

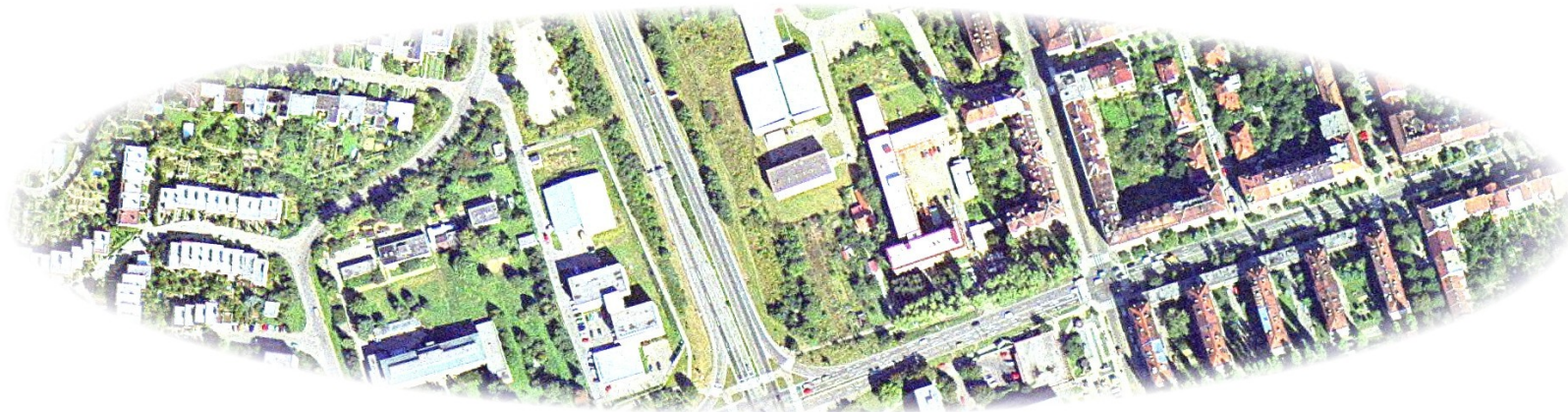


ISI
ASCR

INSTITUTE OF SCIENTIFIC INSTRUMENTS
ACADEMY OF SCIENCES OF THE CZECH REPUBLIC
Department of Coherence Optics, www.isibrno.cz/kgs

Systemy zálohování dat z velkého počtu uživatelských PC a serverů

Ondřej Číp, Ph.D., ÚPT AV ČR
ocip@isibrno.cz



Hlavní problémy při zálohování dat PC a serverů

1. Nárůst kapacity diskových pamětí na lokálních PC i centrálních serverech je v posledních 10 letech několikanásobně vyšší než nárůst kapacity zálohovacích médií
2. Přenosová kapacita LAN i WAN sice roste, ale bývá zahlcena přenosem redundantních zálohovaných dat
3. Při velkém množství zálohovaných dat rostou i nároky na strojový čas a energetický příkon zálohovaných PC a serverů – např. dlouhá doba zálohování jednoho PC
4. V řadě případů jsou zálohovaná data z jednotlivých PC či uživatelských síťových adresářů velmi podobná, při klasickém zálohování se tedy zálohují vícekrát stejná data -> příklady...

Příklady redundantních zálohovaných dat aneb Co často „*bohužel*“ zálohujeme?

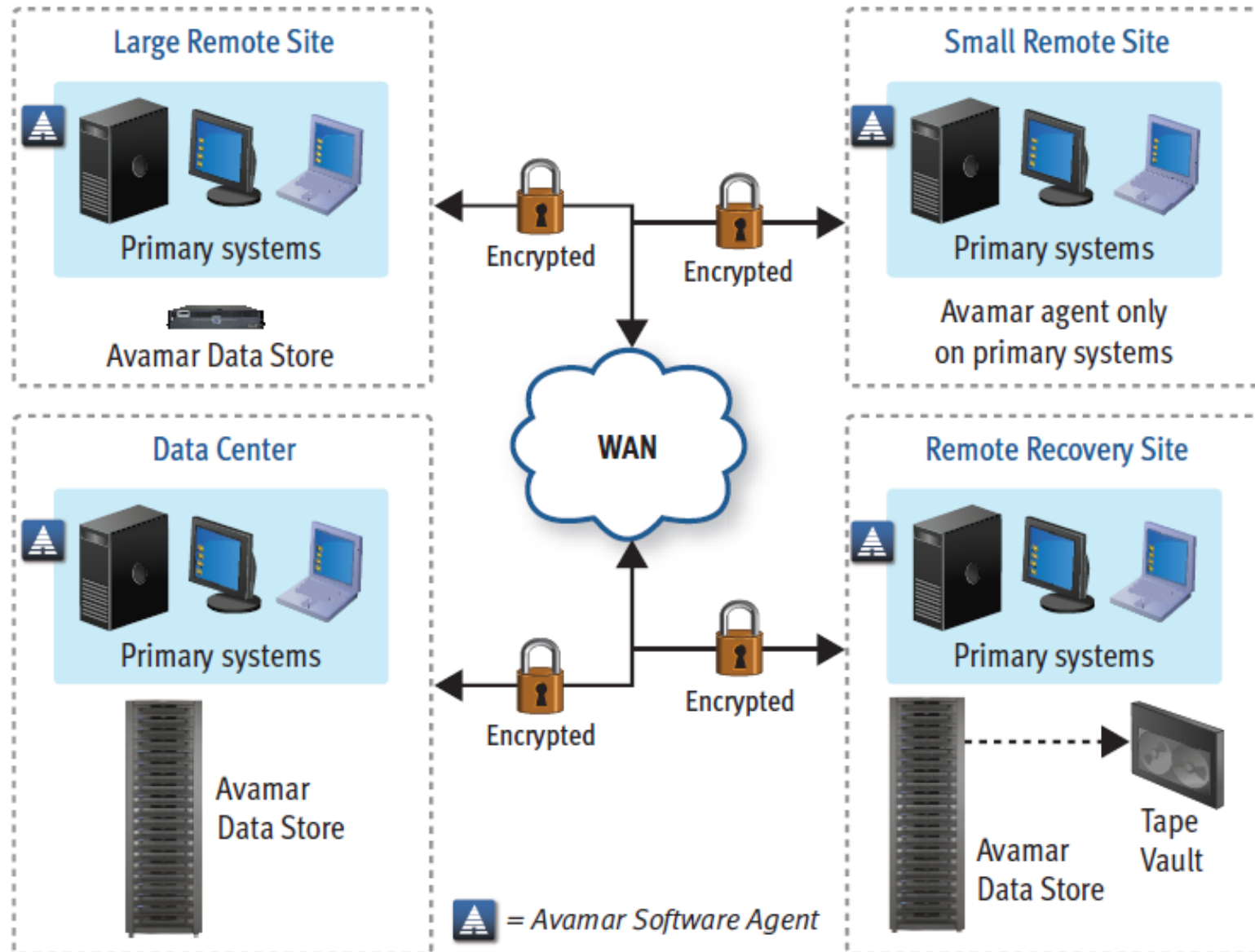
1. Operační systém uživatelského PC – např. WinXP, Win7, Lin RedHat, Mandrake, atd., běžné aplikace uživatele PC – MS Office, OpenOffice, Matlab, LabView, atd.
2. E-mailové zprávy (PC i servery) – hromadně doručená zpráva (např. stejného znění s 10MB videoklipem) se uloží do stovek schránek, např. při počtu 100 adresátů se zálohuje $100 \times 10\text{MB} = 1 \text{ GB}$
3. „Giga“ balíky instalačních souborů aplikací – častý nešvar, atd.
4. Giga soubor, kterému se změnil první 2 byty se zálohuje celý, neboť má jinou délku a časovou značku (platí u inkrementálního zálohování)

Deduplikace

- „kladivo“ na kompresi zálohovaných dat

1. Vyhledání shodné sekvence dat v rámci jednoho souborového systému (např. 1 PC, 1 server) – soubor však není elementární jednotka – základem je rozdělení souborového systému na segmenty a ty se porovnávají a vyhledají se neredundantní **tzv. deduplikované** sekvence
2. Vyhledání shodných segmentů v rámci uživatelských PC a serverů – přenáší se po síti pouze jedinečné deduplikované sekvence a jedinečná data
3. Zálohuje se pouze deduplikovaná a jedinečná data na centrálním či distribuovaném zálohovacím zařízení (pásky, disky, atd.). Systém si pamatuje kde má co být případně zrestaurováno

System EMC² AVAMAR



System EMCC AVAMAR

1. Primárně jde o SW systém, který má distribuovanou strukturu řízení, lze zakoupit i HW se zálohovacím zařízením (není však podmínkou)
2. Avamar software agent se nainstaluje jak na zálohované stanice a servery, tak i na servery se zálohovacím zařízením
3. Systém pracuje na pozadí dle stanoveného postupu, deduplikovaná data jsou přenášena v době menšího zatížení sítě
4. Celkem průhledná licenční politika – platí se za množství zálohovaných deduplikovaných dat (1000GB přibližně 500 tis. – bez HW)

Otázky k diskusi

1. Řešíme problém zálohování dat velkého počtu PC a serverů. Spadá podpora do kompetence KIT?
2. Jaká jsou úskalí tohoto systému? Zejména otázka bezpečnosti dat a jejich možného zneužití
3. Který systém je (nej)lepší? Kromě Avamar je dále IBM a více než 5 dalších systémů
4. Bylo by možno uspořádat v rámci školení správců sítí AV ČR přehlídku nabízených systémů? (školení bývá pravidelně pod záštitou SSC)
5. Zkusit např. pilotní projekt?