

Řešení zálohování s deduplikací

Deduplikace

Podle Gartnera je deduplikace nastupující technologie, která má v IT silné opodstatnění a velice rychle se rozšiřuje její využívání.

*Gartner Feb 2009: Technology Recommendation:
“Consider data deduplication for use in backup and recovery scenarios when speed, cost containment and IT modernization are required.”*

Firma DataDomain vznikla s jednoznačným zadáním – vytvořit storage, která bude deduplikovat všechna uložená data, tzv. single-instance storage neboli uložit unikátní data pouze jednou.

Gartner Dec 2007:

“Data Domain is the deduplication market share and mind share leader” “...most customers and most systems shipped”

Gartner Feb 2009:

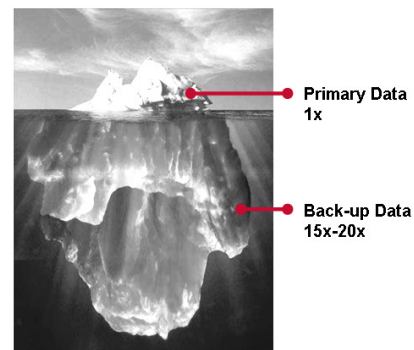
“Data Domain has the largest market share, both in terms of customers and appliances in production. Additionally, Data Domain's in-line deduplication solutions are offered with a variety of price, performance and capacity ranges.”

Zálohování ve firmě

Každá firma má specifická zálohovací schémata, která odpovídají důležitosti jednotlivých aplikací a systémů. Ve finále je nicméně možné obecně popsat problémové oblasti, které jsou podobné pro většinu firem a vyplývají z povahy klasických zálohovacích procesů jako takových.

Typické problémy tradičního zálohování:

- mohutné nárůsty dat
- dodržování SLAs pro zálohu a obnovu
- drahé a nezabezpečené převážení pásek mezi lokalitami
- spotřeba energie, nároky na chlazení, spotřeba prostoru v serverovnách
- problémy se spolehlivostí pásek
- mechanické problémy
- delší doby obnovy
- problémový technologický upgrade



Legitimita páskových řešení

Řešení zálohování pomocí páskových knihoven je stále naprosto legitimním nástrojem, nicméně se stává již jen částí celkového řešení zálohování.

Zejména díky rychlostem obnovy, vstoupil do procesu zálohování další stavební prvek – disková cache. Obvykle kapacita propagovaná na diskovém poli, dedikovaná pro držení nejnovějších záloh, která umožňuje mnohem rychlejší obnovu, ale také zálohu.

Nový stavební prvek – deduplikovaná storage

Dalším stavebním prvkem, který je možno v zálohovacím řešení využít je zálohovací storage s deduplikací a replikací.

Tento prvek je určen k ušetření zálohovaných kapacit, protože ukládá pouze nové, unikátní bloky dat. Tímto způsobem dokáže ušetřit až 90 % stávajících zálohovacích kapacit.

Zároveň jsou tyto zálohovací storage schopny replikovat data, která obsahují na další podobné zařízení pomocí WAN a umožňují tak vytvářet efektivní Disaster Recovery plány.

Moderní celkové řešení zálohování je v současnosti možno vytvořit efektivní kombinací výše popsaných prvků, tedy pásek, diskové cache a deduplikační storage.

Deduplikační storage od DataDomain

DataDomain reálné výsledky

Firma DataDomain byla založena v roce 2001. Od té doby bylo celosvětově prodáno **9500 systémů** celkem **3600 zákazníkům**. Celková kapacita na těchto systémech dosahuje **1650 petabajtů** dat. Napříč všemi systémy je každý den stahován tzv. autosupport report, který kromě informací o kondici boxu, poskytuje i informace o celkovém deduplikačním poměru, který je na boxu dosahován. **Průměrný deduplikační poměr naměřený napříč všemi systémy je: 13,4 : 1, což reprezentuje úsporu zálohovacích kapacit v řádu 92,5 % z objemu zálohovaných dat.**

Architektura DataDomain

Zálohovací storage je možno připojit do současné infrastruktury pomocí 1/10 Gb Ethernet karet nebo pomocí FC HBA 4 Gb, nebo kombinací obou adaptérů.

Přístupovat data je možno pomocí protokolů CIFS, NFS, NDMP, OST a VTL a jejich kombinací.



Deduplikace je prováděna tzv. inline, to znamená rovnou při samotném zálohovacím procesu a všechna data jsou na storage uložena již v deduplikované podobě.

DataDomain je zálohovací storage, proto jako zálohovací kapacitu využívá SATA disky s vysokou densitou zápisu a nízkou cenou za GB.

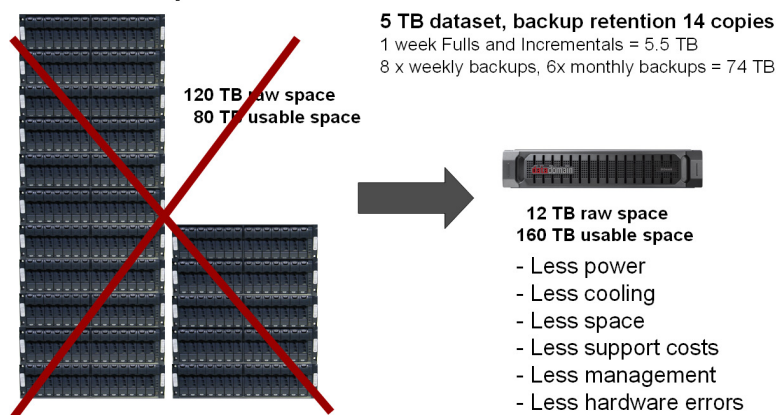
Integrace se zálohovacími softwarý

DataDomain storage je zálohovacím softwarům propagována, buďto jako běžný filesystém nebo jako virtuální pásková knihovna, tzn. že je pro jakýkoliv zálohovací systém naprosto transparentní a není nutné cokoli doinstalovávat.

Jediná podmínka využívání zálohovací storage je odpovídající kapacitní licence na straně zálohovacího software, která bude pokrývat buďto backup to disk nebo VTL licenci dle zvoleného způsobu přístupu.

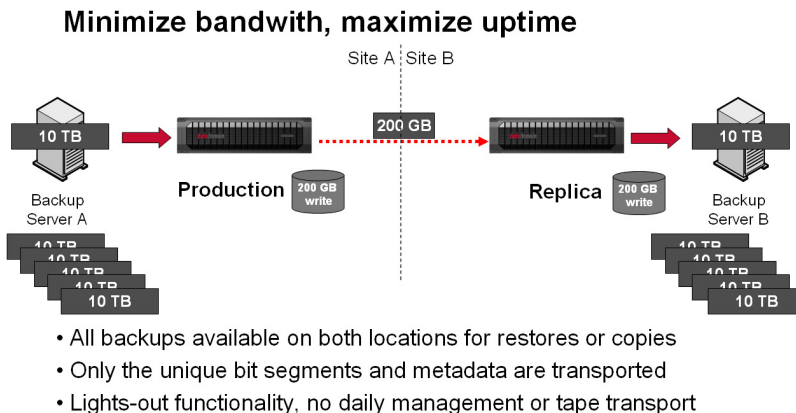
Nastavení zálohovacích jobů určuje, zda bude destinací pro data pásková knihovna nebo deduplikovaná zálohovací storage.

Comparison with S-ATA disks or VTL



Disaster recovery plány a deduplikovaná replikace

Kromě ušetření zálohovacích kapacit pomocí deduplikace, je další výhodou DataDomain zařízení deduplikovaná replikace, tzn. že se přes WAN přenášejí pouze unikátní nové bloky dat.



Hlavní výhody nasazení deduplikované zálohovací storage

- 1) deduplikace dat – ušetření zálohovacích kapacit (redukce počtu pásek a velikosti páskových knihoven)
- 2) spolehlivost obnovy a dodržení SLAs
- 3) replikace přes WAN – deduplikovaná replikace – přenášejí se pouze nové unikátní bloky dat
- 4) replikace přes WAN – vytvoření DR plánu nebo replikace záloh z poboček do centrály
- 5) nezávislost na používaném zálohovacím softwaru
- 6) prověřené řešení

Řešení s DataDomain

Navrhované řešení je kombinací páskových knihoven a deduplikační zálohovací storage. Páskové knihovny mohou sloužit pro data s delší retencí uchování. Deduplikační storage pak slouží, pro masivní úsporu zálohovacích kapacit a pro vytvoření DR konceptu pomocí replikací mezi DD boxy.

