

Lagring: Fra kostbar ressurs til rimelig tjeneste

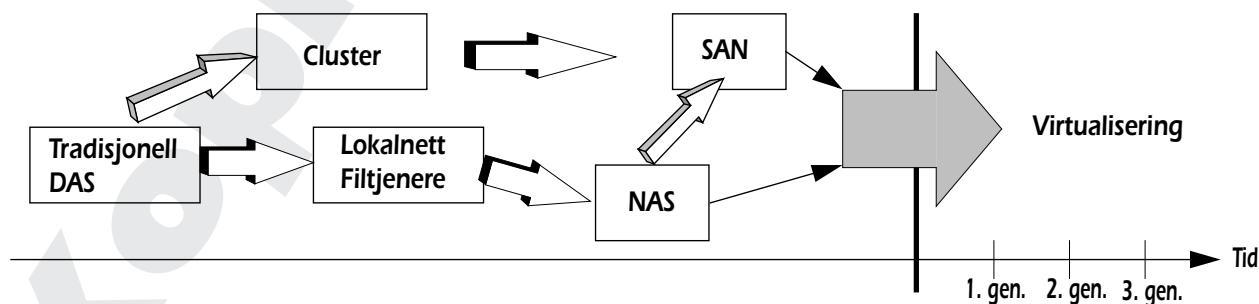
Det er ikke til å unngå at vi fra tid til annen stopper opp et øyeblikk eller to, og sender en tanke tilbake til vår første datamaskin, med 56 kB hukommelse og 300 kB per diskett. Ikke minst når vi med den største selvfølge plugger inn en mikroskopisk 2GB harddisk i digital-kameraet og fyrer løs, dukker slike assosiasjoner opp. Samtidig observerer vi en oppvoksende generasjon for hvem disse ressursene er like selvfølgelige som kalkulator og TV: Husets 18-åring har lite annet enn et hevet øyenbryn å komme med dersom vi foreslår en opprydding for å avhjelpe plassproblemene blant hans gigantiske 280GB masselager.

Vi har lært å tenke annerledes – gang på gang, og tiden er moden for å glemme det meste av hva vi har lært og erfart med hensyn til nettopp masselager. Både båndbredde og lagringskapasitet er blitt så billige at det er lønnsomt å sløse – dersom sløsing fører til at vi sparer tid. Mens ordet 'sløsing' i utgangspunktet smaker vondt, er det verdt å minne seg om at også sløsing er relativt: Dersom sløsing på én front fører til innsparinger på en annen, er det til slutt ressursenes relative kostnader som avgjør hvilke valg som er optimale.

Lagringsnett: En mellomstasjon

Vi har vært igjennom en fase der ulike former for lagringsnettverk har vært den riktige løsningen for både store og små organisasjoner med betydelige lagringsbehov. Over en 10-års periode har teknologi og produkter utviklet og stabilisert seg på et nivå hvor vår avhengighet av selvstendige lagringssystemer er total – økonomisk, driftsmessig og praktisk. Lagringen hører hjemme på nettverket, frikoblet fra applikasjons- og nettverks-tjenere.

Frikoblingen har besørget en ny grad av fleksibilitet med hensyn til hvordan masselageret brukes, aksesseres og hvordan tilhørende tjenester – for eksempel sikkerhetskopiering og replisering – kan organiseres. Dessuten er kapasitet og pris/ytelse hinsides enhver fantasi fra



Figur 2 Utviklingen fra tradisjonell DAS (DIRECT ATTACHED STORAGE) til virtuelle lagringssystemer har tatt ca. 15 år så langt, og vært igjennom en rekke interessante faser underveis.

midten av 90-tallet, hvilket har hatt store konsekvenser for hvordan vi bruker ressursene. Det er ikke nødvendigvis optimalt å rulle sjelden brukte data ut på sekundære og vanskelig tilgjengelige arkivsystemer. Likeledes er replisering i stor skala blitt kosteffektivt i nye sammenhenger.

Datalagring har forlatt rollen som vedheng til tjenere, og etablert seg som en selvstendig tjeneste – en ressurs som på linje med en rekke andre tilbys og leveres via nettverket. Om nettverket er lokalt eller fjernt spiller ingen praktisk rolle. Det er tjenestens effektivitet som avgjør dens kvalitet overfor klienter og brukere. Fysisk avstand er ingen relevant faktor så lenge den ikke fører til forsinkelser. Tilgjengelig båndbredde, dens stabilitet og ikke minst pris er parametrene som avgjør ressursenes optimale plassering.

Likeledes har utvikling og bruk av NAS og SAN lagt grunnlaget for hardware-løsninger og verktøy som på den ene siden sørger for å virtualisere selve masselageret, og på den andre siden gir pålitelighet, robusthet og fleksibilitet på et tidligere ukjent nivå.

Med Gigabit Ethernet (GE) tilgjengelig over lange distanser, og med raskt dalende priser på fiberkapasitet og GE-utstyr, er veien kort til å sette bort – OUTSOURCE – hele lagringsproblematikken, la spesialister overta og kjøpe tilbake en tjeneste med spesifisert mengde og kvalitet – og forutsigbare kostnader.

Er vi klare for en slik forandring? Er teknologien moden? Finnes leverandørene? Disse og flere spørsmål skal vi komme inn på i diskusjonen nedenfor.

Et umettelig behov

La oss imidlertid begynne med det mest grunnleggende forholdet: Behovet for nettopp lagringskapasitet. Den årlige tilveksten av kapasitet overstiger langt det vi for 10-15 år siden anså for å være markedets totalbehov. Videre viser veksten ingen tegn til å avta, tvert imot. Faktum er at behovet forventes å akselerere i årene fremover – og veksten i det profesjonelle markedet er interessant nok temmelig konjunkturavhengig. Våre digitale arkiver ser ingen forskjell på gode og dårlige, nyttige og unyttige, attraktive eller uønskede data. Dårlige nyheter tar like stor plass som gode nyheter, mens regnskaper og årsresultater skal lagres om underskuddene er aldri så store. Filmer, fjernsynsprogrammer, radioprogrammer og CDer – uavhengig av kvalitet – utgjør enorme datamengder som tidligere er arkivert i analogt format på andre media. En kontinuerlig strøm av 'gamle' data – bilder, bøker, filmer, opplæringsprogrammer og lyd – blir måned for måned konvertert til digital form og skal lagres.

Som om ikke dette var nok, fikk bevisstheten rundt katastrofeplanlegging² et skikkelig spark bak etter angrepene mot New York og Washington i fjor høst. Data skal ikke bare sikres, de skal også kunne gjøres

² Se miniserien "Katastrofeberedskap: Hvorfor og hvordan" i Mellvik-Rapporten nr. 92 og 93.

tilgjengelige på rimelig kort varsel selv om det inntreffer en katastrofe. Utilgjengelighet er blitt for kostbart i forhold til hva det koster å sørge for tilgjengeligheten.

Også i privatmarkedet – som naturlig nok er mer konjunkturfølsomt – er behovsutviklingen eksplosiv, i kjølvannet av digitale kameraer, Web-kameraer, DVD-filmer, digitale videokameraer, MP3-musikk og mulighetene for bredbåndsoverføring av alle disse datatypene. At en middels erfaren tenåring selv kan oppgradere sin PC med 120 GB masselager for under NOK 1.500, er kun et første trinn på denne utviklingen. Før eller siden vil også dette markedet oppdage at selv gjort langt fra alltid er velgjort, ei heller på denne fronten. Det er langt mellom private PC-brukere som har noe forhold til sikkerhetskopiering. I disse virus-tider, er antall brukere som har erfart å miste alt de har samlet, raskt voksende.

Konsolidering

I profesjonelle miljøer har konsolidering stått på agendaen siden midt på 90-tallet. 80-tallets PC-eksplosjon førte til en oppløsning av etablerte strukturer som holdt orden på data, datakvalitet, programvare, lisenser og aksesskontroll. Resultatet kunne ingen leve med i lengden, og en massiv tjener-konsolidering parallelt med utbygging av lokalnett bidro til å få situasjonen under kontroll.

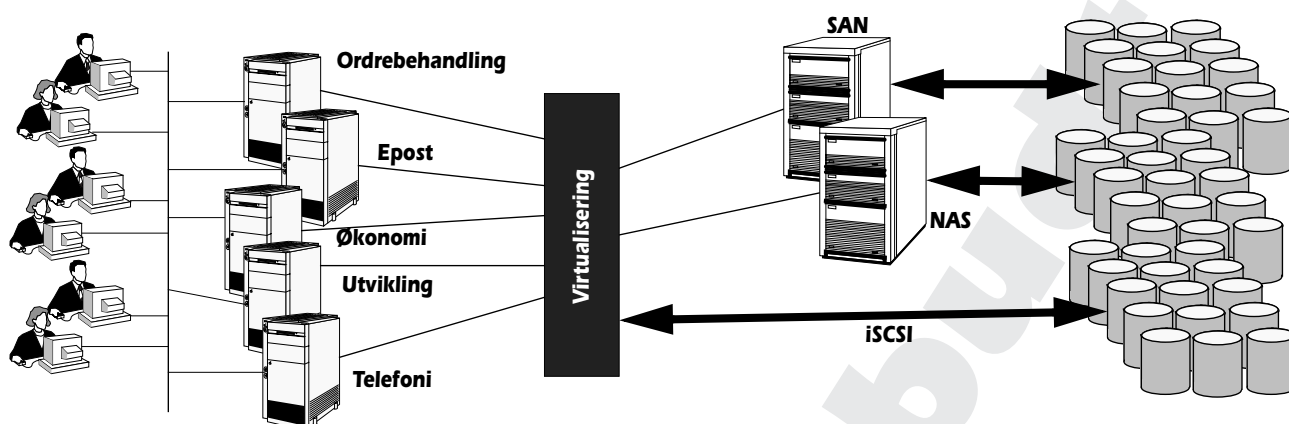
I kjølvannet av denne konsolideringen dukket velkjente problemstillinger fra 70- og 80-årene opp: Pålitelighet og tilgjengelighet ble igjen et sentralt ansvar. Muligheten til å diktere perioder for driftsstans utenfor 'normal arbeidstid' var imidlertid borte. Sammen med lokalnett, Internett og et voksende antall Web-tjenester forsvant distingsjonen mellom dag og natt, kjernetid og kveldstid. Å stenge av systemene over natten, eller kanskje sågar en hel weekend for vedlikehold, sikkerhetskopiering eller oppgraderinger, var ikke lenger mulig. Resultatet ble utilfredstillende vedlikehold, sikkerhetskopiering og beredskap.

Tiden var moden for ytterligere konsolidering – denne gangen på lagringssiden. Det som begynte med *cluster*- (klynge-) teknologi sent på 80-tallet ble til lagringsnettverk (SAN) og nettverkslagring (NAS) på 90-tallet, og la grunnlaget for dagens virtualiseringstrend på lagringssiden (figur 2).

SAN – Storage Area Network
NAS – Network Attached Storage

Virtualisering

Vi ser at nøkkelen til konsolidering av masselager er frikoblingen mellom lagring og systemer. Arkitektorene vi ser konturene av i dag har mye til felles med avanserte lagringsnettverk, men er for det første enklere – fordi nettverksteknologien som benyttes ikke lenger er spesiell (typisk FibreChannel), men basert på standarder TCP/IP, iSCSI og Gigabit Ethernet (eller 10 Gigabit Ethernet). For det andre blir systemene, som vi også var inne på ovenfor, tilført et ekstra nivå av programvare som fullstendig gjemmer den underliggende hardware fra nettverket og dermed klient-systemene.



Figur 3 Mens dagens lagringsnettverk representerer et viktig skritt mot å gjøre lagring til en selvstendig tjeneste, er kontakten med og avhengigheten av den underliggende teknologien fortsatt stor. Ved å innføre et virtualiseringsnivå mellom klientene (applikasjonstjenester) og de fysiske lagringssystemene, tar vi et langt skritt i retning av å fjerne den siste teknologi-skansen og gjøre lagring til en regulær tjeneste.

Fra drøm ...

Dette programvarenivået åpner for en grad av fleksibilitet og dynamikk som tidligere ikke har vært mulig. Med referanse til figur 3 ovenfor, har vi introdusert et nytt nivå mellom klientsystemene og det fysiske maskelageret – som tilbyr generaliserte lagringstjenester til sine klienter. Hva som finnes bakenfor, er både usynlig og irrelevant sett fra klientenes side. I teorien kan det finnes en vilkårlig blanding av teknologier bak virtualiseringsnivået: NAS og SAN kombinert med ferske iSCSI-baserte systemer, og kanskje sågar tradisjonelle systemer (DAS, *Direct Attached Storage*).

Slik virkeligheten ser ut i dag, er fleksibiliteten betydelig mindre. Enkelte leverandører virtualiserer kun sine egne produkter, mens andre leverer det vi kan kalle generisk virtualisering, som er vesentlig mer i harmoni med den idealiserte modellen ovenfor. Ingen har så langt presentert et produkt – enn si et praktisk konsept – som kan gi fullstendig virtualisering, og fungere som en sentral for all lagringstrafikk.

... til virkelighet

Praktisk talt samtlige aktører i markedet for lagringssystemer har lansert eller annonsert virtualiserings-systemer eller -produkter i løpet av det siste året. Dette er åpenbart et marked alle vil ha en bit av, og aktiviteten er deretter. De som ikke har tid, kompetanse eller kapasitet til å utvikle sine egne løsninger, lisensierer eller kjøper andres. Enkelte bygger funksjonaliteten direkte inn i lagringssystemene, mens andre lager selvstendige enheter som typisk kan håndtere flere produktgrupper eller -typer.

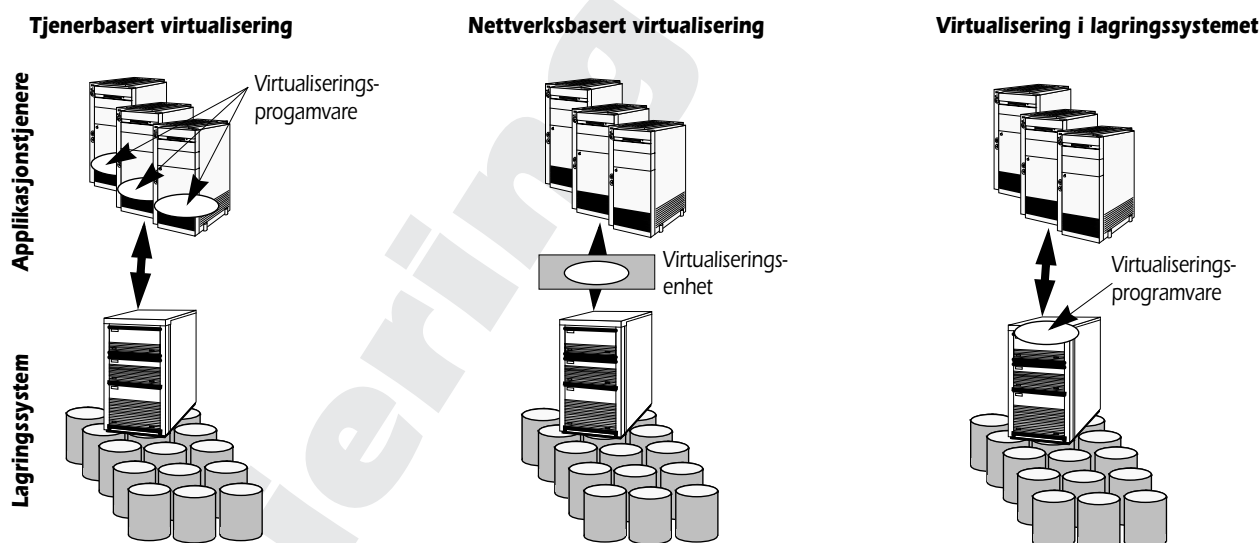
Om produktene som er tilgjengelige i dag er aldri så avanserte, er det et faktum at de representerer første generasjon. Videre viser en gjennomgang av produkter og strategier fra de største aktørene i lagringssegmentet – HP/Compaq, EMC, Sun, Network Appliance, IBM og Hitachi – at de fleste fortsatt har mye å gjøre på bakrommet før pla-

nene om reell virtualisering kan konkretiseres utover 'plansje'-nivå. 2002 har vært et hektisk år for disse 6 og deres kommende konkurrenter, nystartede selskaper med gode ideer og evne til å sette dem ut i livet.

Ikke overraskende er det blant nykommerne vi finner de mest interessante ideene og produktene. DataCore [www.datacore.com], FalconStor [www.falconstor.com] og Vicom [www.vicom.com] er gode eksempler og plasserer seg tett opp mot den idealiserte skissen ovenfor. Programvare som kjøres på standard hardware, og kan virtualisere et bredt spektrum av ulike lagringsløsninger. Sammen med et halvt dusin andre nykommere representerer disse tre en spennende samling oppkjøpskandidater: Sannsynligheten for at minst halvparten av dem blir overtatt i løpet av de neste 12 månedene – av større aktører som mangler teknologi, men har penger.

Hvilken arkitektur?

Diskusjonen ovenfor avslører blant annet at det er stort sprik leverandørene imellom med hensyn til valg av arkitektur når virtualisering skal introduseres. Enkelte velger å plassere programvaren på tjenersiden, andre i nettverket og atter andre i selve lagringsystemene. Det kan argumenteres godt for alle variantene, og på kort sikt vil de leve side om side (figur 4).



Figur 4 Virtualisering kan gjøres på tre grunnleggende ulike måter, alle med sine egne karakteristika, fordeler og ulemper. Enkelte leverandører (for eksempel HP) har eller har hatt produkter i alle tre kategorier, mens mye taler for at den nettverksbaserte varianten på sikt kommer best ut.

Utviklingen begynte med tjener-basert virtualisering, i produkter som HPs OpenView Storage Allocator og Veritas Volume Manager, som fortsatt brukes og representerer et nærliggende progresjonstrinn for miljøer som allerede benytter disse produktene. Den viktigste ulempen er at virtualiseringen ikke er transparent på tjenersiden, men krever programvare på hver tjener, som i sin tur skal styres og vedlikeholdes. I små og oversiktlige miljøer er dette kurant nok, mens det raskt blir komplisert i store og dynamiske omgivelser.

På lenger sikt er det sannsynlig at faktorer som enkelhet og fleksibilitet vil avgjøre hvilken variant som blir dominant, eller om en helt ny angrepsvinkel og dertil hørende arkitektur er svaret på utfordringene. For eksempel er det åpenbart en stor fordel sett fra klient-systemene (typisk applikasjons-tjenere) at de kan forholde seg til lagringssystemene uten introduksjon av ekstra programvare lokalt. Videre er det investert store summer de siste 5 årene i avanserte lagringssystemer som ikke går ut på dato de første årene, verken funksjonelt eller kapasitetsmessig. Dermed blir det en åpenbar fordel for nye produkter å kunne innlemme eksisterende løsninger i systemet uten store komplikasjoner.

Virtualisering i lagringssystemet

Virtualisering i lagringssystemet og på tjenersiden har tallrike felles trekk, blant annet fordi det er snakk om å introdusere ekstra programvare på systemer som allerede finnes. En åpenbar fordel med å gjøre forandringen på lagringssystemet i stedet for på hver enkelt tjener, har med administrasjon å gjøre: I de fleste sammenhenger er færre bedre. På den andre siden begrenser vi fleksibiliteten når selve lagringssystemet også skal gjøre virtualiseringen: Vi får tilleggskompleksitet innenfor de rammene den enkelte leverandør setter.

De tjenerbaserte løsningene vi nevnte ovenfor representerte en åpenbar trussel for leverandørene av lagringssystemer gjennom å dekke over produktenes egenskaper – usynliggjøre dem. Derfor tok det ikke lang tid før også denne gruppen hadde 'virtualisert' sine produkter. I tillegg til å beskytte seg mot trusselen fra generelle programvare-løsninger, var interessen stor for å begrense fleksibiliteten og binde de nye tjenestene så tett som mulig til fysiske produkter. Deres viktigste argument var – og er fortsatt – muligheten til å 'trimme' ytelse og besørge høy pålitelighet gjennom tett kobling mellom virtualisering og underliggende utstyr.

At det var her utviklingen startet, betyr imidlertid ikke at noen av disse kombinasjonene er en optimal kombinasjon. Det er vel og bra at lagringssystemene får sofistikerte egenskaper som optimaliserer pålitelighet, sikkerhet og effektivitet, men det gir – som vi allerede har vært inne på – beskjeden fleksibilitet å plassere virtualiseringen så nær det fysiske lageret. Videre er det naturlig å spørre om intelligensen i lagringssystemene har nok kapasitet til å håndtere kompleks virtualisering uten å introdusere forsinkelser. Erfaringer med dagens produkter har gjort det naturlig å stille et slikt spørsmål.

Nettverksbasert virtualisering

Etter den første bølgen med tjener-basert virtualisering, der spesialiserte aktører som HP, EMC og Network Appliance introduserte grader av virtualisering i sine lagringsprodukter, har pendelen svingt over mot nettverks-basert virtualisering.

Plattformen er gjerne standard Intel- eller SPARC-basert hardware, med NT, W2k, Linux eller Solaris som operativsystem, og spesialisert

programvare som gjemmer lagringssystemene bakenfor. Arkitekturen gir den fleksibilitet vi etterlyste ovenfor, men introduserer samtidig en potensiell flaskehals: Alle masselager-transaksjoner må gjennom virtualiserings-enheten, som med standard hardware og ditto operativsystem knapt kan sies å være optimalisert for oppgaven. Spesielt Intel-baserte produkter har vært utsatt for kritikk i så henseende. Dagens PC-plattform er vel og bra i sin allsidighet, men til slike oppgaver er den samme allsidigheten en ulempe, ikke en fordel. Det samme kan anføres for operativsystemenes del: De er generelle, allsidige, tunge og ressurskrevende – og slett ikke laget for spesialiserte oppgaver som dette.

Leverandørene svarer med å introdusere sofistikerte caching-mekanismer som skal redusere svartidene. Selv om slike tiltak påviselig hjelper, er det fortsatt et faktum at virtualiserings-enhetene i mange tilfeller har mindre ressurser å rutte med enn selve lagringsenhetene. Mens leverandørene lover at flaskehalsene skal bort og at ytelse ikke skal være noe problem, er det vanskelig å se for seg at dagens nettverksbaserte løsninger kan skalere tilstrekkelig.

Neste trinn

Gjennomgangen leder oss naturlig til neste trinn i utviklingen. På samme måten som nettverksrutere, NAS-produkter og en rekke andre produktkategorier har gått fra spesial-løsninger på generelle plattformer til skreddersydd hardware, vil den nære fremtid bringe nye alternativer til lagringssegmentet. Vi ser konturene av en serie svitsje-produkter, i fagkretser kalt 'virtualiserings-svitsjer', med de egenskapene vi har etterlyst ovenfor, både med hensyn til ytelse og ikke minst fleksibilitet overfor de bakenforliggende lagringssystemene.

Spørsmålet er ikke om disse produktene kommer, men når – og ikke minst hvem som inntar ledertrøya i feltet. De etablerte leverandørene av lagringssystemer har mye å beskytte, og dermed en egen interesse av å trenere utviklingen. Samtidig ser de uten tvil hvilken vei det bærer, og deres valg av tidspunkt for å hoppe på toget mot fremtiden blir kritisk.

Oppsummering

De faktiske forhold gir oss grunnlag for to relativt opplagte konklusjoner. For det første befinner vi oss i en utviklingsmessig brytningsfase. Lagrings-virtualisering har passert '*point of no return*' i den forstand at det ikke hersker noen tvil om konseptets fremtid. De praktiske øvelsene har av naturlige årsaker vært gjennomført på standard-plattformer, hvilket karakteriserer dagens produkter. Neste generasjon, som er på beddingen i disse dager, vil gi produkter som er vesentlig mer effektive, sofistikerte og fleksible. Det betyr også at fremtiden ikke kan kjøpes i dag, og at de løsningene vi anskaffer på kort sikt, neppe vil ha lang levetid, i alle fall ikke på egen hånd. Mens dette er en uønsket situasjon, er den knapt uvanlig, og behovsutviklingen stopper ikke opp selv om vi gjerne skulle ha ventet 12 eller 18 måneder med nye investeringer.

Det er viktig å være klar over at dagens virtualiserings-produkter allerede har vesentlige fordeler å by på fremfor tradisjonelle lagringsløsninger. De representerer et trinn i en spennende og positiv retning ikke bare for leverandørene, men i like stor grad for markedet. At utsiktene til et arkitekturskifte i overskuelig fremtid ligger der, er like naturlig som at vi for to år siden kjøpte 1 GHz PCer, vel vitende om at 2 GHz ville bli vanlig om et år eller to.

Ingen tjenesteleverandører i sikte

Den andre konklusjonen er at veien frem til spesialiserte, kommersielle lagringstjenester blir et reelt alternativ, fortsatt er lang og kan måles i år – spesielt på våre kanter. I USA og enkelte andre land begynte disse tjenestene å røre på seg allerede for to-tre år siden, og har bidratt til verdifull erfaring både markedsmessig og produktmessig. En rekke aktører har satset stort og gått over ende, mens andre har funnet riktige nisjer og forøvrig vært dyktige nok til å få hjulene til å gå.

Avhandlinger kan skrives om erfaringer, feil og suksesser i den forbindelse, mens vi i denne omgang skal nøye oss med konklusjonen og dens følger på kort sikt: At eksterne tjenester ikke vil bli tilgjengelige i nær fremtid betyr ikke at vi skal la problemstillingen ligge. Det trengs ikke tredjeparts-leverandører for å frikoble og virtualisere lagringssystemene. Faktum er at de fleste IT-miljøer er tjent med å starte prosessen på egen hånd, gjøre lagring til en tjeneste og ta med de fordelene dette medfører – gjerne som et trinn på veien mot en fremtidig *outsourcing*. Dermed blir vi kjent med problemstillingene, arkiturene, tjenestene og de reelle kostnadene – erfaringer som er gull verdt på egen hånd, og minst like mye verdt i en forhandlings situasjon med potensielle tjenesteleverandører.

Med hensyn til spørsmålene vi stilte innledningsvis, er svaret 'nei': Vi er ikke klare, leverandørene er ikke klare og produktene er ikke modne. Det er for tidlig å hoppe i land, men vi har land i sikte, og tiden er inne til å forberede seg på den fremtiden som kommer mot oss. At vi må legge til side mange dyrt ervervede erfaringer og kunnskaper, kan være smertefullt og ulystbetont, men fremtiden er alltid spennende – for den som ikke bare ønsker å være med, men også å styre den. ■