

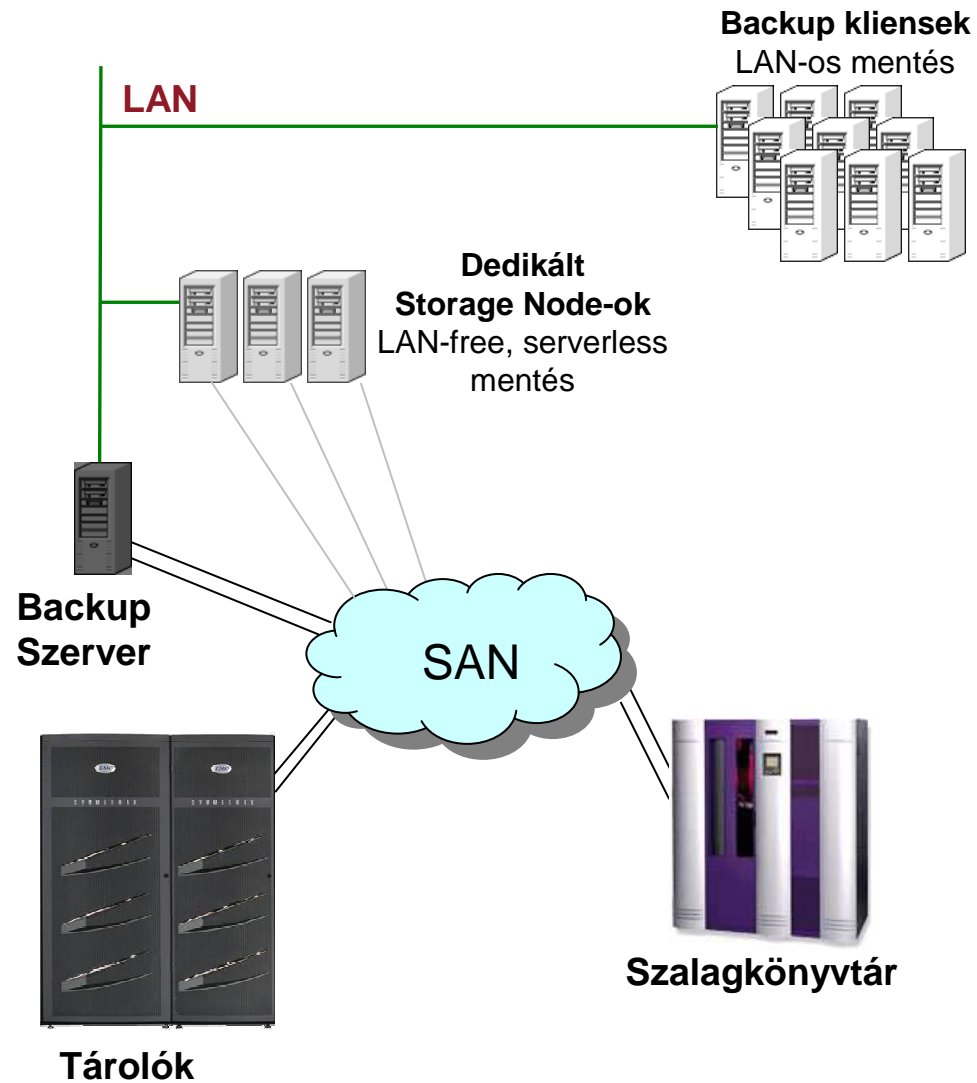
Adatmentés és deduplikáció

Jelinkó Patrik
technológiai tanácsadó
EMC Magyarország

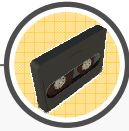
Email: jelinko_patrik@emc.com

- Angolban jobban elkülönülő fogalmak
 - Backup (restore)
 - Data recovery
 - Archiving
 - Active/intelligent archiving

Hagyományos mentési környezet



Nehézségek a szalagos mentési rendszerekben



Teljesítmény

- Mentési ablak túl kicsi
- Visszaállítás hosszadalmas
- Mentési SLA betartása nehézkes
- Szalagos médiák megbízhatósága nem megfelelő
- Távoli telephelyek központi mentése nehézkes



Költségek

- Hatalmas kellékanyag szükséglet
- Nehézkes Off-line szalag kezelési és tárolási teendők
- Folyamatos meghajtó és tárhely bővítés
- Folyamatos migráció
- Adatok hosszú távú tárolása
- Média elhasználódása



Menedzsment és biztonság

- Médiák elvesztése
- Folyamatos problémamenedzsment
- Jogviták
- „Egy szalag nem szalag”

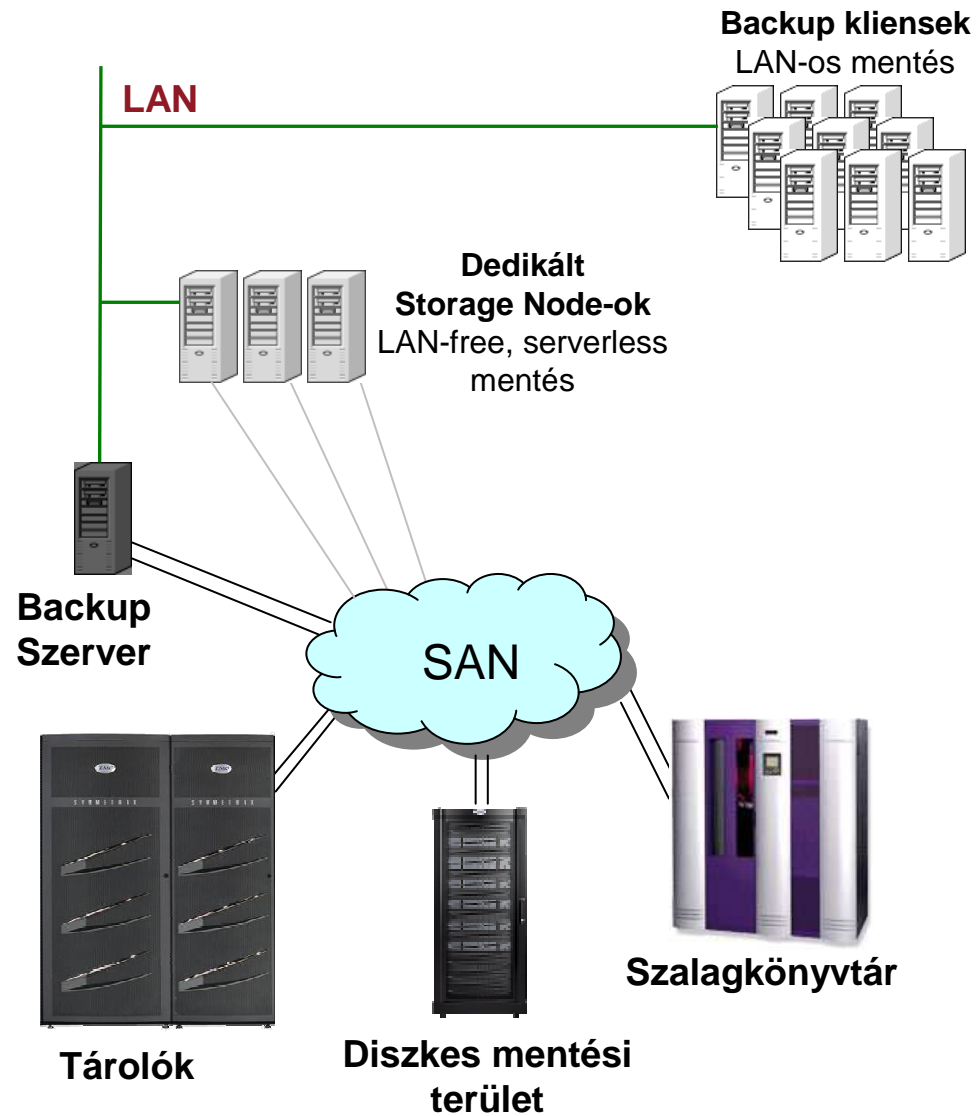
Szalagos rendszerrel együtt járnak a nehézségek

Diszkes mentési rendszerek előnyei

- Általában nagyobb mentési és visszaállítási sebesség
 - Mentési ablakba beleférünk
 - SLA-nak megfelelünk
- Megbízhatóbb magasabb rendelkezésre állás
 - Adatvédelem a megszokott lemezes technológiákkal (RAID5, RAID6)
- Egyszerűbb rendszermenedzsment
 - Automata finomhangolás
 - Központi média menedzsment
- Kisebb média kezelési veszély
 - Hatékony távoli replikáció médiák szállítása helyett
 - Adatok online tartása, offline médiák visszakeresése helyett
 - Szállítási költségek eliminálása



Hagyományos mentési környezet diszkes mentéssel



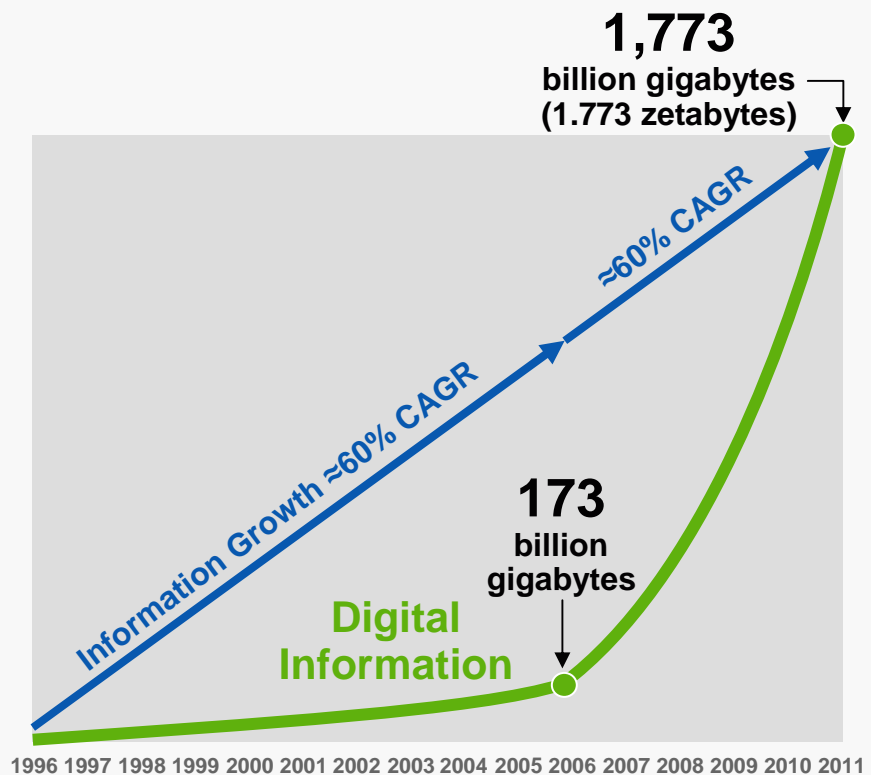
Hagyományos diszk vs. Virtuális szalagkönyvtár

	Diszk	VTL
Hardver felépítés	+	-
Implementáció	+	+
Adattárolási hatékonyság	-	+
Nagyvállalati alkalmazhatóság	+	++
Menedzsment	+	++
Teljesítmény	+	+
Kompatibilitás	+	++

Miért érdekes a deduplikáció?

- Az adatnövekedés elkerülhetetlen
- Exponenciálisan növekvő mentési rendszerek
 - Tipikusan a produktív rendszer 4-30-szorosára kell méretezni
 - Klasszikus napi/heti/havi (full) mentések megtartása évekre
- Egyre hosszabb ideig szükséges az adatokat megőrizni
 - Költséges médiamenedzsment, migrációs költségek, üzemeltetési költségek
- Folyamatos üzemű adatközpontok
 - Nincs jó időpont a mentésre
 - Sáv szélesség limitációk
 - Erősödő konszolidáció új megközelítést igényel

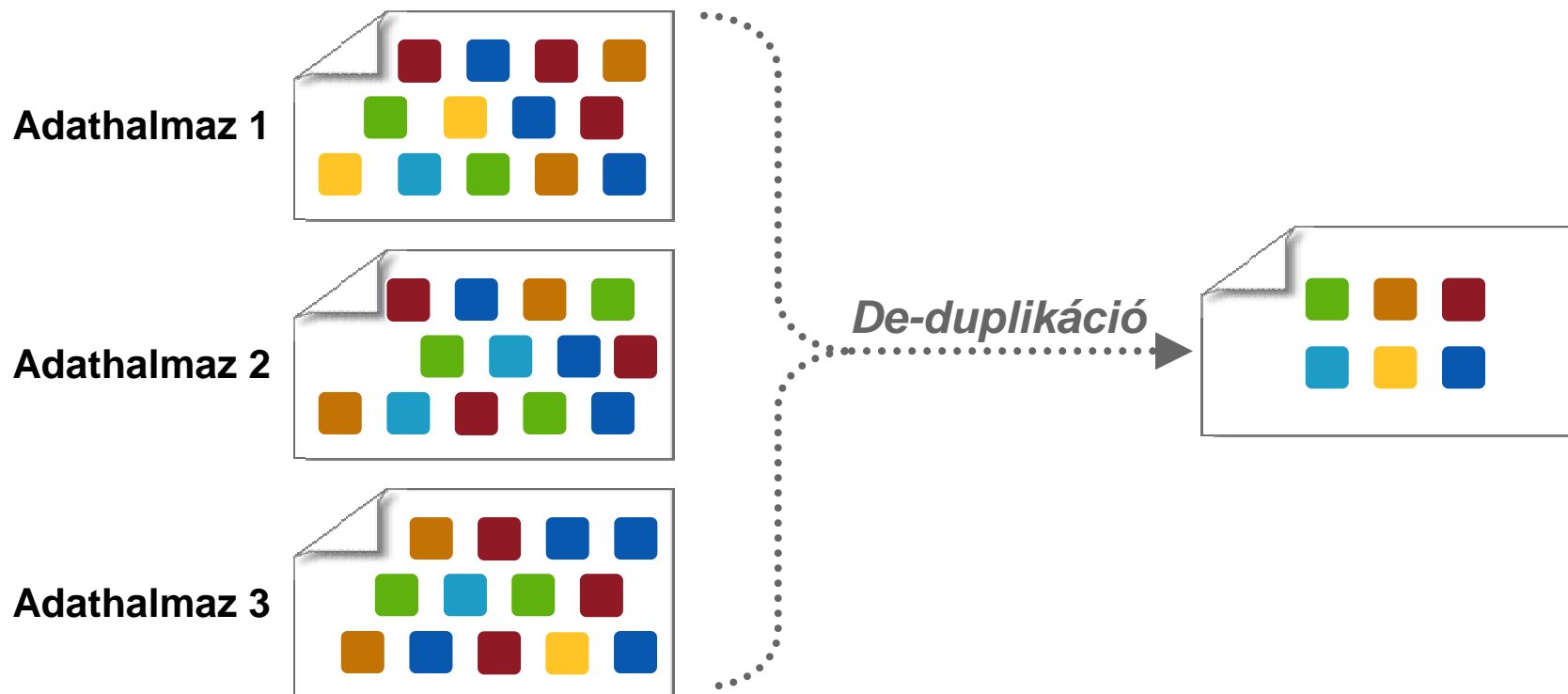
AMOUNT OF DIGITAL INFORMATION CREATED AND REPLICATED EACH YEAR



Source: IDC White Paper, "The Diverse and Exploding Digital Universe", March 2008 – Sponsored by EMC

EMC definíciója a deduplikációra

“Az a folyamat, amely során megkeresik és azonosítják az egyedi adatrészleteket a kezelt adathalmazon belül, és a tárolás vagy adatmozgatás esetén megszüntetik a redundanciát.”



Előtte: 39 szegmens

Utána: 6 egyedi szegmens

Deduplikáció felhasználási területei

- Mentés – nagy hatékonyság a sok azonosság miatt
- Archiválás – kisebb hatékonyság, de hasznos lehet
 - Email vagy file archiválás
- Elsődleges tárolók – korlátozott felhasználási terület
 - File szerverek
 - Performancia problémák léphetnek fel
- Replikáció – sávszélesség jobb kihasználása végett

Deduplikáció példa

1 Első változat

4/1/2006 EMC Corporation, Technology Solutions

Table of Contents	A	B
Introduction		
Report Organization		
Executive Summary		
Findings Highlights		
Recommendation Highlights		
Assessment Overview		
Assessment	C	D
Assessment Findings		
Server and Data Amounts		
Amount of Data Growth		
Time of Data on Servers		

2 Duplikált változat

4/1/2006 EMC Corporation, Technology Solutions

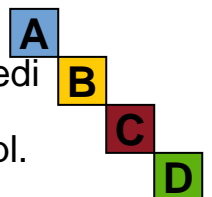
Table of Contents	A	B
Introduction		
Report Organization		
Executive Summary		
Findings Highlights		
Recommendation Highlights		
Assessment Overview		
Assessment	C	D
Assessment Findings		
Server and Data Amounts		
Amount of Data Growth		
Time of Data on Servers		

3 Módosított verzió

2/8/2006 EMC Corporation, Technology Solutions

Table of Contents	E	B
Introduction		
Report Organization		
Executive Summary		
Findings Highlights		
Recommendation Highlights		
Assessment Overview		
Assessment	C	D
Assessment Findings		
Server and Data Amounts		
Amount of Data Growth		
Time of Data on Servers		

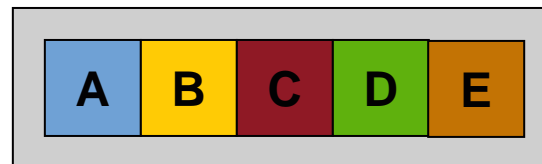
Csak az egyedi részeket mozgatjuk (pl. mentjük)



Adatmozgatás nem történik

E

Az megváltozott szegmenst eltávolítjuk



De-duplikált tárolás

- Objektum szintű: több azonos objektum esetén egy másolatot tárol (single instancing)
 - Csak egy másolat kerül eltárolásra és mozgatásra
 - A deduplikált állomány helyén csak egy „csonk” marad egy pointerként működik

- Rész-objektum szintű: Egyes objektumok részszegekből is csak az egyedi részek kerülnek eltárolásra
 - Minden objektumot szeletekre osztanak, és egyedi azonosítóval látnak el.
 - Minden egyedi szelet eltárolásra kerül az eredeti állományok visszaállításához szükséges leíró adatokkal
 - Blokk méret:
 - Fix hosszúságú, kevésbé hatékony, egyes esetekben jól használható
 - Változó hosszúságú, hatékonyabb a strukturálatlan adatoknál
 - Mindkét esetben általában a kisebb blokkméret hatékonyabb deduplikációt eredményez

- Mi lehet egy objektum?
 - Egy levél, fájl, mentési egység (save set), file rendszer, egy LUN...

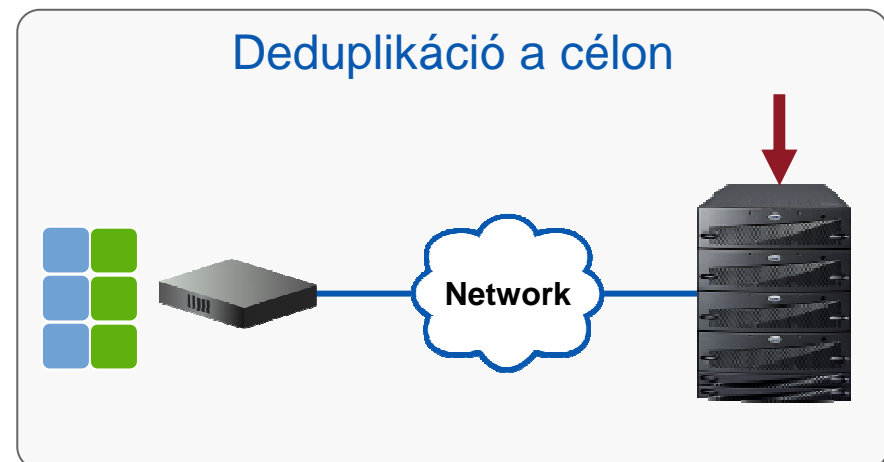
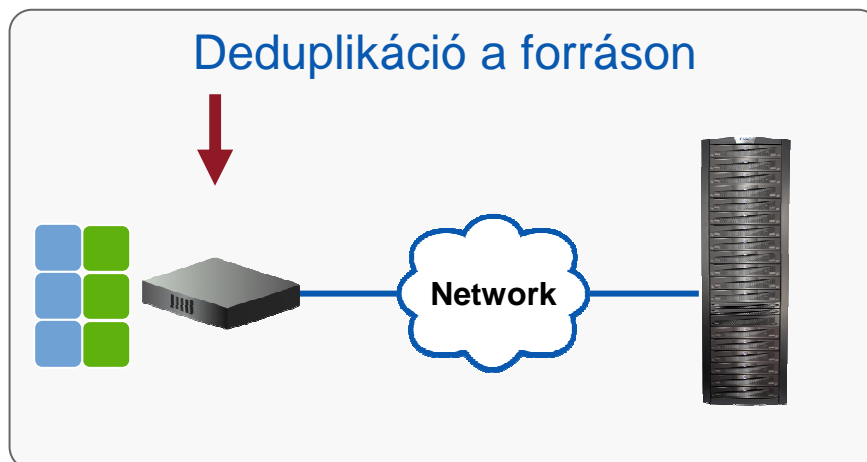
Hol történik a deduplikáció?

Forráson

- Kliens szoftver azonosítja az ismétlődő adatrészleteket már a forráson
- Csak az új egyedi adatrészletek kerülnek továbbításra a hálózaton majd ezeket az egyedi adatokat tárolják a mentési rendszerben
- Gyorsabb mentést és kisebb hálózati sávszélességet eredményez

Célon

- Mentési rendszer az eredeti adatokat küldi a tárolóra
- A deduplikáció csak akkor kezdődik ha az adatok elérték a tárolót
- Deduplikáció vagy mentés közben vagy utólag történhet
- VTL vagy B2D rendszereknél alkalmazott
- Mentési rendszereknek nem érzékelhető megoldás



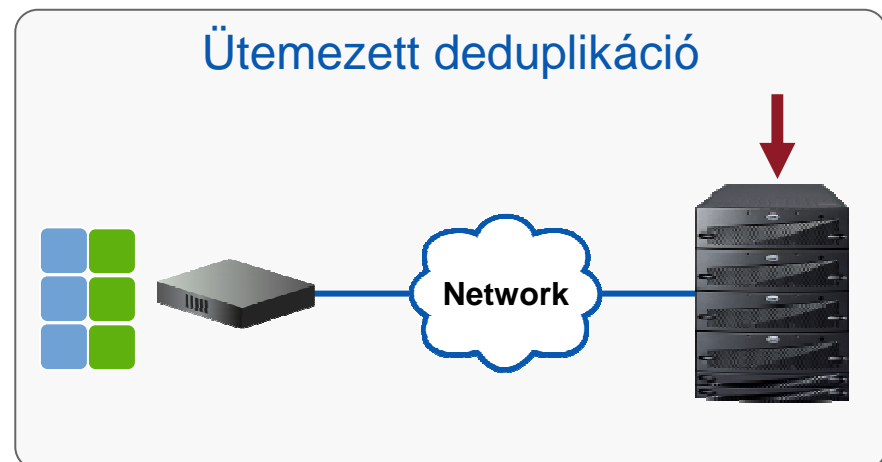
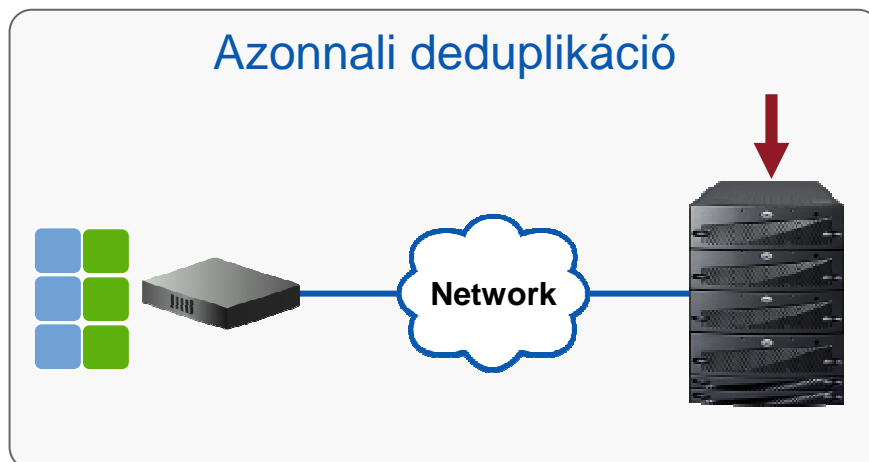
Mikor történik a deduplikáció?

Azonnal – A mentési folyamat közben

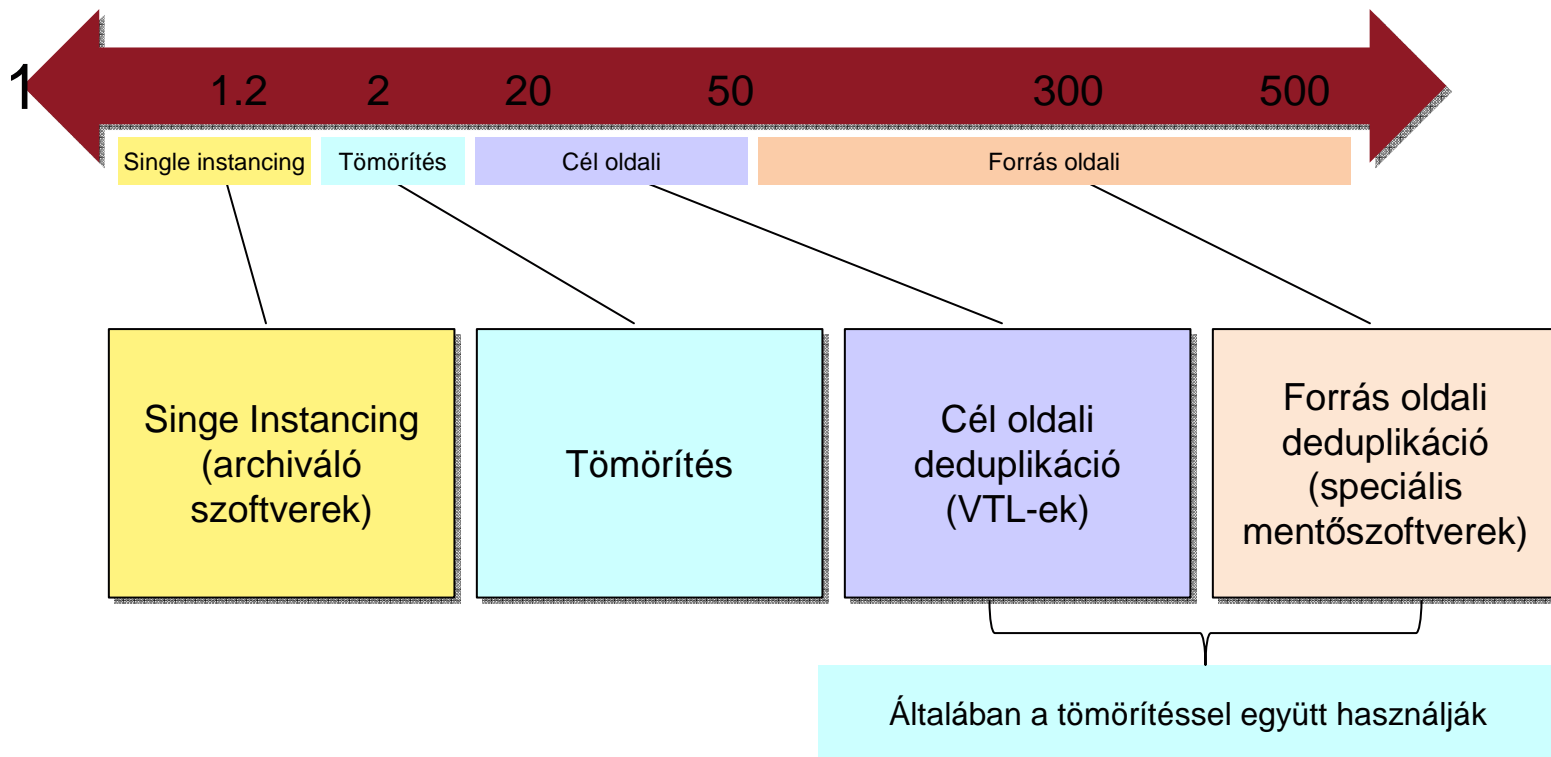
- Ideális olyan esetekben ahol a mentési ablak nem korlátozza a mentést
- Kapacitás optimalizáció maximális

Utólag – A mentési folyamat befejezése után

- Tartalom eredeti formában letárolódik, majd utólag kerül deduplikálásra
- Ideális olyan esetekben ahol a sebesség a legkritikusabb



A deduplikálás hatékonysága



Hatékonyságot befolyásoló tényezők

- Adat típusa
 - „Természetes” adatforrás (audió, videó, szkennelt képek) egyedi; Felhasználók által előállított (doc, ppt, pst...) gyakorta tartalmaz duplikációt (templates, etc.)
 - Titkosított és tömörített adatok nem jól deduplikálhatók
- Adat változási ráta
 - Kis adatváltozási ráta nagyobb mennyiségű deduplikálható adatot eredményez a teljes mentésekben
- Adattárolási szabályok
 - Ha az adatok hosszabb ideig megőrzésre kerülnek
Nagyobb a valószínűsége hogy többször kerülnek azonos adatok lementésre
- Mentési stratégia
 - Több teljes mentés növeli az azonos adatok előfordulását

Deduplikációs hatékonyságot számos tényező befolyásolja – még egy kicsiny eltérés az egyes rész-területen is jelentős változást hozhat.

Deduplikáció hatása a sávszélesség használatra

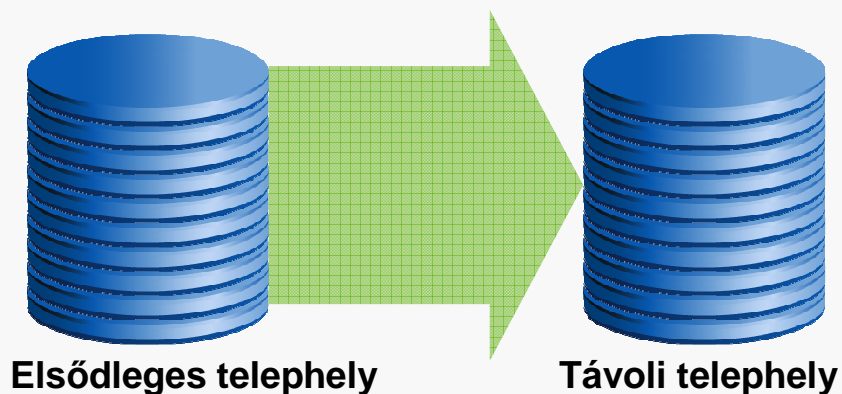
Deduplikáció nélkül

- Nem csökken a helyi tárhely igény
- Nincs sávszélesség megtakarítás
- Nincs megtakarítás a másodlagos telephelyen

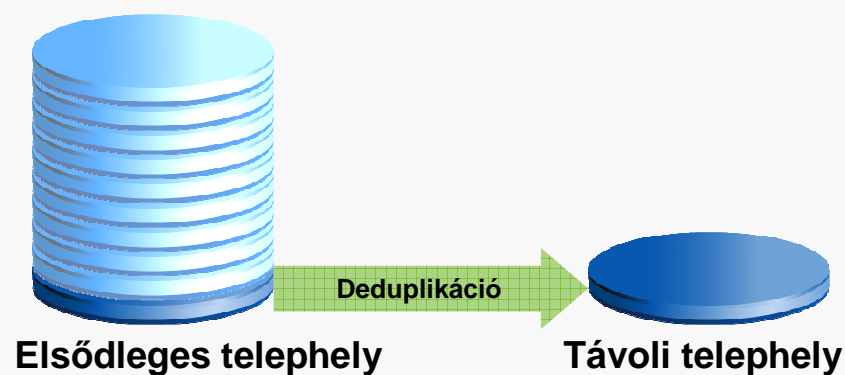
Deduplikációval

- Mentési rendszer tárhely optimalizáció
- Sávszélesség és idő megtakarítás
- Tárhely megtakarítás a másodlagos telephelyen


Távoli replikáció deduplikáció nélkül

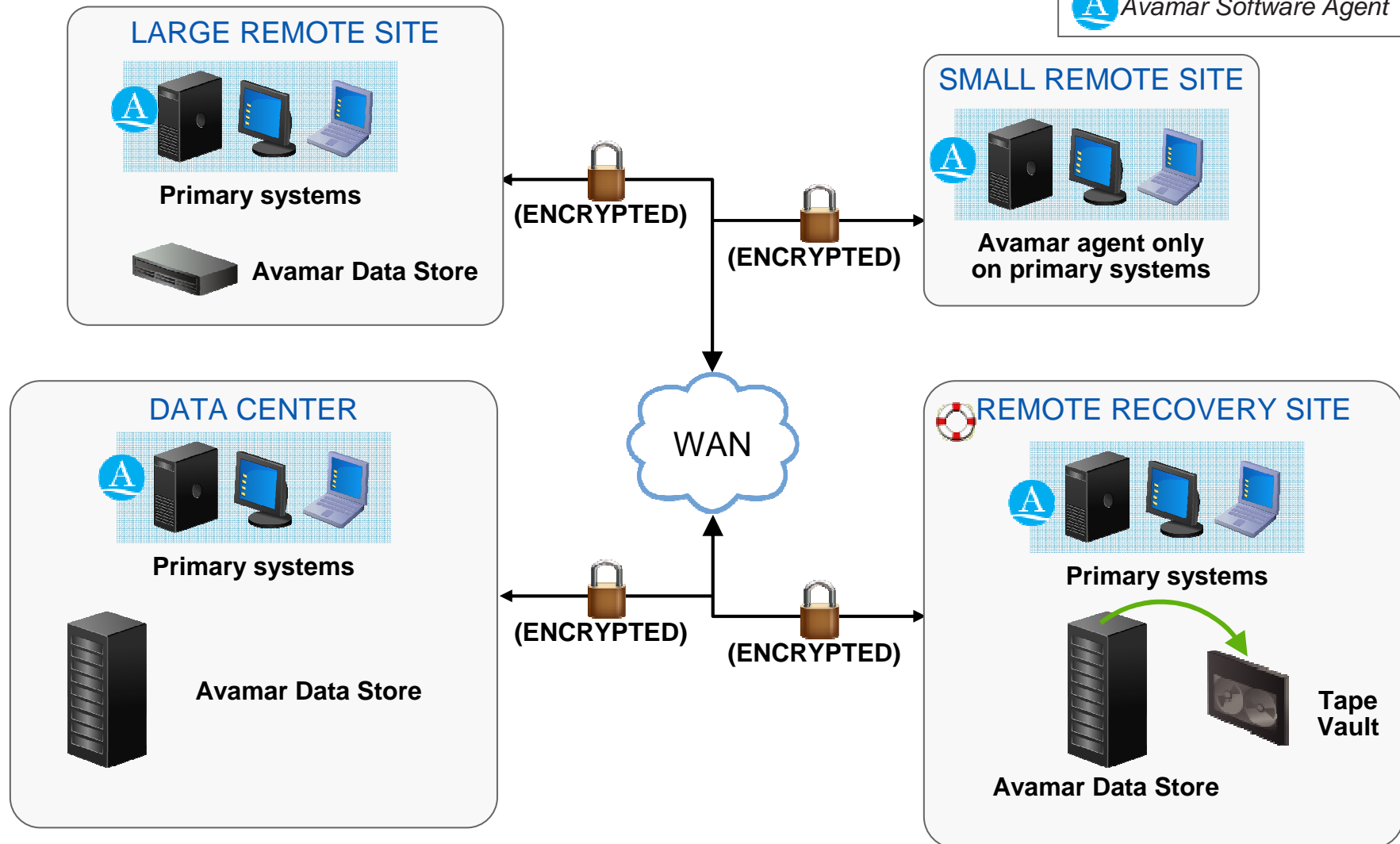


Replikáció deduplikáció után



Példa: deduplikációs mentőszoftver

 Avamar Software Agent



Néhány konkrét deduplikációs eredmény (Avamar)

Ügyfél	Adat összetétel	Redukciós faktor	Éles adat mennyisége (GB)	Mozgatott adat mennyisége (GB)
A	Windows fájl rendszer	586:1	3,573	6.1
B	Windows, Linux és UNIX fájl rendszerek	436:1	5,097	11.7
C	Mérnöki/gyártási adatok NAS tárolórendszeren	135:1	3,265	24.2
D	20% adatbázis, 80% fájl rendszerek (Windows and UNIX)	120:1	9,583	80.0
E	Linux fájl rendszerek és adatbázisok	75:1	7,831	104.2

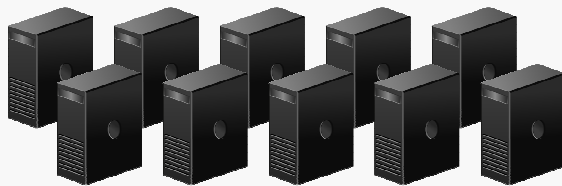
Source: EMC

Deduplikáció és szerver virtualizáció

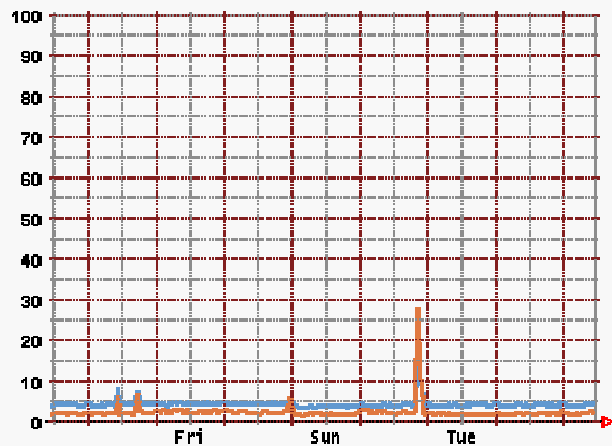
A szervervirtualizáció megköveteli a mentési rendszerek felülvizsgálatát

Hagyományos szerver

Általában kevésbé terhelt hardver, backup számára elegendők az erőforrások



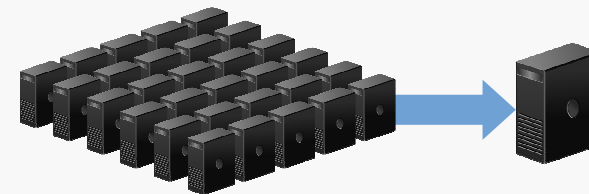
% CPU



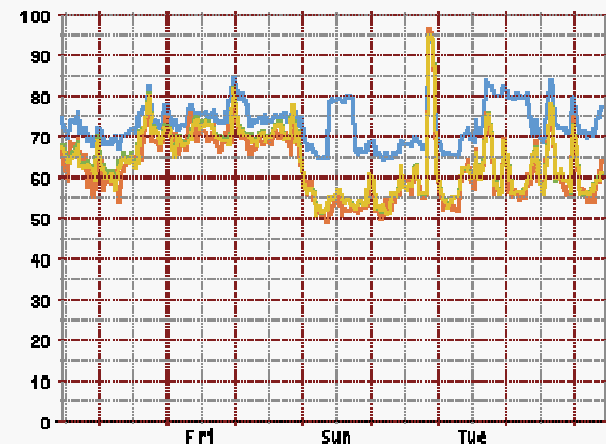
vcpu0 ready0

Virtualizált szerverek

A hardver folyamatosan erősen leterhelt, a mentésre nincs sok erőforrás



% CPU Busy (8 LCPUs w/HT)



cpu0 cpu1 cpu2 cpu3

EMC deduplikációs mentési megoldások különböző igényekhez

EMC Avamar

- Piacvezető mentési rendszer a piacon létező leghatékonyabb forrás oldali deduplikációval
- Egyszerű visszaállítás
- Teljesen újszerű technológia

EMC NetWorker

- Átfogó mentési megoldás
- Teljes-körű integrációs biztosít a különböző deduplikációs megoldásokhoz

Disk Library termékcsalád

- Piacvezető diszk alapú mentési megoldás deduplikációs képességekkel
- Együttműködik a meglévő mentési rendszerekkel
- Rugalmas megoldás a kisvállalatoktól a nagyvállalatokig
- Nagy teljesítmény és megbízhatóság

Támogatott kliensek

Microsoft Windows.net Server 2003 • Red Hat • Solaris • Suse • AIX • HP • Mac OS • VMware • Windows • Vista Novell

Mentési rendszerek

EMC NetWorker • Symantec NetBackup • Symantec Backup Exec • IBM TSM CommVault Galaxy • HP Data Protector • CA ArcServ • and more

Avamar Software



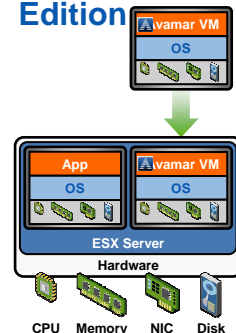
Telepíthető kvalifikált szerverekre

Avamar Data Store



Integrált hardver-szoftver megoldás

Avamar Virtual Edition



VMWare virtuális gép támogatás

NetWorker



DL3D 1500



Belépő szintű deduplikációs mentőeszköz

DL3D 3000



Középkategóriás deduplikációs mentőeszköz

DL3D 4000



SAN alapú VTL deduplikációva

EMC²[®]

where information lives[®]

DL3D 1500 és DL3D 3000

Az új disk alapú deduplikációs mentési megoldás LAN-os rendszerekhez

- DL3D 1500
 - CLARiiON CX3-10 háttértár
 - 4–36 TB kapacitás
 - 720 GB/órás csúcsteljesítmény (SAN)
- DL3D 3000
 - CLARiiON CX3-40 háttértár
 - 8–148 TB kapacitás
 - 1,44 TB/órás csúcsteljesítmény (SAN)
- Paraméterezzhető deduplikáció
 - Objektum szintű változó hosszúságú deduplikáció
 - Eltávolítja a redundáns adatokat a mentésekből és így csökkenti a szükséges tárhelyet. Ezáltal több mentést lehet online tartani, illetve lecsökkenti a replikációhoz szükséges sávszélességet
- Replikálás a tartalom deduplikálása után
 - Akár 10 különböző rendszer is képes egy központira replikálni az adatokat
 - Ki- és bekapcsolható adat titkosítás: 128-bit AES



DL3D 1500

6 GE port
CIFS/NFS
2 FC port SAN
ports (VTL)
4 TB induló
kapacitás
3 év garancia



DL3D 3000

8 GE port
CIFS/NFS
4 FC port SAN
(VTL)
8 TB induló
kapacitás
3 év garancia

DL4000 termékcsalád

A legnépszerűbb SAN alapú VTL-je most már deduplikációs képességekkel

- CLARiiON CX3-80 alapú tároló rendszer
 - Single vagy dual engine systems
 - 99,999% rendelkezésre állás
 - RAID 6
- PB-nál is nagyobb használható bruttó kapacitás
 - 1 TB SATA diszkek
 - 930 diszkiig bővíthető
- Nagy teljesítmény
 - Hardver tömörítés
 - End-to-end 4 Gb/s rendszer
- Hatékony deduplikációs képességek
 - Nagy sebesség a hatékony adattárolás megtartásával
- Energiatakarékosság
 - Alacsony fogyasztású diszkek spin-down képességekkel



DL4000 Series

*Industry's only virtual tape library,
built from the ground up with
4 Gb/s components*

EMC Centera — Az archiválási megoldás

- Tartalom címzésű adattároló (CAS)
- Objektum szintű deduplikáció
- Hosszútávú hiteles adattárolás online elérhetőséggel
 - Adattárolás a diszkek sebességével és az optikai rendszereknél megszokott hitelességgel
 - Adatmegőrzés akár évtizedekre
- Egyszerű menedzselhetőség
- Nagyfokú integráltság
 - Központi archívum akár különböző rendszerekhez
 - Dokumentumkezelő, képalkotó, mentési és archiválási rendszerek, ERP rendszerek támogatása
 - Nyílt szabványos csatolófelületek (pl. FTP, NFS...)
- Különböző állami és uniós megfelelési tanúsítványok



Adat deduplikáció NetWorkerrel

Deduplikáció forráson vagy akár a célon

- Integrált deduplikációs megoldások
- Avamarral integrált NetWorker kliens
 - Forrás oldali deduplikáció
- VTL támogatás és integráció
 - LAN-alapú mentési rendszerek (DL1500 vagy DL3000)
 - SAN-alapú deduplikációs rendszerek DL4000
- Teljes körű mentési rendszer menedzsment
- NetWorker FastStart – Deduplikáció kész rendszer!

