisk bruker. Dessuten kan vi stå overfor driftsmessige mareritt, for eksempel at lagring må allokeres manuelt på to nivåer.

<sup>6</sup> Hovedkort med slik funksjonalitet er allerede kommet på markedet, og bidrar til både å senke (produksjons-)kostnader og presse priser. Derfor er det temmelig risikofritt å spå at dette blir regelen snarere enn unntaket på tjenersiden i overskuelig fremtid.

## Neste fase

Løsningen på utfordringene ligger i overgangen til rent iSCSI-utstyr, med tilhørende administrative verktøy, som ikke bare skal ta hånd i hanke med selve lageret – allokering og bruk av ressurser, men også problemstillinger knyttet til sikkerhet, som vi ikke skal komme inn på i denne sammenhengen.

2004 blir et gjennombruddsår i så henseende – både med hensyn til produkter og verktøy, og derfor et tema vi skal re vitisere her i Mellvik-Rapporten – i 2. halvår. I mellomtiden er det viktigste praktiske poenget å ikke anskaffe SAN eller SAN-produkter uten at de er forberedt for å håndtere IP, Ethernet og iSCSI. FC forblir dominant i markedet i inneværende år på grunn av opparbeidet massefart og den trygghet dette skaper. iSCSI står imidlertid klar til å ta over, og får gode skussmål fra miljøer som allerede har tatt steget. I kjølvannet vil prisene i løpet av kort tid halveres. På sikt fortsetter fallet. **Det** er gode nyheter. ■