

Záloha a obnova dat datových center

Backup and Restoring Data in Data Centers

Bc. Jiří Kavan

Diplomová práce
2012



Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Jiří KAVAN**
Osobní číslo: **A10870**
Studijní program: **N 3902 Inženýrská informatika**
Studijní obor: **Informační technologie**

Téma práce: **Záloha a obnova dat datových center**

Zásady pro vypracování:

1. Zpracujte teoretický přehled problematiky zálohování a obnovy dat, vytvořte přehled v současnosti používaných řešení.
2. Proveďte analýzu současného stavu zálohování dat v rámci příslušného organizačního celku.
3. Na základě předcházející analýzy navrhnete způsob zálohování a obnovy dat těchto zdrojů.
4. Realizujte řešení zálohování podle návrhu a popište použité technologie.
5. Uvedte zálohování do provozu v rámci konkrétního organizačního celku a otestujte její funkčnost.
6. Vypracujte závěrečné hodnocení a navrhnete směry budoucího vývoje.

Rozsah diplomové práce:

Rozsah příloh:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

1. **LEBER, Jody. Windows NT: zálohování a obnova dat : příručka pro správce systémů. Praha : Computer Press, 1998. 282 s. ISBN 8072261231.**
2. **LEIXNER, Miroslav. PC - zálohování a archivace dat : Nestůjte za dveřmi. Praha : Grada Publishing, a.s., 1993. 394 s. ISBN 8085424738.**
3. **KASTNER, Aleš. Zálohování a archivace. Praha : GComp, 1997. 128 s. ISBN 8085649586.**
4. **BASL, Josef. Podnikové informační systémy: podnik v informační společnosti. Praha : Grada Publishing a.s., 2008. 283 s. ISBN 8024722798.**
5. **PRESTON, W. Backup and recovery. Sebastopol, CA: OReilly, c2007. 729 s. ISBN 978-059-6102-463.**
6. **DE GUISE, Preston. Enterprise systems backup and recovery: a corporate insurance policy. Boca Raton: CRC Press, c2009. 306 s. ISBN 14-200-7639-6.**
7. **NELSON, Steven. Pro data backup and recovery. New York: Springer Science Business Media, 2011. 280 s. ISBN 14-302-2662-5.**
8. **ENGLEWOOD, Cliffs. Data storage management: a systems approach. NJ: Prentice Hall, 1998. ISBN 978-013-8901-875.**

Vedoucí diplomové práce:

Ing. David Malaník

Ústav informatiky a umělé inteligence

Datum zadání diplomové práce:

24. února 2012

Termín odevzdání diplomové práce:

21. května 2012

Ve Zlíně dne 24. února 2012

prof. Ing. Vladimír Vašek, CSc.

děkan



doc. Mgr. Roman Jašek, Ph.D.

ředitel ústavu

ABSTRAKT

V dnešní době je velmi důležité, ne-li přímo nutné, se dobře starat o počítačová data a chránit je před jejich ztrátou. Způsob, jak data chránit, je jejich zálohování. Právě zálohováním a obnovou dat se zabývá tato diplomová práce se zaměřením na vytváření velkokapacitních záloh. V první části práce jsou popsány možné způsoby a strategie zálohování a obnovy dat, používané principy a technologie. V praktické části je popsán konkrétní realizovaný projekt zaměřený především na zálohu velkého množství dat a víceúrovňovou bezpečnost uložení záloh. Je zde také kladen důraz na co nejkratší dobu obnovy malých částí.

Klíčová slova: záloha, obnova, velkokapacitní záloha, strategie zálohování

ABSTRACT

Nowadays it is very important, if not necessary, to take good care of computer data and protect them from their loss. The way to protect data is their backup. The aim of this dissertation work is to describe the ways of backing-up and restoring of data and it is focused on creating high-capacity backups. The first section describes possible methods and strategies for backing-up and restoring data, principles and technologies used by. The practical part describes the specific project implemented primarily to backup of large amounts of data and multi-level security of backup saving. There is also an emphasis on the shortest recovery time of small parts.

Keywords: backup, restore, high-capacity backup, backup strategy.

Na tomto místě bych rád poděkoval panu Michalovi Hrdličkovi z firmy MHM computer a.s. za rady, odbornou pomoc a ochotu, které se mi dostalo během období vzniku této práce. Můj dík patří také vedoucímu mé diplomové práce panu Ing. Davidu Malaníkovi, Ph.D. za vedení a důležité připomínky při řešení diplomové práce.

Prohlašuji, že

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji,

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně

.....
podpis diplomanta

OBSAH

ÚVOD	9
1 TEORETICKÁ ČÁST	10
1 HISTORIE ZÁLOHOVÁNÍ	11
2 DŮVODY K ZÁLOHOVÁNÍ	13
2.1 NEJČASTĚJŠÍ PŘÍČINY ZTRÁTY DAT.....	13
2.2 NÁSLEDKY ZTRÁTY DAT.....	14
2.2.1 Záchrana dat z pevného disku	14
2.2.2 Záchrana dat z diskového pole RAID	14
2.2.3 Záchrana dat z Flash disku a paměťové karty.....	15
3 ZÁKLADNÍ PŘEHLED PROBLEMATIKY ZÁLOHOVÁNÍ	16
3.1 ZÁKLADNÍ POJMY	16
3.1.1 Zálohování.....	16
3.1.2 Archivace	16
3.1.3 Synchronizace	16
3.1.4 Obnova	16
3.1.5 Komprese	17
3.1.6 Deduplikace.....	17
3.1.7 Časové okno	17
3.2 TYPY ZÁLOHOVÁNÍ	18
3.2.1 Nestrukturovaná	18
3.2.2 Plná záloha	18
3.2.3 Diferenciální záloha	19
3.2.4 Inkrementální záloha.....	20
3.2.5 Zrcadlová + Reverzně přírůstková záloha	20
3.2.6 Průběžná ochrana dat	21
3.3 ZÁLOHOVACÍ MÉDIA	21
3.3.1 Magnetická páska.....	21
3.3.2 Páskové technologie.....	22
3.3.3 Pevný disk	25
3.3.4 VTL – Virtual Tape Library.....	25
3.3.5 sVTL	26
3.3.6 DAS.....	27
3.3.7 NAS	28
3.3.8 SAN.....	29
3.4 NEJPOUŽÍVANĚJŠÍ ZÁLOHOVACÍ METODY	30
3.4.1 D2T (Disk-To-Tape)	30
3.4.2 D2D (Disk-To-Disk)	31
3.4.3 D2D2T (Disk-To-Disk-To-Tape).....	31
3.5 ROTACE ZÁLOHOVÁNÍ.....	32
3.5.1 Round Robin (schéma s jednou páskou pro každý den)	32
3.5.2 Grandfather-Father-Son (GFS)	33
3.5.3 Tower of Hanoi (Hanojská věž).....	33

3.6	ZABEZPEČENÍ PŘED VÝPADKEM SLUŽEB DATOVÝCH CENTER	34
3.6.1	Lokalita datového centra	34
3.6.2	Redundance	34
3.6.3	Failover	35
3.6.4	Cluster	35
3.6.5	Geocluster	37
3.7	PLÁN OBNOVY	38
3.7.1	Testování obnovy	38
3.7.2	Testování čitelnosti záloh	39
II	PRAKTICKÁ ČÁST	40
4	ANALÝZA PROSTŘEDÍ	41
4.1	ZADÁNÍ	41
4.2	STÁVAJÍCÍ STAV SYSTÉMŮ A JEJICH ZÁLOHOVÁNÍ	41
5	NÁVRH A REALIZACE ZBŮSOBU ZÁLOHOVÁNÍ	44
5.1	NÁVRH ZÁLOHOVÁNÍ DATOVÉHO CENTRA	44
5.1.1	Tvorba návrhu se splněním podmínek	44
5.1.2	Datové centrum – hardware	46
5.1.3	Výběr zálohovacího software	49
5.2	REALIZACE NÁVRHU	50
5.3	ZÁLOHOVACÍ PROSTŘEDÍ V PROVOZU	53
5.3.1	Instalace a GUI	53
5.3.2	Nastavení zálohovacích jednotek	56
5.3.3	Nastavení zálohovacích médií	56
5.3.4	Správa klientů	56
5.3.5	Vytvoření zálohy (procesu)	57
5.3.6	Nastavení způsobu zálohování	58
5.3.7	Logy a stav zálohovacích procesů	58
5.3.8	Notifikace	59
5.3.9	Obnova – ověření správného provedení zálohování	59
6	ZHODNOCENÍ PROJEKTU A SMĚR DALŠÍHO VÝVOJE	61
6.1	ZHODNOCENÍ	61
6.2	DALŠÍ VÝVOJ	62
6.2.1	Offline geocluster	62
6.2.2	Online geocluster	63
	ZÁVĚR	64
	ZÁVĚR V ANGLIČTINĚ	65
	SEZNAM POUŽITÉ LITERATURY	66
	SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK	69
	SEZNAM OBRÁZKŮ	71
	SEZNAM TABULEK	72
	SEZNAM PŘÍLOH	73

ÚVOD

V dnešní době již téměř neexistuje průmyslové odvětví nebo vědní obor, ve kterém by se nepoužívaly počítače. Zároveň se počítače staly nedílnou součástí každodenního života a to i v domácnostech. Nejsou to jen počítače samotné, které jsou tak důležitými a cennými pomocníky, ale, troufám si tvrdit, že tím nejdůležitějším a nejcennějším jsou informace ve formě dat, které se na těchto počítačích zpracovávají a ukládají. Právě proto je velmi důležité, ne-li přímo nutné, se o tato data dobře starat a chránit je před jejich ztrátou. Většinou platí, že čím větší společnost, tím větší množství dat, o které je nutné se postarat. Způsob, jak informace chránit, je jejich zálohování. Tato práce má za úkol přiblížit čtenáři problematiku zálohování a zdůraznit její důležitost. Popisuje možné způsoby a strategie zálohování a obnovy dat. Možností realizace zálohování a obnovy dat je několik. Je proto potřeba si před použitím konkrétní strategie a zakoupením technologie zároveň stanovit míru dostupnosti dat, jak dlouho zpětně budeme chtít data uchovávat, maximální dobu jejich případné obnovy a další parametry. V závěru práce je popsán konkrétní realizovaný projekt zaměřený především na zálohu velkého množství dat a víceúrovňovou bezpečnost uložení záloh. Důraz je také kladen na co nejkratší dobu obnovy malých částí informací.

Pro příklad využití zálohování nemusíme chodit nijak daleko. Vezměme typický příklad z domácnosti. Pokud se například po návratu z exotické dovolené, kde jsme si nafotili spoustu krásných památečních fotografií digitálním fotoaparátem, porouchá záznamové zařízení, obvykle paměťová karta, ještě před tím, než jsme si stihli udělat jejich kopii čili zálohu, zbydou nám, místo zaznamenaných vzpomínek akorát oči pro pláč. Ale to samozřejmě ještě není nic proti tomu, když se ve velké korporaci několik let pracuje na nákladném projektu, z něhož jsou veškeré informace zaznamenávány a ukládány v počítači, který, těsně před dokončením projektu, vypoví službu a společnost tak přijde o veškerá data na něm uložená. V takovémto případě jsou ztráty nedozírné a přitom stačilo tak málo, zvolit vhodný způsob a data pravidelně zálohovat.

I. TEORETICKÁ ČÁST

1 HISTORIE ZÁLOHOVÁNÍ

Již při uvedení do provozu prvních počítačů vznikla otázka, jak a na jaké médium ukládat získané informace pro potřebu jejich opětovného využití. Tenkrát se ještě jednalo o relativně malé množství dat. Zálohování dat je vlastně vytváření jejich kopií nebo jejich přesun na nové úložiště. V polovině minulého století byla data uchovávána na děrných štítcích nebo děrných páskách. Tato média představovala v tehdejší době základní nosič informací pro v té době používané výpočetní stroje. Děrné štítky a pásky se používaly pro vkládání dat do počítačích strojů, ale také jako záloha, kterou šlo jednoduše přenést či přemístit do trezoru a zde uchovat.

Už v šedesátých letech dvacátého století se začala používat náhrada papírových štítků a pásek ve formě magnetických pásek. Tyto pásky byly o mnoho řádů rychlejší a měly mnohem větší kapacitu než děrné štítky. Později byly děrné štítky, jako záznamové zařízení, nahrazeny magneoptickými pevnými disky, které v mírně pozměněné podobě používáme v podstatě dodnes. Záznam na magnetickou pásku však dodnes zcela nevyrazil. Dnešní páskové jednotky byly výrazně vylepšeny a stále tak představují pro daný typ záloh jedno z nejpoužívanějších zálohovacích médií.

S dalším vývojovým typem záznamového zařízení přišla na konci padesátých let minulého století firma IBM, byly to, až dodnes zřejmě nejrozšířenější médium, pevné disky. První pevné disky samozřejmě nedosahovaly ani z daleka takových parametrů jako ty dnešní, a to jak parametrů kapacitních, tak i rychlostí ukládání a čtení uložených dat. V osmdesátých letech se stal pevný disk nedílnou součástí každého počítače.

I tato technologie je již překonána a nástupcem pevných disků v oblasti ukládání dat se stávají SSD (Solid-State Drive). Tato technologie je zase o něco dále v rychlosti přenosu dat. V jejím masivnějším rozmachu však brání zatím dost vysoké náklady na jejich pořízení.

Krátce po vzniku pevných disků byly vynalezeny diskety, vyvíjené především jako médium pro přenos dat mezi počítači, a proto vhodné pro zálohování. Z původních kapacit disket kolem 80 kB vznikly postupným vývojem diskety s nominální kapacitou 1,44 MB. Vznikali i další vývojové obdoby ve formě 100MB floppy a jednotky ZIP. V porovnání s disketami však měly jen krátké trvání.

Mezi další používaná média po disketách, které se dnes již téměř nevyužívají, byly a u osobních počítačů stále jsou CD a DVD, případně HD DVD a Blu-ray disky, paměti typu flash ve formě různých USB klíčenek, SSD disků, MP3 přehrávačů a jiných. Moderním způsobem zálohování je také ukládání dat přes počítačové sítě na speciální hybridní úložiště tvořená pevnými disky, magnetickými páskami nebo opět výše uvedenými optickými médii.

Na jakémkoliv nosiči informací ovšem hrozí riziko ztráty nebo poškození dat. Ruční vyhledávání a kopírování souborů s daty pro zálohování je poměrně pracné, a proto se pro zajištění snadného způsobu zálohování vyvíjí různé zálohovací aplikace a systémy.

Z hlediska požadavku na uchování dat není rozdíl v tom, zda je zálohování prováděno na počítačích doma nebo ve firmách či institucích. Rozdíl přichází ve způsobu provedení záloh. Realizují se různé typy záloh za využití různých technologií.

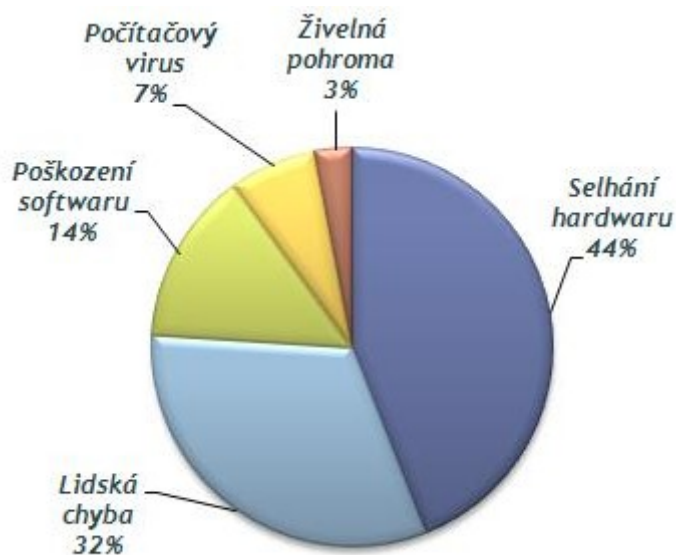
Většina programů pro zálohování umožňuje zálohu vybraných souborů a složek, případně nastavení systému. Pro běžného domácího uživatele počítače je to plně dostačující způsob zálohování. Ve veřejném či firemním prostředí to již však obvykle nestačí. Pokud je potřeba zajistit zálohování celého systému, který navíc musí běžet nepřetržitě, pak s pouhým zálohováním souborů nevystačíme a přicházejí na řadu jiné daleko složitější technologie. [9]

2 DŮVODY K ZÁLOHOVÁNÍ

Hlavní důvod k zálohování dat je jednoznačný, riziko jejich ztráty. O důležitá data bychom tedy měli pečovat s vědomím možných rizik, a to nejen při jejich úplné ztrátě, ale také při jejich poškození, krátkodobé nedostupnosti nebo jen prodlení při jejich dlouhotrvající obnově. Ke ztrátě dat může dojít z velkého množství příčin a některé mnohdy nejsme schopni ovlivnit.

2.1 Nejčastější příčiny ztráty dat

- Uživatelské chyby
- Chyby správce
- Hardwarové závady
- Softwarové chyby
- Viry a průniky do systémů
- Přírodní a jiné pohromy



Obrázek 1 Příčiny ztráty dat [21]

2.2 Následky ztráty dat

Každá ztráta informací – dat je nepříjemná. Čím důležitější data jsou, tím je i jejich ztráta citelnější, u menších a středních firem až existenční. Podobně to platí i o množství dat, o které z nějakého důvodu přijdeme.

Pokud příčina ztráty dat nenastala přímo fyzickou destrukcí úložného zařízení, máme ještě teoretickou šanci na jejich obnovu. Většinou to nezvládneme svépomocí, ale musíme to provést zprostředkovaně prostřednictvím specializovaných firem zabývajících se touto problematikou. Tento proces je samozřejmě velmi nákladný a při větším množství dat také velmi zdlouhavý. Navíc nám ani to ještě nezaručí stoprocentní obnovu všech ztracených dat.

Zde pro ukázkou uvádím velmi zobecněný přehled nejčastějších typů poškození dle náročnosti (ceny jsou převzaty z ceníku firmy zabývajících se touto problematikou):

2.2.1 Záchrana dat z pevného disku

Tabulka 1 Orientační ceny záchrany dat z pevného disku [10]

položka	cena od
lehké poškození např. poškození souborového systému FAT/NTFS	99 Kč
střední poškození např. vadná elektronika, úprava firmware	900 Kč
těžké poškození např. vadné hlavy, zadřená ložiska, spálený motor, poškozená záznamová vrstva, práce se servisní oblastí, souborový systém jiný než FAT/NTFS	3 000 Kč

Z tabulky 1 je patrné, že už u záchrany dat z lokálního počítače hodně záleží na míře závady, která nás o data připravila, ale ještě to nejsou žádné závratné částky. Tedy v případě, že je záchrana ještě možná.

2.2.2 Záchrana dat z diskového pole RAID

Tabulka 2 Plán záchrany dat z diskového pole RAID [10]

položka	cena od
Cena záchrany dat z diskového pole RAID je stanovována vždy individuálně.	individuálně
Posouzení možnosti záchrany dat z diskového pole je zdarma.	zdarma

Stanovisko specializované firmy v tabulce 2, řešit vše individuálně, je samozřejmě způsobeno obrovským množstvím rozličných typů polí od jejich složitosti přes množství modelů, jejich rozličné velikosti po jejich přístupnost. Malou útechou může být alespoň to, že je firma ochotna posoudit reálnost záchrany dat bezplatně.

2.2.3 Záchrana dat z Flash disku a paměťové karty

Tabulka 3 Orientační ceny záchrany dat z Flash disku a paměťové karty [10]

položka	cena od
SW závada např. USB flash disky, paměťová karta	600 Kč
HW závada např. USB flash disky, paměťová karta	1 000 Kč

Dodatek firmy k orientačním cenovým nabídkám:

„Uvedené ceny jsou **nejnižší možné**. Přesná cena bude určena po provedení prvotního posouzení časové a technologické náročnosti zakázky, které je zdarma.“ [10]

Samozřejmě čím víc budete na pokus o záchranu dat spěchat, tím víc si pochopitelně za služby spojené se záchranou důležitých informací připlatíte.

Jak je vidět i záchrana, byť jen jednoho jediného souboru nebo menšího množství dat něco stojí, a to ne jen po finanční stránce. Další věcí je čas strávený nad řešením problému okolo vyřizování možnosti záchrany dat a pokus o záchranu samotný. V neposlední řadě hraje roli i psychologické hledisko z obavy před ztrátou důležitých informací, které může ovlivnit nejen jednotlivce v rámci organizace, ale v samotném důsledku může ohrozit i firmu jako takovou.

3 ZÁKLADNÍ PŘEHLED PROBLEMATIKY ZÁLOHOVÁNÍ

3.1 Základní pojmy

3.1.1 Zálohování

Zálohování anglicky „backup“ je proces, při němž vzniká kopie zdrojových dat. Kopie zdrojových dat neboli datová záloha je obvykle uložena do jiného úložiště, než se nacházejí zdrojová data. Datová záloha může být komprimovaná nebo nekomprimovaná. V případě komprimované zálohy jsou obvykle záložní data uložena do souboru, jako jsou zip, rar, arj a další. Oproti archivaci dat, je při zálohování dat obvykle kladen důraz na možnost rychlé obnovy dat.

3.1.2 Archivace

Oproti zálohování dat, při archivaci není kladen důraz na možnost rychlé obnovy. Hlavním účelem archivace dat je dlouhodobé uchování takových dat, které již nejsou potřeba pro každodenní využití a u kterých se již nepředpokládá jejich další editování. Při archivaci dat se obvykle využívá komprimace. Archivace nemusí být chápána pouze jako dlouhodobé uložení, ale jako uložení s certifikací neměnnosti z pohledu zákona o archivnictví, k tomu jsou využity speciální archivační zařízení.

3.1.3 Synchronizace

Účelem synchronizace dat je udržování přesné kopie zdrojových dat v úložišti. Synchronizace může být prováděna online nebo podle časového plánu. Online synchronizace promítá do úložiště v reálném čase změny provedené ve zdrojových datech. Synchronizace podle časového plánu v nastavených intervalech kontroluje změny provedené ve zdrojových datech a tyto změny promítne do úložiště. [11]

3.1.4 Obnova

Obnovou dat, anglicky „restore“, se rozumí proces, při kterém dochází k obnovení souborů či složek nebo obojího ze zálohy nebo archivu. Je to proces přenášející data opačným směrem než při zálohování či archivaci. Je to stěžejní záležitost, kvůli které

zálohování vůbec provádíme, a která nám zajistí minimální negativní dopady v případě ztráty dat.

3.1.5 Komprese

Kompresa dat, také komprimace dat, je speciální postup při ukládání nebo transportu dat. Úkolem komprese dat je zmenšit datový tok nebo zmenšit potřebu zdrojů při ukládání informací. Obecně se jedná o snahu zmenšit velikost datových souborů.

Kompresa má v oblasti zálohování velký význam. Umožňuje provádět zálohování dat při současné úspoře místa na paměťovém médiu. Kompresní algoritmy obvykle využívají určité míry redundance dat v datových souborech.

3.1.6 Deduplikace

Deduplikace je speciální technika komprese dat, která zabraňuje ukládání stejných datových bloků na jednom úložišti. Deduplikační jednotka ukládá referenční informace o datové struktuře a díky tomu je schopná při zpětném čtení deduplikovaných dat zpět obnovit původní, komplexní informaci. Účelem deduplikace je úspora místa na datovém úložišti. Kromě varianty blokové deduplikace, existuje ještě deduplikace na úrovni souborů, kdy je ukládána pouze jedna kopie souboru.

3.1.7 Časové okno

Tento pojem označuje dobu, po kterou je možné provádět zálohování s ohledem na co nejmenší omezení produkčního prostředí společnosti. Používá se u automatických zálohovacích systémů, kde je jako jedna z konfiguračních položek kterou určuje provoz. Jeho výběr je závislý na požadavcích provozu a s rostoucím objemem jde o kritický parametr a je základním stavebním kamenem při řešení zálohovací architektury. Zpravidla se časová okna nastavují na noční hodiny mimo pracovní dobu. Aby byl totiž proces zálohování efektivní, je nutné mít k dispozici cílový počítač, pokud možno s ukončenými aplikacemi. Uzavřené aplikace jsou žádoucí z prostého důvodu. Pokud zálohovací server přistupuje k počítači, na němž je spuštěna aplikace s rozpracovaným souborem, ve kterém nejsou změny uloženy, tak pro něj nejsou změny v otevřeném souboru viditelné a nemůže je tudíž zazálohovat.

Perioda časových oken, tedy jak často budou časová okna po sobě následovat, je taktéž na administrátorovi, obvykle se nastavuje na každý pracovní den.

3.2 Typy zálohování

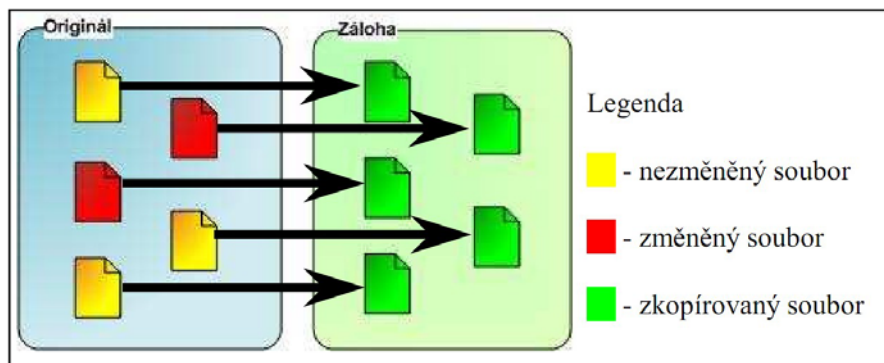
Způsobů zálohování existuje několik. Výběr vhodné metody zálohování vždy závisí na požadavcích, které vyplývají z mnoha ohledů. Záleží na tom, jakého charakteru data jsou, na kolik jsou důležitá nebo citlivá, jak velký rozsah dat je pro nás důležitý pro zajištění fungování organizace, která data jsou pro nás důležitá a která naopak můžeme postrádat, jak velký máme prostor pro zálohy a kolik času proces zálohování využije, kde bude záloha umístěna. Je rozdíl, zda zálohujeme doma nebo v zaměstnání, server nebo pracovní stanici. Je důležité vědět, jak velký výpadek pracovní činnosti si můžeme dovolit v případě potřeby obnovy dat po jejich ztrátě. Pokud se jedná o server, jak dlouho může být mimo provoz. Je třeba vhodně určit periodu zálohování. Ještě by se dalo pokračovat dál. Parametrů pro zvolení typu zálohování je mnoho. V každé situaci je potřeba se rozhodovat individuálně. Zde jsou dnes nejčastěji používané možnosti.

3.2.1 Nestrukturovaná

Nestrukturovanou zálohou je míněno například manuální nesystematické kopírování dat na různá média. Obvykle je pro zálohu použito větší množství médií (CD, DVD). Jednoduše řečeno, zálohuje se na to, co je zrovna po ruce. O záloze je tak uchováno, kromě dat, jen minimum informací. Jedná se o nejjednodušší způsob zálohování a nevyžaduje žádné speciální softwarové prostředky. Tento způsob zálohování je velmi obvyklý na domácích počítačích. Ve firemním sektoru se s touto zálohovací strategií obvykle nesetkáme. [12]

3.2.2 Plná záloha

Plná záloha, anglicky „full backup“, je základním a nejjednodušším typem zálohování vůbec. Během zálohování jsou všechny soubory uloženy do zálohových setů bez ohledu na to, zda byly změněny v době od poslední zálohy či nikoliv. Každá úplná záloha obsahuje vždy všechny soubory, složky či agendy. Princip plné zálohy je znázorněn na obrázku 2. Po technické stránce jsou uloženy všechny definované soubory. [11]

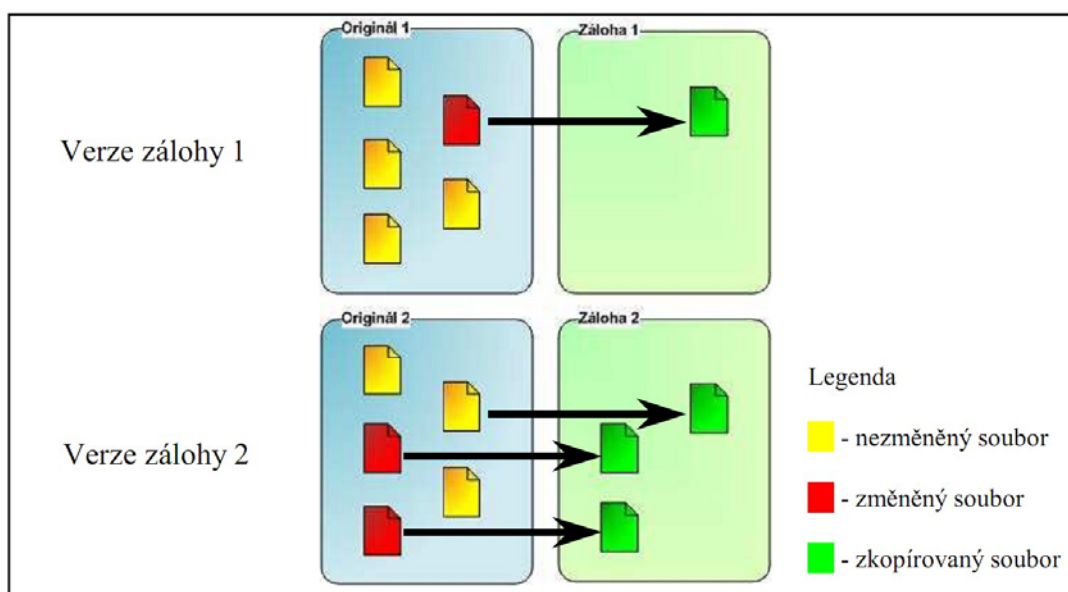


Obrázek 2 Princip plné zálohy

Obnova dat se provádí z libovolné úplné zálohy (obvykle z poslední) podle data, ke kterému se chceme pomocí obnovy dat vrátit. Nevýhodou je, že pokaždé musíme pracovat s velkým množstvím dat, protože každá záloha obsahuje všechna data. Proto se tato metoda většinou kombinuje s jiným typem záloh.

3.2.3 Diferenciální záloha

Diferenciální záloha neboli rozdílová záloha je taková záloha, která zaznamenává pouze data, která se změnila od plné zálohy. K plné záloze je možno vytvořit libovolný počet diferenciálních záloh, přičemž tyto diferenciální zálohy nejsou na sobě závislé. Tato metoda zálohování umožňuje mazat jednotlivé diferenciální zálohy, které již nejsou potřeba. Princip diferenciální zálohy je znázorněn na obrázku 3.

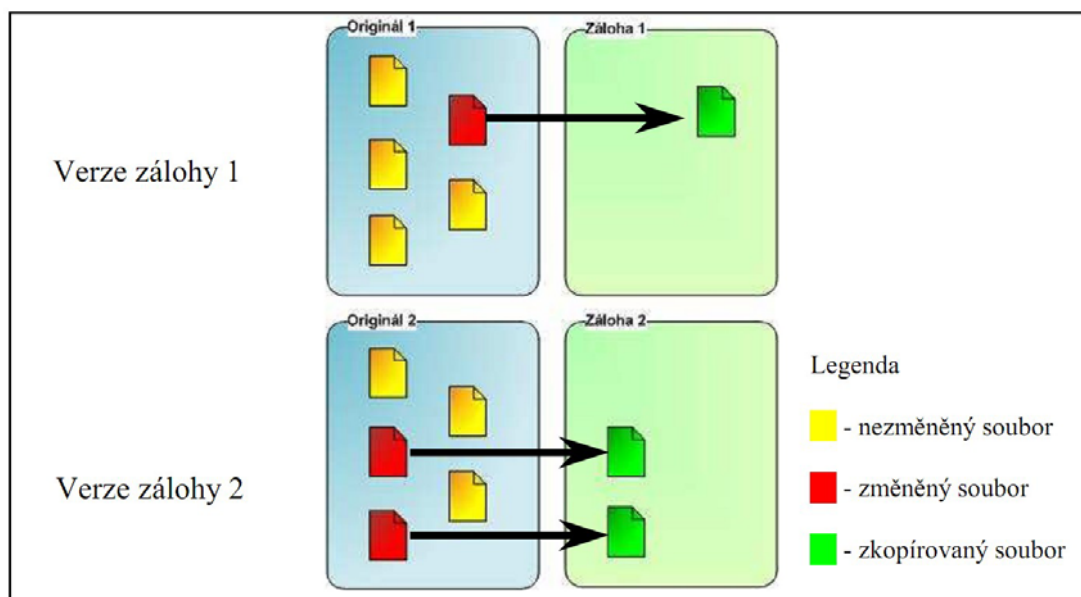


Obrázek 3 Princip diferenciální zálohy

Obnova dat se provádí z poslední úplné zálohy a poslední rozdílové. Obnova dat je tedy více odolná vůči poškození. Poškození jedné diferenciální zálohy nemá vliv na žádnou jinou zálohu. [11]

3.2.4 Inkrementální záloha

Inkrementální záloha neboli přírůstková záloha je taková záloha, která zaznamenává pouze data, která se změnila od posledního zálohování. Inkrementální zálohou je tedy možné zazálohovat nové změny, které ještě nejsou obsaženy v plné záloze nebo předchozí inkrementální záloze. K plné záloze je možno vytvořit libovolný počet inkrementálních záloh. Plné a inkrementální zálohy tvoří jeden řetězec, není tudíž možné smazat předchozí inkrementální zálohy nebo dokonce plnou zálohu. Princip inkrementální zálohy je znázorněn na obrázku 4.



Obrázek 4 Princip inkrementální zálohy

Obnova dat se provádí z poslední úplné zálohy a všech přírůstkových záloh od poslední úplné. Pokud je jedna inkrementální záloha poškozena, pak není možno obnovit ani zálohy následující za touto poškozenou zálohou. [11]

3.2.5 Zrcadlová + Reverzně přírůstková záloha

Tento model obsahuje zrcadlo reflektující stav systému po poslední záloze a historii přírůstkových záloh. Výhodou je, že máme neustále k dispozici aktuální plnou zálohu

a ukládáme pouze historii změn. Každé zálohování se automaticky promítá do zrcadla a soubory, které byly změněny, jsou přesunuty do přírůstkové zálohy. Tato metoda se nehodí pro přenosná media, protože každá záloha musí být provedena pomocí srovnání se zrcadlem. [12]

3.2.6 Průběžná ochrana dat

Tato metoda využívá místo plánovaných periodických záloh okamžitý zápis každé změny do žurnálu změn (logu). Provádí se ukládáním bytů nebo celých bloků dat místo ukládání celých změněných souborů. Průběžný záznam změn v žurnálu umožňuje získat obraz dat v minulosti. Naproti tomu prosté zrcadlení dat na druhý disk stav v minulosti nezachycuje. [12]

3.3 Zálohovací média

Z pohledu možnosti volby médií pro zálohování není příliš mnoho na výběr. Vývoj poslední doby ukazuje, že budoucnost na tomto poli patří magnetickému zápisu pro uchovávání dat. Lze sice najít i média používající jiné či hybridní technologie záznamu pro uchovávání dat, jako například magneto-optické disky nebo flash paměti, ale ty do trhu zabývajícím se zálohovacím médii prakticky nezasahují. Smysl má proto popisovat pouze nejčastěji využívaná média, tedy magnetické pásky a pevné disky.

3.3.1 Magnetická páska

Přestože je technologie magnetických pásek nejstarší technologií používanou pro záznam a uchování dat, díky svým výhodám je stále v oblibě. Poprvé byla pro záznam dat použita v roce 1951 na počítači UNIVAC I. Pásky zůstávají konkurenceschopnou alternativou pevným diskům vzhledem ke své nízké ceně. Ačkoli hustota záznamu je nižší než u pevných disků, dostupné místo na pásce je obvykle větší. Během celé své existence pásky poskytovaly velkou cenovou výhodu oproti magnetické diskové paměti, obzvláště pro vytváření záložních dat, kde je přenositelnost médií také velmi důležitá. Rychlý rozvoj velikosti disků a jejich ceny společně s pomalejší inovací pásek snížil podíl produktů magnetických pásek na trhu. Moderní kazety jsou ve formátech DAT/DDS, AIT, Digital Linear Tape a Linear Tape Open. Velikou výhodou pásek je poměrně dlouhá životnost. [13]

3.3.2 Páskové technologie

Existuje řada páskových zálohovacích technologií. Díky spolehlivosti a bezpečnosti uložených dat zvítězila lineární technologie zápisu do podélných stop (SLR, DLT, SDLT, LTO) nad technologií helical scan - tedy zápisem rotační hlavou (DAT/DDS, AIT).

Jednoznačně zvolit nejlepší z lineárních technologií záznamu dat asi není možné, protože vždy budeme muset nahlížet na volbu zálohovacího zařízení jak z finančního, tak technologického hlediska. [14]

V současné době jsou používány tyto technologie:

- **DAT/DDS** – Patří mezi nejpoužívanější zálohovací média z důvodu nízké pořizovací ceny zálohovacího systému a také nízké ceny pásek určených k zálohování. Proto je používána pro každodenní zálohy dat. Většina serverových operačních systémů obsahuje pro DAT/DDS (Digital Audio Tape / Digital Data Storage) drive ovladače. Tím je dána snadná přenositelnost dat mezi systémy. Nevýhodou tohoto média je na dnešní dobu nízká přenosová rychlost a malá kapacita médií. [10]

Tabulka 4 Kapacity a rychlosti DAT/DDS pásek [25]

Generace	Kapacita (GB)	Rychlost (MB/s)	Rok
DDS-1	1.3/2.0	0.18	1989
DDS-2	4.0	0.6	1993
DDS-3	12.0	1.1	1996
DDS-4	20.0	3.2	1999
DAT-72	36.0	3.2	2003
DAT-160	80	6.9	2007
DAT-320	160	12	2009

- **AIT** – jde o inovovanou technologii zápisu na 8mm pásku, kterou na trh uvedla firma Sony. Nový formát zápisu AIT (Advanced Intelligent Tape) ale není kompatibilní se staršími 8mm páskami firmy Exabyte.

Tabulka 5 Kapacity a rychlosti AIT pásek [26]

Generace	Kapacita (GB)	Rychlost (MB/s)	Rok
AIT-E Turbo	20	6	2004
AIT-1	25, 35	3, 4	1996
AIT-1 Turbo	40	6	2004
AIT-2	36, 50	6	1999
AIT-3	100	12	2001
AIT-3Ex	150	18	2006
AIT-4	200	24	2005
AIT-5	400	24	2006
AIT-6	800		<i>zrušeno</i>

- **SLR** – Historicky nejstarší technologie QIC byla zásluhou Tandberg Data zdokonalena v technologii SLR (Scalable Linear Recording). Kazeta se vyznačuje robustní mechanickou konstrukcí, která zajišťuje nejen přesné usazení v mechanice, ale znamená i lepší odvádění tepla. Je minimalizována možnost zničení záznamového média, páska zůstává stále uvnitř kazety. [14]

Tabulka 6 Kapacity a rychlosti SLR pásek [24]

Generace	Kapacita (GB)	Rychlost (kB/s)	Rok
SLR1	0,25	84.8	1986
SLR2	0,525	199	1988
SLR3	1	197	1990
SLR4	2,5	296	1992
SLR5	4	387	1997
SLR24, SLR6	12	1200	1997
SLR32, MLR1	16	1500	1996
SLR50, MLR3	20	2000	1997
SLR7	20	3000	1997
SLR60	30	4000	1998
SLR75	38	4300	2000
SLR100	50	5000	2001
SLR140	70	6000	2003

- **DLT** – Je považována za nástupce DAT pásky. Většímu rozšíření DLT (Digital Linear Tape) brání hlavně vysoké pořizovací náklady a poměrně drahá média. Předností je naopak značná kapacita pro uložení dat a vysoká rychlost zápisu. Uplatňují se hlavně tam, kde je prioritou zálohovat vysoké objemy ve velice krátké době. [10]

Tabulka 7 Kapacity a rychlosti DLT pásek [25]

Generace	Kapacita (GB)	Rychlost (MB/s)	Rok
DLT1	40	3	1999
DLT-VS80	40	3	2001
DLT-VS160	80	8	2003
DLT-V4	160	10	2005

- **SDLT** – Patří do skupiny zálohovacích mechanik s kapacitou pod 100 GB. Je to odlehčená verze nové generace DLT (Super DLT). Používá pásek s půlpalcovým formátem. Kazety mají jen jednu cívku, ta druhá je pevnou součástí streameru. [14]

Tabulka 8 Kapacity a rychlosti SDLT pásek [25]

Generace	Kapacita (GB)	Rychlost (MB/s)	Rok
SDLT 220	110	11	1998
SDLT 320	160	16	2002
SDLT 600	300	36	2004
SDLT 600A	300	36	2005

- **LTO** – Seagate, ve spolupráci s partnery Hewlett-Packard a IBM vytvořil LTO (Linear Tape-Open) technologii. Je to nová, výkonná, škálovatelná a otevřená pásková architektura, která bude následovat rostoucí požadavky na ukládání dat dosahujících kapacit až 1,6 TB na jednu cartridge v generaci 5 Ultrium formátu. Jde o otevřený formát a díky tomu je zde patrný konkurenční tlak na cenu. V současnosti jde o nejperspektivnější formát. Od LTO 3 je možnost využití nepřepisovatelných médií WORM (Write Once Read Many).

V současné době je LTO nejprodávanější a celosvětově nejpodporovanější řešení. Mechaniky LTO mají řadu benefitů. Jednou z důležitých vlastností je vzájemná kompatibilita mezi jednotlivými generacemi i mezi jednotlivými výrobci. Pokud podnik v současnosti použije pro zálohování dat mechaniku generace LTO2, budou jeho investice chráněny i do budoucna. Pokud situace za několik let vynutí upgrade na v budoucnu používaný formát (LTO6), má jistotu, že své LTO4 archivy přečte. [14]

Tabulka 9 Kapacity a rychlosti LTO pásek [25]

Generace	Kapacita (GB)	Rychlost (MB/s)	Rok
LTO-1	100	20	2000
LTO-2	200	40	2003
LTO-3	400	80	2005
LTO-4	800	120	2007
LTO-5	1600	180	2010

3.3.3 Pevný disk

Pevný disk, zkratka HDD (Hard Disk Drive), je zařízení, které se používá v počítačích a ve spotřební elektronice k dočasnému nebo trvalému uchování většího množství dat pomocí magnetické indukce. První komerční pevné disky se objevily v roce 1956, nejprve pro sálové počítače. Předchůdcem pevných disků je magnetická páska a magnetický bublen. Jejich současnými největšími konkurenty jsou SSD a USB flash disk, které využívají nevolatilní (stálé) flash paměti. Označení pevný disk, se uchytilo jako obecný pojem a obsahuje i SSD a další.

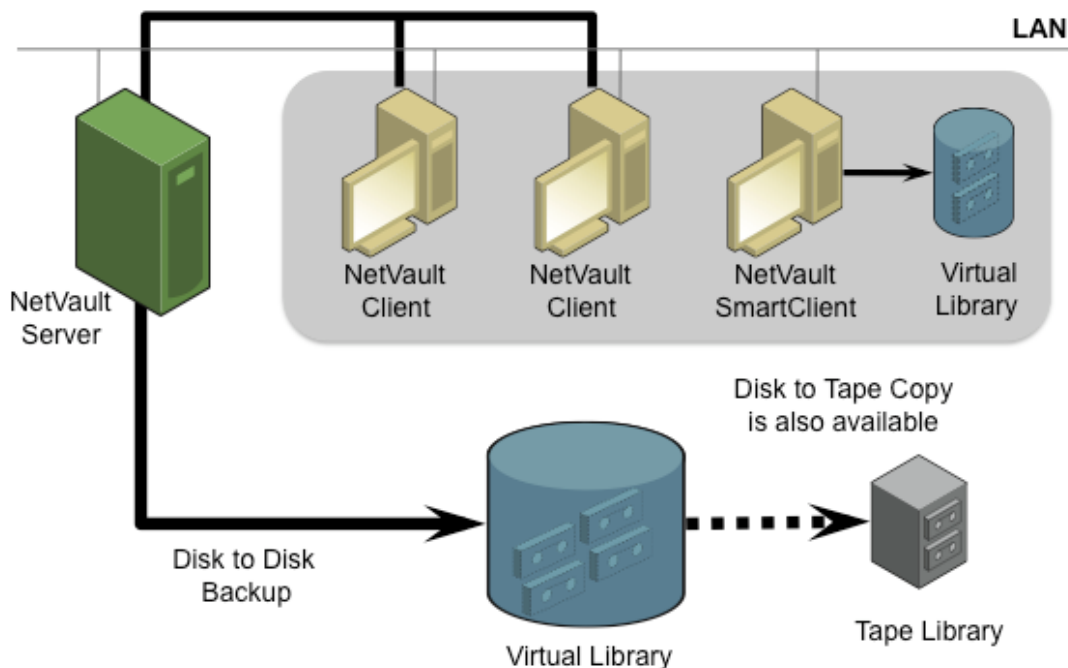
Hlavním důvodem velkého rozšíření pevných disků je velmi výhodný poměr kapacity a ceny disku doprovázený dostatečnou rychlostí čtení a zápisu dat. Hlavní nevýhodou je mechanické řešení, které má vysokou spotřebu elektrické energie, je náchylné na poškození při nešetrném zacházení (otřesy nebo náraz při zápisu/čtení dat; v offline stavu je odolný do 350G) a vyšší hmotnost.

Pro zvýšení spolehlivosti a snížení rizika poškození pevného disku a tím následné ztrátě dat se pevné disky spolu kombinují a zapojují se do tak zvaných diskových polí neboli RAID polí. Tím je možné zvýšit bezpečnost dat, nebo výkon diskového subsystému.

3.3.4 VTL – Virtual Tape Library

Jedná se o zařízení pro zálohování diskové povahy, které se vůči okolí prezentuje jako pásková knihovna s robotikou, médii, sloty a může nahradit skutečnou páskovou knihovnu, aniž by to pro stávající implementaci zálohovacího SW znamenalo rozdíl. U VTL se jako médium využívá namísto pásek zpravidla diskové pole. Tato technologie je mimořádně vhodná pro řešení náročných zálohovacích úloh jako velké množství malých souborů. Použití VTL znamená efektivnější zpracování zálohovacího procesu, spuštění

více zálohovacích úloh zároveň, vyšší spolehlivost záloh, rychlejší provedení záloh i obnovy dat atd. VTL se v rámci zálohovací strategie často kombinuje s klasickou páskovou knihovnou, která v takové topologii slouží obvykle k archivaci.



Obrázek 5 Schéma zapojení VTL

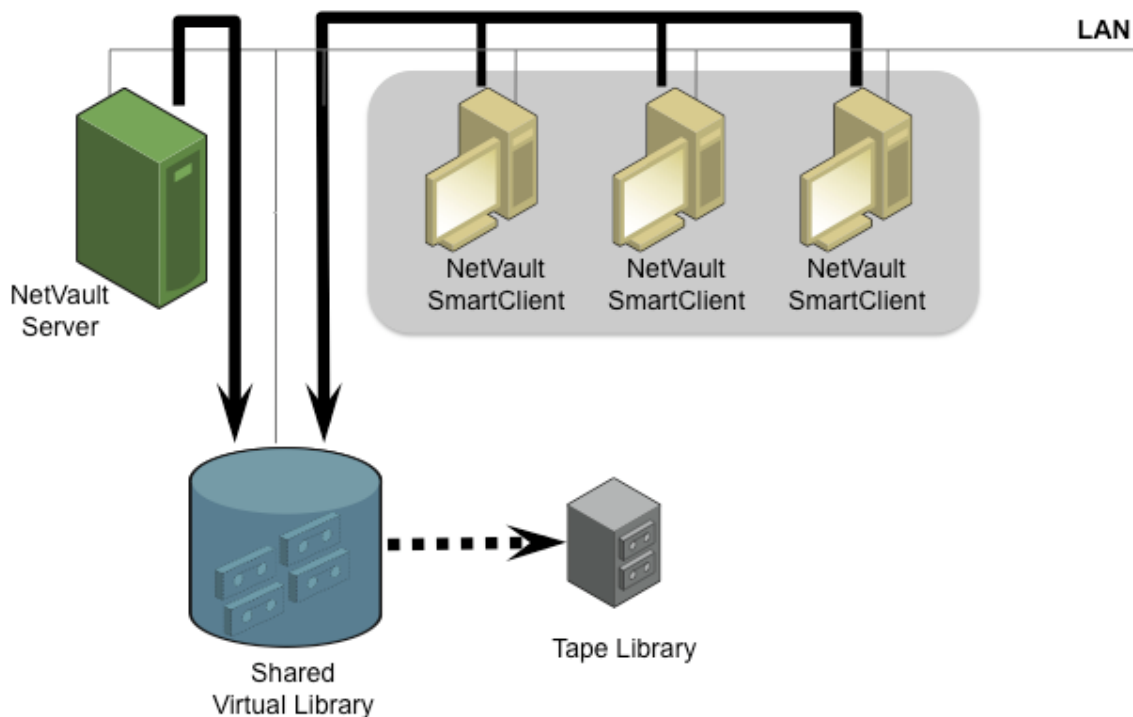
VTL využívá v zásadě technologii, která mívá menší výkon, než primární datová úložiště. Na výkon disků VTL zpravidla nebývají kladeny tak vysoké nároky, jako v případě primárních úložišť. Jedním z pozitivních efektů tohoto principu je nižší cena diskové kapacity.

Spolu s VTL je často využívána technologie **deduplikace** dat, která dále snižuje nároky na výkon disků a kapacitu VTL úložiště.

V současné době jsou disková pole tvořena SATA nebo PATA disky. K vytváření záloh je možné využít softwarové nástroje obvyklé pro zálohování na fyzické pásky.

3.3.5 sVTL

sVTL (shared VTL) je v podstatě stejné zařízení jako VTL, ale navíc umožňuje sdílený přístup více klientů. Zálohovací servery nebo klienti mohou současně v mnoha úlohách zálohovat nebo obnovovat data ze sdílené virtuální knihovny.



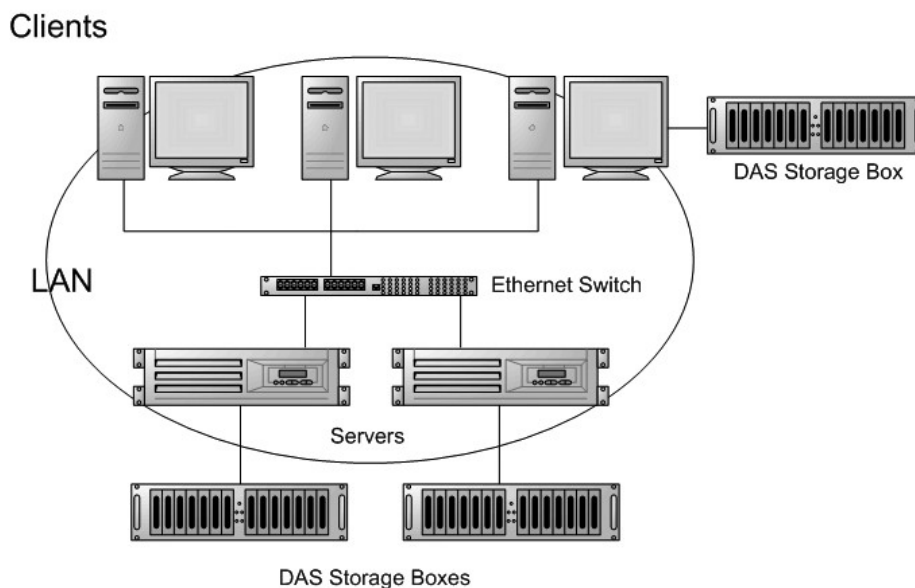
Obrázek 6 Schéma zapojení sVTL

3.3.6 DAS

Direct Attached Storage (DAS) je externí datové úložiště, které je přímo připojeno k pracovní stanici nebo serveru. DAS řeší celou škálu manažerských problémů pro rostoucí firmy. Jednoduchým způsobem řeší potřebu přidání druhého pevného disku k počítači, využití externího disku nebo použití rozšiřujícího zařízení, které obsahuje několik pevných disků, připojených k serveru. Jedná se o jednoduché a cenově dostupné zařízení pro ukládání dat ideální pro lokální sdílení souborů v prostředí s jedním nebo více servery. DAS umožňuje rychle přidat úložnou kapacitu bez plánování nákladů.

DAS se k počítači připojuje přes rozhraní ATA, SATA, SCSI nebo Fiber Channel, což znamená, že nepracuje v síti jako síťové zařízení, ale jako součást počítače, ke kterému je připojeno. Obecně tedy rozšiřuje kapacitu serveru nebo pracovní stanice při zachování vysoké datové propustnosti a nízké přístupové doby. [15]

Direct Attached Storage

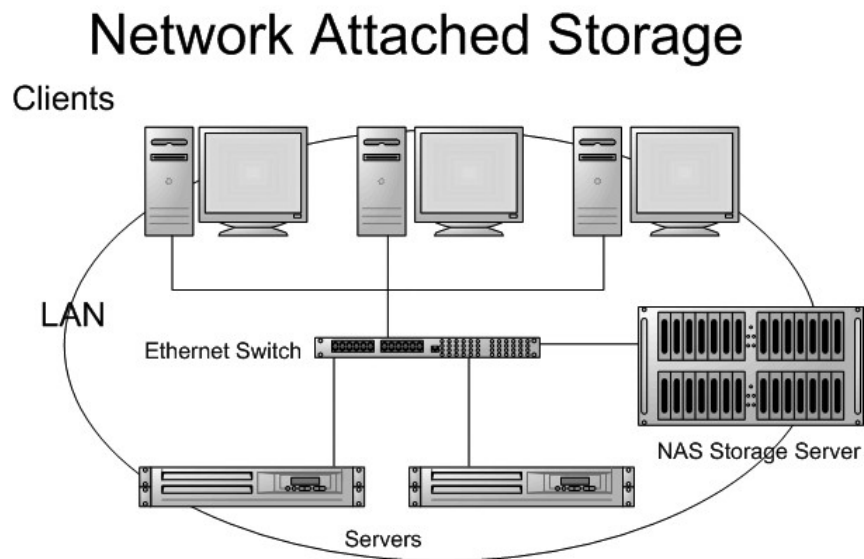


Obrázek 7 Schéma připojení DAS k počítačům [15]

3.3.7 NAS

Network Attached Storage (NAS) je datové úložiště připojené k místní síti LAN. V porovnání mezi jinými síťovými úložišti je NAS rychlejší, má lehčí administraci a snadnější nastavení. NAS obsahuje jeden a více pevných disků, které se můžou slučovat do větších datových struktur nebo můžou vytvořit RAID pole.

NAS je ideální úložiště pro podniky, které potřebují sdílení dat mezi více uživateli. Pomáhá zvýšit produktivitu a chránit informace centralizací dat na jediné místo, kam k nim mohou všichni uživatelé přistupovat a upravovat je. Zároveň je tím velmi usnadněno zálohování sdílených dat. NAS nabízí úroveň sdílení dat přes síť a je ideálním pro společnost, která hledá jednoduchý a nákladově efektivní způsob, jak umožnit přístup k datům a jejich sdílení pro více klientů v síti. NAS systémy lze integrovat do každého prostředí a umožňuje sdílení informací napříč všemi provozovanými platformami. S přidanými funkcemi, jako je RAID a za provozu vyměnitelné disky, nabízí NAS solidní ochranu dat a jednoduchost správy. [15]



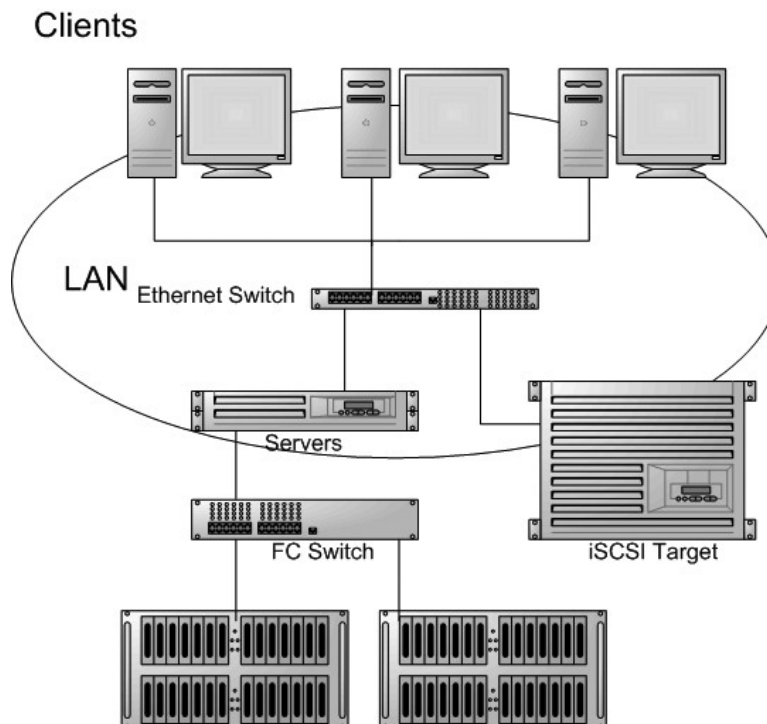
Obrázek 8 Schéma připojení NAS k síti [15]

3.3.8 SAN

Storage area network (SAN) je dedikovaná datová síť, oddělená od sítí LAN, WAN apod., která slouží pro připojení externích zařízení k serverům, jako jsou disková pole, páskové knihovny a jiná zálohovací zařízení. SAN vznikla hlavně kvůli narůstajícím potřebám na zabezpečení a konsolidaci dat. U SAN systémů se využívá blokový přenos dat. Bloky se přenášejí přímo mezi servery a úložnými prostory. Na rozdíl od TCP/IP je SAN od samého začátku navržena na rychlý přenos obrovských množství dat na velké vzdálenosti.

Díky poměrně vysokým pořizovacím nákladům se SAN budují hlavně ve větších společnostech, které vyžadují vysokou dostupnost svých služeb a co největší škálovatelnost rozšiřitelnost. V poslední době se však díky klesající ceně začínají malé SAN rozšiřovat i do stále menších společností. SAN zařízení tak postupně nahrazují DAS technologii. [15]

Storage Area Network



Obrázek 9 Schéma zapojení SAN sítě [15]

3.4 Nejpoužívanější zálohovací metody

3.4.1 D2T (Disk-To-Tape)

Jde o nejstarší a nejjednodušší způsob zálohování v serverových systémech na magnetické pásky. Dříve se jednalo o standard. V dnešní době však již vykazuje řadu omezení. To základní vyplývá ze skutečnosti, že zálohovací SW migruje data ze zdroje přímo na páskovou mechaniku či páskovou knihovnu. Pásková mechanika se však neumí zcela lineárně přizpůsobit kolísajícímu datovému toku, navíc potřebuje jeho určitou minimální úroveň k tomu, aby korektně fungovala. V případě nedostatečného přísunu dat nastává efekt, kdy dochází k procesu opakovaného vyprazdňování cache a zastavení mechaniky, čekání na data, opětovného startu mechaniky a tak stále dokola, což velmi negativně ovlivňuje životnost mechanik.

Také existuje řada situací, například zálohování velkého množství malých souborů nebo záloha dat ze vzdálených poboček přes WAN, při kterých zdroj, infrastruktura,

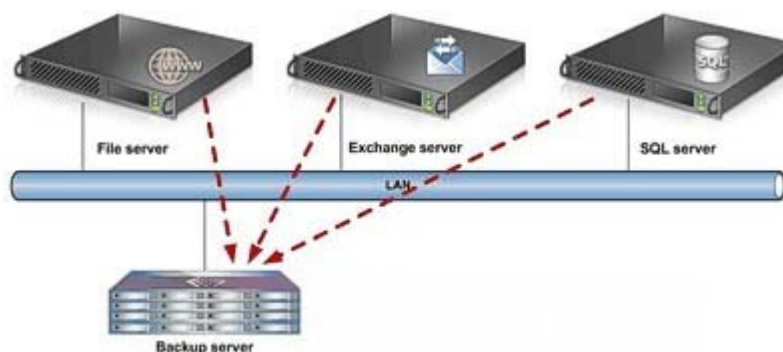
server, zálohovací SW apod. nedovede pracovat dostatečnou rychlostí tak, aby se plně využily možnosti současných zálohovacích mechanik. [16]



Obrázek 10 Schéma zálohování typu D2T [16]

3.4.2 D2D (Disk-To-Disk)

Zálohování na disk a obnova dat z disku je relativně nový způsob ochrany dat. Jde o jednostupňové zálohovací schéma, kdy záloha probíhá do diskového pole. Řeší řadu nevýhod konvenčního způsobu zálohování, podstatně zkracuje zálohovací okno, využívá v plné šíři rychlost zálohovacího systému. A protože zálohovaná data jsou uložena na stále připojených discích, jsou neustále k dispozici. Díky tomu lze dosáhnout velice krátkých časů obnovy. Při zálohování na disk navíc není nutné spravovat média a ukládat je externě. Tato jednoduchost je atraktivní pro manažery IT na všech úrovních. [16]

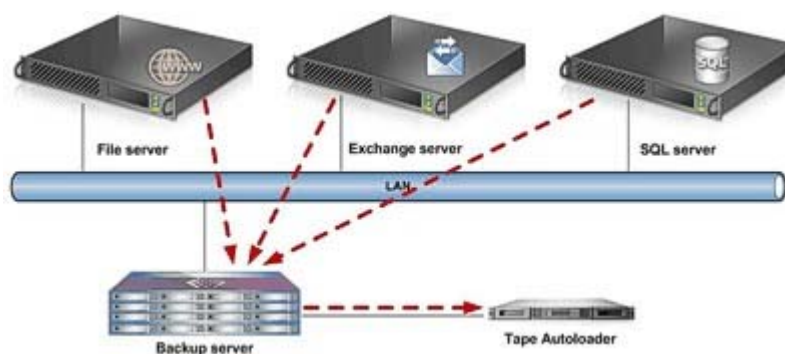


Obrázek 11 Schéma zálohování typu D2D [16]

3.4.3 D2D2T (Disk-To-Disk-To-Tape)

Nejrozšířenější zálohovací strategií je D2D2T. V různých prostředích může být realizována několika způsoby, ale jedno je vždy stejné – zálohování se provádí dvěma etapami z disku na disk a následně z disku na pásku. První záloha je směřována na disky

a využívá hlavní výhody těchto médií – schopnosti přizpůsobit rychlost zápisu objemu právě přenášených dat. Zálohovací proces zálohuje stejně obří soubory i logy databází a malé soubory filesystemů, tedy rychlost a objem zaslaných dat se různě mění. Nejhorším případem je častý zápis malého množství dat. Dříve se běžně stávalo, že roztočená páska marně čekala na data a pak zpomalovala. Nakonec i to zpomalení nebylo dost pomalé a při obnovení datového toku páska musela najet na poslední místo zápisu couváním. Využitím technologických vlastností diskového prostoru tak lze většinou zmenšit zálohovací okno a dát více výkonu aplikacím. Nebo naopak provádět zálohy častěji a tím uspořit mnoho práce uživatelům při případné obnově dat. Pokud by byla několikahodinová ztráta dat kritická, pomůže disková záloha efektivně využít vlastnosti inkrementálního či rozdílového backupu. [17]



Obrázek 12 Schéma zálohování typu D2D2T [16]

3.5 Rotace zálohování

Jedním z klíčových prvků každého zálohování dat je určení rotačního cyklu. Rotace médií je nutná kvůli rovnoměrnému opotřebení a zároveň je vhodné rozdělit zátěž mezi několik médií. Nejlepší rotační schéma médií je takové, které zaručí co možná nejdelší, nejrozsáhlejší a nejrozdílnější kopie podnikových dat.

3.5.1 Round Robin (schéma s jednou páskou pro každý den)

Nejjednodušší schéma rotace pásek získáme, když vyhradíme pro každý den pracovního týdne jednu pásku. Pásky jsou označeny (pondělí, úterý, středa, čtvrtek, pátek). Každý den je provedena na danou pásku plná záloha všech dat vyhrazených pro zálohování. Tato rotace umožňuje obnovu dat s maximálním posunem zpět jeden týden. Schéma je vhodné pro použití v menších společnostech s využitím interní nebo

externí páskové mechaniky anebo s využitím zařízení NAS s vytvořenou VTL, která může sloužit jako primární úložiště dat. Toto řešení je vhodné také tam kde je možné provádět každý den plnou zálohu a časový posun o 1 týden zpět je dostačující. [19]

3.5.2 Grandfather-Father-Son (GFS)

Mezi nejběžněji používané schémata zálohování patří metoda GFS. Toto schéma využívá denní (Son), týdenní (Father) a měsíční (Grandfather) media sety. Čtyři media sety jsou označeny pro každodenní zálohu v pracovním týdnu (obvykle pondělí až čtvrtek). Na tyto media sety (označeny ve schématu GFS jako Son) pak probíhají inkrementální zálohy. Tyto media sety (Son) jsou opětovně přepisovány v dalším týdnu. Další skupinou čítající pět media setů, která je obsažena v zálohovacím schématu GFS jsou media sety označeny jako Week 1, Week 2, atd. (Father). Na tyto media sety (Father) probíhají každý týden plné zálohy, media sety ze setu Son nejsou použity a pro skupinu Father je nastavena doba expirace jeden měsíc. Poté následuje jejich opětovný přepis. Finální media set Grandfather se skládá ze 3 media setů (media set může být tvořen jak jednou páskou tak i několika páskami) a je označen Month 1, Month 2, Month 3, atd. Na tyto media sety pak probíhá opětovný zápis jednou za tři měsíce a více měsíců (záleží kolik media setů je vyčleněno pro skupinu Grandfather. U těchto media setů je nastavena expirace (tj. možnost dalšího přepisu) dle počtu media setů ve skupině Grandfather. Každý media set ze skupiny pásek (Son, Father nebo Grandfather) je buď samostatná páska, nebo skupina pásek, což je závislé na velikosti zálohovaných dat. Celkový počet použitých media setů v zálohovacím schématu GFS je dvanáct. Z důvodů opotřebení pásek a z důvodů zachování delší historie (archivace) záloh je doporučeno media sety v daném časovém intervalu měnit za nové. [19]

3.5.3 Tower of Hanoi (Hanojská věž)

Schéma Hanojské věže vychází z logické hry, které má svůj původ v Číně. Hra je postavena na přesunutí pěti kotoučů z jednoho kolíku na některý další, a to s minimálním počtem tahů. Přitom hráč nikdy nesmí mít v ruce více než jeden kotouč a nikdy nesmí dát větší kotouč na menší. Je dokázáno, že nejmenší počet tahů je 31.

Metoda Hanojské věže využívá pro zálohování pět media setů. Media set A je použit každý druhý den, media set B je použit každý čtvrtý den, media set C je použit

každý osmý den, media sety D a E jsou použity střídavě každý šestnáctý den. Sety, které jsou používány častěji, pak obsahují novější zálohy a ty, které jsou používány méně, naopak starší. Nevýhodou je, složitost manuální obsluhy. Tuto variantu je vhodnější využívat za podpory zálohovacího softwaru, kde je umožněno plánování (scheduling) celého procesu, a automatické knihovny médií. [19]

3.6 Zabezpečení před výpadkem služeb datových center

3.6.1 Lokalita datového centra

Datové centrum by mělo poskytovat datům ochranu proti zásadním rizikům. Za zásadní rizika jsou považovány přírodní katastrofy (seismická aktivita, bouře, hurikány, tornáda, sesuvy půdy, záplavy, požáry), znečištění ovzduší, elektromagnetické rušení, otřesy, politické klima, dopravní nehody především letecké. Tyto rizika znamenají hrozbu vyřazení činnosti celého datového centra. Zavedením redundance na úrovni datového centra se lze bránit rizikům uvedeným výše. Základní požadavky na redundantní datová centra:

- **Riziková událost nesmí vyřadit z běhu obě datová centra současně.**
- Datová centra musí být umístěna v infrastruktuře, která umožňuje dostatečné kvalitní vzájemné datové spojení.
- Vzájemná dostupnost mezi datovými centry musí být v rozumném časovém horizontu, nejlépe více dopravními prostředky. Toto pravidlo umožňuje rychlý přesun odborného personálu v případě ohrožení. [18]

3.6.2 Redundance

Základním předpokladem úspěšné ochrany kritických dat, služeb a aplikací před výpadkem je zavedení redundance. Pokud uvažujeme, že se poskytované služby (například sdílení dat) skládají z komponent, pak redundance je duplikace kritických komponent systému za účelem zvýšení dostupnosti. Pokud duplikovaný komponent selže, nahradí jej jeho kopie. Existují tři základní typy redundance:

Hardwarová – duplikuje zařízení. V systému jsou duplikovány kritické hardwarové komponenty.

Softwarová – duplikuje aplikace. Paralelní běh aplikací, které mají stejný účel, ale jsou různě implementované.

Informační – duplikuje informace. K datům je připojena redundantní informace, která umožňuje obnovu původních dat.

3.6.3 Failover

Failover je automatické přepnutí na redundantní nebo pohotovostní server, systém nebo síť při poruše nebo abnormálním ukončení dříve aktivního serveru, systému nebo sítě. Technologie failover je obvykle nasazována na serverech, systémech nebo sítích, které vyžadují nepřetržitou dostupnost a vysoký stupeň spolehlivosti.

3.6.4 Cluster

Redundance umožňuje budovat služby, které splňují kritéria vysoké dostupnosti. Vysoce dostupná služba (High Availability Service), poskytuje klientům s pevně garantovanou dobou odstavení. Dostupnost služeb se udává procentuálně, je definována jako podíl času dostupnosti za pevně stanovený časový úsek. Nejčastěji se dostupnost vztahuje k roku či měsíci. Tabulka 10 obsahuje základní třídy vysoké dostupnosti vztahené na různé časové úseky. Vysoká dostupnost s číslem obsahující pouze číslice devět se přezdívá N devítek (např. pro 99,99 % = 4 devítky).

Tabulka 10 Třídy vysoké dostupnosti [22]

Dostupnost	Výpadek za rok	Výpadek za měsíc	Výpadek za týden
90%	36,5 dne	72 hodin	16,8 hodin
95%	18,25 dne	36 hod	8,4 hod
99%	3,65 dne	7,2 hod	1,68 hod
99,9%	8,76 hod	43,2 minut	10,1 min
99,99%	52,6 min	4,32 minut	1,01 min
99,999%	5,26 min	25,9 s	6,05 s
99,9999%	31,5 s	2,59 s	0,605 s

Řešení umožňující budování vysoce dostupných služeb za využití redundance se nazývají HA clustery (High availability clusters). HA cluster je skupina výpočetních uzlů, která se podílí na běhu služeb. Cluster se chová pro vnějšího klienta jako jeden

výpočetní uzel. Clustery se dělí dle způsobu implementace výše uvedených operací a rozlohy, nad kterou pracují do následujících typů:

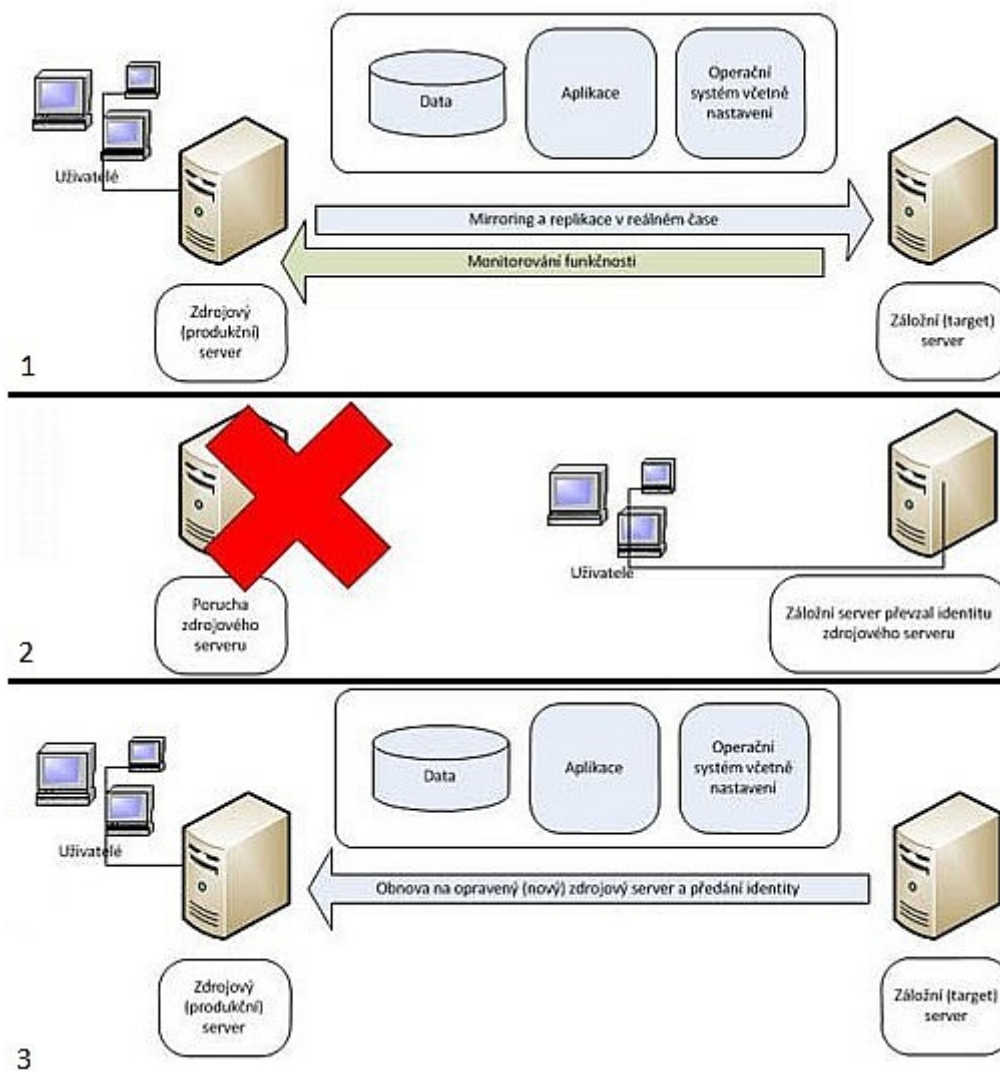
Lokální – Uzly clusteru jsou umístěny v jednom datovém centru, rozprostírají se v rámci jedné budovy, řádově desítky až stovky metrů.

Kampus – Uzly jsou rozprostřeny v rámci několika budov v blízkém okolí, řádově stovky metrů. Cluster je rozprostřen v rámci dvou či více datových center. Datová centra jsou vzájemně propojena dostatečně kvalitními linkami, které umožňují replikaci dat a kontrolu běhu.

Metro – Základní vlastnosti shodné s kampus clusterem, uzly jsou rozprostřeny ve vzdálenosti řádově desítek km.

Regionální – Základní vlastnosti shodné s metro clusterem, uzly jsou ale rozprostřeny ve vzdálenosti řádově stovek km.

Kontinentální – Předchozí typy pracovaly v rámci jedné logické jednotky, clusteru. Kontinentální cluster je tvořen několika nezávislými clustery, jež jsou umístěny v různých datových centrech. Často se jedná o kooperaci clusterů lokálního typu. Clustery jsou od sebe vzdáleny řádově tisíce km.



Obrázek 13 Schéma fungování clusteru [23]

3.6.5 Geocluster

Kampus, metro, regionální a kontinentální clustery jsou označovány jako geocluster či disaster tolerant cluster. Se zvyšující se vzdáleností uzlů se zvyšuje míra ochrany aplikace před výpadkem. Možné hrozby jsou uvedeny v kapitole Lokalita datového centra.

Geocluster se skládá ze dvou datových center, DC1 a DC2. DC1 hostuje data A, DC2 hostuje data B. Datová centra jsou propojena širokopásmovou sítí za účelem replikace a kontroly běhu. V použité technologii na propojení datových center se jednotlivé typy geoclusteru liší. Pokud kontrola zjistí nedostupnost služeb v datovém centru, provede failover služeb do druhého datového centra. Tento krok ukazuje druhá část

obrázku 13. V DC1 (na obrázku reprezentováno jako zdrojový server) došlo k problémům, které znemožní spojení klientů a serveru. Následuje detekce a failover služeb do DC2 (na obrázku reprezentováno jako záložní server) a uživatelé, většinou aniž by změnu postřehly, využívají služeb na záložní straně. Po obnovení funkcionality DC1 jsou data, služby a aplikace replikovány zpět. Jakmile dojde k dokončení replikací, jsou i uživatelé přepojeni na původní DC1.

3.7 Plán obnovy

Zálohou na discích nebo páskách to nekončí. Je potřeba mít Disaster Recovery plán neboli plán obnovy. Disaster Recovery (DR) plán je definice postupů pro zajištění obnovy IT služeb po požárech, živelných pohromách či jiných zásadních událostech. Existence těchto dopředu promyšlených postupů nejen minimalizuje dopady nepříjemných událostí na obchodní výsledky firmy, ale i optimalizuje využití fungujících zdrojů firmy a vynakládání financí na zajištění opětovného fungování procesů.

3.7.1 Testování obnovy

DR plán je velmi úzce spjat s organizací IT a technologickou infrastrukturou. Proto je důležitá pravidelná aktualizace DR plánu a testování procesů v reálném prostředí. Problematika testování recovery procesů v provozu s vysokou dostupností je složitá a nevhodně konstruované a provedené testování může samo o sobě fungování ohrozit. Proto je nedílnou součástí DR plánu i metodika testování recovery procesů a školení jednotlivých týmů s cílem zajistit vysokou profesionální připravenost.

Doporučení je testovat rychlost obnovy, vyzkoušet obnovu serveru či aplikace a měřit u tuho čas. Zapisovat si postup kroků, které jsme udělali. Když seznam budeme mít a obnova bude hotova, zkusme obnovu provést znova s tím, že použijeme přesný postup tak, jak jsme jej zaznamenali. Potom získáme jasnější představu o délce obnovy. Nesmíme se nechat se uchlácholit tím, že se ten server nainstaluje za hodinu, a potom není co řešit. To je velmi krátkozraké a přesto je spousta lidí, kteří v těchto představách žijí. Abych jim ale nekřivdil, žil jsem v těchto představách také, až do prvního reálného testu. [20]

3.7.2 Testování čitelnosti záloh

Samozřejmě kromě aktuálního plánu obnovy je třeba mít také jistotu, že je vůbec z čeho obnovu provádět. Je pravda, že v dnešní době již většina moderních zálohovacích zařízení a zařízení pro ukládání dat mají jakousi sebe detekci správné funkcionality a v případě potřeby dopředu upozorní na nutnost výměny disku nebo pásky z důvodu opotřebení či na potřebu vyčištění páskové mechaniky. Ale nemusí a mnohdy to ani není dostačující způsob kontroly čitelnosti dat ze zálohovacích médií. Proto je doporučeno pravidelně jednou za čas alespoň namátkově zkusit obnovit menší objem dat ze zálohy pro kontrolu, že vše funguje jak má, a že jsou naše důležitá zazálohovaná data bezpečně uložena.

II. PRAKTICKÁ ČÁST

4 ANALÝZA PROSTŘEDÍ

Po obdržení zadání téměř jakéhokoliv projektu je velice důležité nejprve provést analýzu stávajícího prostředí před samotným návrhem řešení a následnou realizací daného projektu.

4.1 Zadání

Mým úkolem bylo navrhnout a následně realizovat způsob zálohování a obnovy dat právě budovaného datového centra v naší části organizačního celku. Zadání mělo několik podmínek, které bylo potřeba vzít do úvahy při navrhování řešení.

Seznam zadaných nutných podmínek, které byly kladeny na řešení:

- zrychlení zálohovacího procesu i při větším množství dat než bylo doposud,
- zrychlení obnovy dat (především časté obnovy malých objemů dat),
- zvýšení bezpečnosti zálohovaných dat,
- sjednocení systému zálohování napříč všemi informačními systémy v organizaci,
- nativní podpora zálohování databázových systémů MS SQL, Informix, MySQL,
- nativní podpora zálohování poštovních serverů MS Exchange.

4.2 Stávající stav systémů a jejich zálohování

Před zahájením výstavby datového centra byla naše lokalita provozována na hardwarových technologiích převážně od společnosti Hewlett-Packard (HP). V rámci celé organizace je podobně vybavených lokalit více, ale ne na všech došlo nebo dojde k rozšíření o datové centrum.

Hardwarové parametry stávající lokality (stěžejní prvky):

- *Servery:* HP ProLiant 380 Generace G4 a G5 (12 kusů)
- *Diskové pole:* HP MSA 1500 (2 kusy)
- *SAN Switch:* HP Storage Works 4/8V SAN switch (2 kusy)
- *LAN Switch:* CISCO Catalyst 2960(70) 24 port (3 kusy)
- *Zálohovací knihovna:* HP Storage Works MSL 6030 (1 kus)

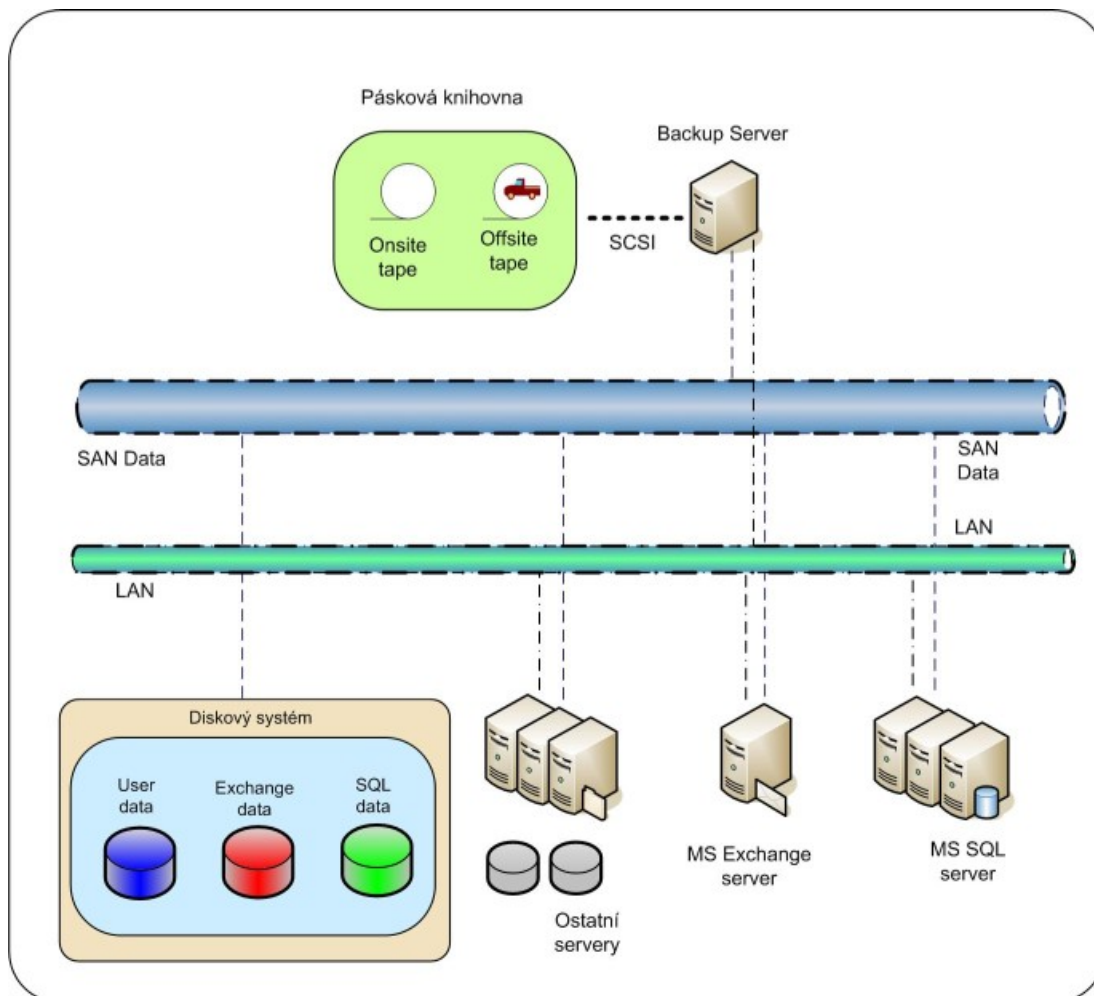
- *UPS*: HP R3000 XR s přídatným ERM (6 kusů)

Celé řešení samozřejmě obsahovalo více hardwarových prvků, jako například KVM switche, PDU jednotky, RACKy a podobně, ale ty pro tento popis nejsou důležité, tak je zde konkrétně neuvádím.

Na názorném obrázku 14 je přehledně schematicky znázorněné zapojení stávajících systémů a jejich propojení na zálohovací systém. Servery, včetně zálohovacího serveru, používají pro běžnou komunikaci mezi sebou a pro komunikaci směrem k uživatelům standardní síť Ethernet LAN. Ta je pro tuto běžnou komunikaci naprosto dostačující. Pro přístup k diskovým polím je však již použito rychlejšího datového komunikačního média v podobě technologie SAN. Tato technologie je použita i pro zálohování dat z diskových polí. Zálohovací server (Backup Server) se prostřednictvím serverového zálohovacího SW spojí s klientskou částí zálohovacího SW nainstalovanou na produkčních serverech (MS Exchange server, MS SQL server a ostatní servery) a zjistí, co je potřeba zazálohovat a kde jsou data uložena na diskovém poli. Tato komunikace probíhá prostřednictvím sítě LAN. Přenos samotných zálohovaných dat pak již probíhá po rychlejší síti SAN do zálohovacího serveru. K němu je připojená robotická zálohovací knihovna prostřednictvím rychlého komunikačního rozhraní SCSI, přes které jsou data posílána do páskových mechanik v robotické knihovně a zaznamenávána na zálohovací pásy.

Tato technologie má velkou nevýhodu v tom, že veškerá zálohovaná data musí procházet přes zálohovací server. To samozřejmě výrazně zatěžuje nejen zálohovací server samotný díky nutnosti zpracování těchto dat, ale také komunikační linky. Z důvodu déle trvajícího přenosu a ukládání zálohovaných dat je zároveň časové okno, čili samotná doba provádění záloh, poměrně veliké. To je samozřejmě nežádoucí jev, protože obvykle je snahou časové okno zkrátit na co možná nejkratší dobu. V případě menšího množství zálohovaných dat to není až tak velký problém a je to plně dostačující a dostupné řešení pro zálohování. V případě citelného nárůstu dat, která je potřeba zazálohovat již toto řešení nestačí a potřeba zvolit jinou technologii zálohování a to je právě náš případ.

Problém není ale pouze v dlouhotrvajícím procesu zálohování, ale možná ještě větší nevýhoda tohoto řešení je v případě nutnosti obnovování dat. Pokud by se jednalo o menší množství dat, o které jsme z nějakého důvodu přišli, obnova by netrvala ani příliš dlouho. Pokud by však ztráta dat byla větší nebo bychom dokonce museli obnovovat úplně celou lokalitu, doba obnovy mohla být až nepříjemně dlouhá.



Obrázek 14 Schéma zapojení zálohování ve stávajícím stavu

Díky rozšíření lokality o datové centrum, některé z původních komponent již dále nebyly zapotřebí, protože byly nahrazeny právě novou instalovanou technologií. Byly to prvky pro ukládání dat, tedy disková pole a prvky pro zálohování dat, tedy zálohovací robotická knihovna. Disková pole byla plně nahrazena a výrazně rozšířena velkým datovým úložištěm v datovém centru, a protože by tak velké množství dat nebylo možné zálohovat stávající zálohovací knihovnou, i ta musela být nahrazena větším rozšiřujícím způsobem.

5 NÁVRH A REALIZACE ZBŮSOBU ZÁLOHOVÁNÍ

5.1 Návrh zálohování datového centra

U takovýchto již poměrně velkých systémů je velmi důležité neudělat žádnou chybu při zapojení a realizaci jak samotného datového centra, tak jeho zálohování. Proto je nanejvýš vhodné ne-li přímo nezbytné před tímto krokem vypracovat podrobný návrh řešení. Tento návrh je dobré několikrát důkladně projít a konzultovat jej s dalšími nejlépe zkušenějšími odborníky na tuto problematiku. Vzhledem k těmto okolnostem a také tomu, že s takovými nákladnými a složitými systémy jsem doposud neměl téměř žádné zkušenosti, jsme se v rámci organizace rozhodli nespoléhat pouze na jednotlivce, ale na kompletním návrhu a řešení spolupracovat s organizací, která se na realizaci a provoz takovýchto systémů specializuje. Toto řešení se z počátku může jevit jako zbytečně nákladné, ale výsledkem vede spíše k úsporám před ještě nákladnějším řešením případných problémů spojených právě s chybně navrženým systémem a jeho realizací. V oblasti informačních technologií již několik let spolupracujeme s firmou MHM computer a.s., se kterou jsme po všech stránkách velmi spokojeni, a proto jsme si ji, také vybrali i pro spolupráci na tomto projektu.

5.1.1 Tvorba návrhu se splněním podmínek

Nyní je tedy na řadě tvorba komplexního návrhu podle zadaných kritérií. Při návrhu konkrétního řešení jsem se zaměřil především na splnění bodů zadání.

- Prvním bodem byla podmínka zrychlení zálohovacího procesu i při větším množství dat než bylo doposud. Vybudováním datového centra se výrazným způsobem rozšířila kapacita pro ukládání pracovních dat, což byl samozřejmě žádoucí jev, ale tím se zároveň zvýšil objem dat, která je potřeba zálohovat. Stávající způsob zálohování byl navrhován na kapacity, bez datového centra, takže novým podmínkám by již nevyhověl jak z kapacitních důvodů, tak i délkou časového okna, které by pravděpodobně zasahovalo i do doby produkce a tím výrazně ovlivňovalo dostupnost systémů. Pro tuto podmínku se jako ideální způsob zálohování jeví zálohovací metoda typu D2D, která je dostatečně rychlá, čímž se zálohovací časové okno zkrátí na minimum a to i při větším objemu zálohovaných dat. Navíc díky rychlému přenosu dat

lze provádět zálohování touto technikou i v době plného provozu bez většího omezení dostupnosti systémů. Návrhem této metody byl tedy první bod podmínek splněn.

- Další podmínkou bylo zrychlení obnovy dat. To se týká především takových činností obnovy dat, která jsou prováděna u většiny systémů nejčastěji a tou je obnova malého množství malých dat o velikosti jednoho adresáře nebo několika málo souborů či snad úplně nejčastěji jediného souboru. Ve stávajícím způsobu zálohování lze takovou operaci obvykle provést do několika desítek minut. Ovšem při několikanásobném nárůstu objemu dat, které vybudováním datového centra vzniklo, by tato operace mohla trvat nesrovnatelně delší dobu, což je nevyhovující. I této podmínce velmi dobře vyhovuje zálohovací metoda D2D z důvodu již zmíněných vlastností této metody. Tento bod byl tedy návrhem této metody také splněn.
- Již delší dobu se v naší organizaci řeší problém bezpečnosti zálohovaných dat. Stávající metoda využívající pro zálohování dat magnetické pásky je sice vysoce spolehlivá, ale pouze pro určité typy nežádoucích jevů vedoucích ke ztrátě dat. Proto byl mezi podmínky zařazen bod zvýšení bezpečnosti zálohovaných dat. Důležitou součástí bezpečnosti je totiž také bezpečnost fyzická, která zároveň pomáhá řešit zabezpečení zálohovaných dat před hrozbami typu živelných katastrof, požárů a podobně. Pro splnění této podmínky jsem navrhl rozšíření dříve navrhované zálohovací metody D2D na více vrstvou a ještě bezpečnější metodu D2D2T. Při prvním kroku provádění zálohování dat (D2D) jsou data přenášena v rámci jedné budovy a jedné místnosti. Z toho vyplývá, že například v případě požáru můžeme přijít jak o produkční data, tak i o jejich zálohy. Z tohoto důvodu je metoda D2D nevyhovující. Proto byla tato metoda rozšířena o další krok a tím je D2T. Samozřejmě, že pro zvýšení zabezpečení zálohovaných dat včetně toho fyzického je vhodné, aby druhá fáze zálohování probíhala na dislokované úložiště. Proto jsem tedy navrhl umístit zálohovací knihovnu do jiné budovy a propojit budovy rychlým datovým spojem. Tím je dosaženo nejen vyšší bezpečnosti po stránce fyzické, ale zároveň také máme hned dvě kopie zálohovaných dat, což zvyšuje bezpečnost záloh i pro případ poruchy hardware u prvního zálohovacího stupně D2D. Tato podmínka byla tedy také splněna.
- Protože nové datové centrum bude využíváno různými informačními systémy postavených na různých platformách, bylo další podmínkou umožnění použití jednotného systému zálohování napříč všemi informačními systémy v organizaci.

S tímto požadavkem úzce souvisí i další dvě podmínky a těmi jsou podmínky nativní podpory provozovaných softwarových technologií a systémů jak databázových, tak prostředků elektronické komunikace. V této fázi se v podstatě jednalo především o to najít vhodný zálohovací software, který tyto podmínky bude splňovat. Výběru zálohovacího softwaru se věnuji podrobněji v jedné z dalších kapitol.

5.1.2 Datové centrum – hardware

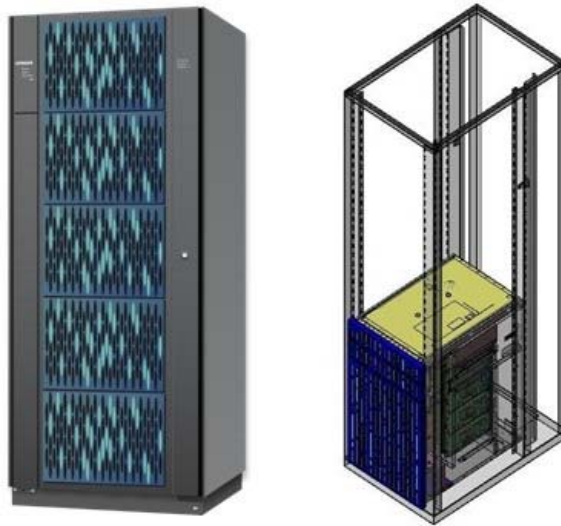
Základním stavebním kamenem dobrého a spolehlivého systému je výběr kvalitního a osvědčeného hardware. Při výběru zařízení pro sestavování našeho nového datového centra jsme dali na rady zkušenějších a zvolili technologie, které firma MHM computer a.s. již řadu let používá, které se jim osvědčily a s jejichž provozováním mají dlouholeté zkušenosti.

Použité technologie při výstavbě datového centra:

❖ Diskové pole Hitachi USP VM

- USP-VM je jediný virtual storage controller s možností přidání internal disk storage
- Umožňuje nondisruptive (bez nutnosti přerušení provozu) migraci dat mezi heterogenními storage systémy
- Maximální cache: 128 GB
- Počet host portů: max. 48 4/8 Gbps
- Interní disky: 0 – 240
 - Typ: Flash 73, 146 a 200GB
 - Typ: Fibre Channel 146, 300, 400, 450 a 600GB
 - Typ: SATA II 1 a 2TB
- Maximální raw kapacita
 - Fibre Channel: 138 TB (600GB jeden disk)
 - SATA: 472 TB (2TB jeden disk)
 - Interní a externí: 96 PB

- Agregované IOPS: 1,2 milion
- Agregovaná šířka pásma: 13,3 GB/s
- Open systémy
 - HP-UX, Tru64 UNIX, Open VMS; AIX®
 - Microsoft®Windows Server: 2000, 2003, 2008
 - Novell: NetWare
 - Linux: SUSE: Linux Enterprise Server; Red Hat: Enterprise Linux AS
 - SGI: IRIX; Sun Microsystems: Solaris
 - VMware: ESX Server



Obrázek 15 Hitachi USP-VM [27]

❖ Diskové pole Hitachi AMS 2300

- Symetrický active/active design
- Počet kontrolerů 2
- Hitachi Dynamic Load Balancing (DLB) kontroler
- Cache řadiče (na systém) 4GB – 8GB
- Fibre Channel: 8Gb/s
- iSCSI: GigE
- Maximální počet disků 240

- Maximální Raw kapacita 472TB SATA, 138TB SAS
- Interní disky
 - 300GB SAS (15K RPM)
 - 450GB SAS (15K RPM)
 - 600GB SAS (15K RPM)
 - 1TB SATA II (7200 RPM)
 - 2TB SATA II (7200 RPM)
- Podpora RAID polí
 - RAID-0, RAID-1, RAID-1+0,
 - RAID-5, RAID-6
- High Availability (prvky vysoké dostupnosti)
 - Nondisruptive (bez nutnosti přerušení provozu) výměna komponent
 - Nondisruptive výměna hot-pluggable disků
 - Nondisruptive update mikrokódu
- Podpora OS: MS Windows, VMware, RHEL, SUSE, Oracle Enterprise Linux, Sun Solaris, Novell Netware, HP True64 ,HP-UX, IBM AIX, Apple MAC OS, IBM z/OS,



Obrázek 16 Hitachi AMS 2300 [28]

❖ Pásková knihovna actiLib LTO Series 4U

- Počet slotů pro pásky: 48
- Maximální počet mechanik: 4x LTO4, LTO5
- Maximální kapacita:
 - při použití LTO4 - 38,4 TB (bez komprese) / 76,8 TB (s kompresí)
 - při použití LTO5 – 72 TB (bez komprese) / 144 TB (s kompresí)



Obrázek 17 actiLib Library LTO Series 4U [29]

5.1.3 Výběr zálohovacího software

Pro bezproblémovou funkčnost a kvalitní zabezpečení dat a jejich záloh je také, kromě kvalitního hardware, potřeba zvolit neméně kvalitní zálohovací software. Ten navíc v našem případě musel splňovat řadu podmínek, které nejsou až tak obvyklé při pořizování podobného softwarového vybavení. To bylo dáno tím, že jsme nebudovali datové centrum jako výchozí technologii, ale tímto systémem jsme vlastně rozšiřovali stávající technologie, které již byly nasazené a provozované a které bylo potřeba zachovat. Zároveň však bylo potřeba sjednotit způsob zálohování a centralizovat správu a umístění zálohovaných dat. Ideálním řešením se jevílo plně přenést zálohování na nové datové centrum. Tato centralizace zálohování má velké výhody, ale měla také několik úskalí. Obrovskou devízou tohoto řešení je maximální zjednodušení správy, zpřehlednění a zrychlení zálohování. S tím samozřejmě souvisí i zjednodušení a velmi výrazné zrychlení procesu obnovy dat v případě její potřeby.

Zálohovacích softwarů, pro rozsáhlá řešení, jaké je to naše, je na trhu několik a většinou jsou ve své podstatě velmi podobné. Základní vlastnosti mají téměř totožné, liší se vlastně pouze přidávanými vylepšeními a podporou zálohování některých speciálních typů

programů. Dále jsou samozřejmě rozdíly cenové, které v dnešní době hrají také výraznou roli při výběru, a v neposlední řadě je to podpora produktu výrobcem zálohovacího SW. Výběr mezi nimi potom spočívá především v tom, aby nasazený produkt co možná nejlépe vyhovoval požadavkům organizace a zároveň byl pro tuto organizaci cenově dostupný. Před tímto krokem se v naší organizaci využívalo hned několik programových vybavení pro zálohování, od těch nejjednodušších, které jsou dnes často již součástí systému, jako například NTBackup, využívaný především pro zálohování stanic a menších samostatných serverů, po složitější, jako například EMC Legato NetWorker nebo Bakbone NetVault, využívané pro zálohování stávajících větších systémů.

Pro realizaci našeho projektu byl nakonec vybrán zálohovací software NetVault od firmy Bakbone. Jednak asi nejlépe pokrývá všechny naše oblasti potřebné k zálohování, potom také, asi jako jediný, podporuje zálohování databázového systému Informix od firmy IBM, který je u nás v rámci jednoho z provozovaných informačních systémů nasazen, a v neposlední řadě s ním již máme praktické zkušenosti, protože už před zřizováním nově nasazovaných technologií jsme jej využívali u systémů stávajících.

Tímto krokem byly splněny všechny podmínky stanovené při zadání projektu, takže návrh je kompletní.

5.2 Realizace návrhu

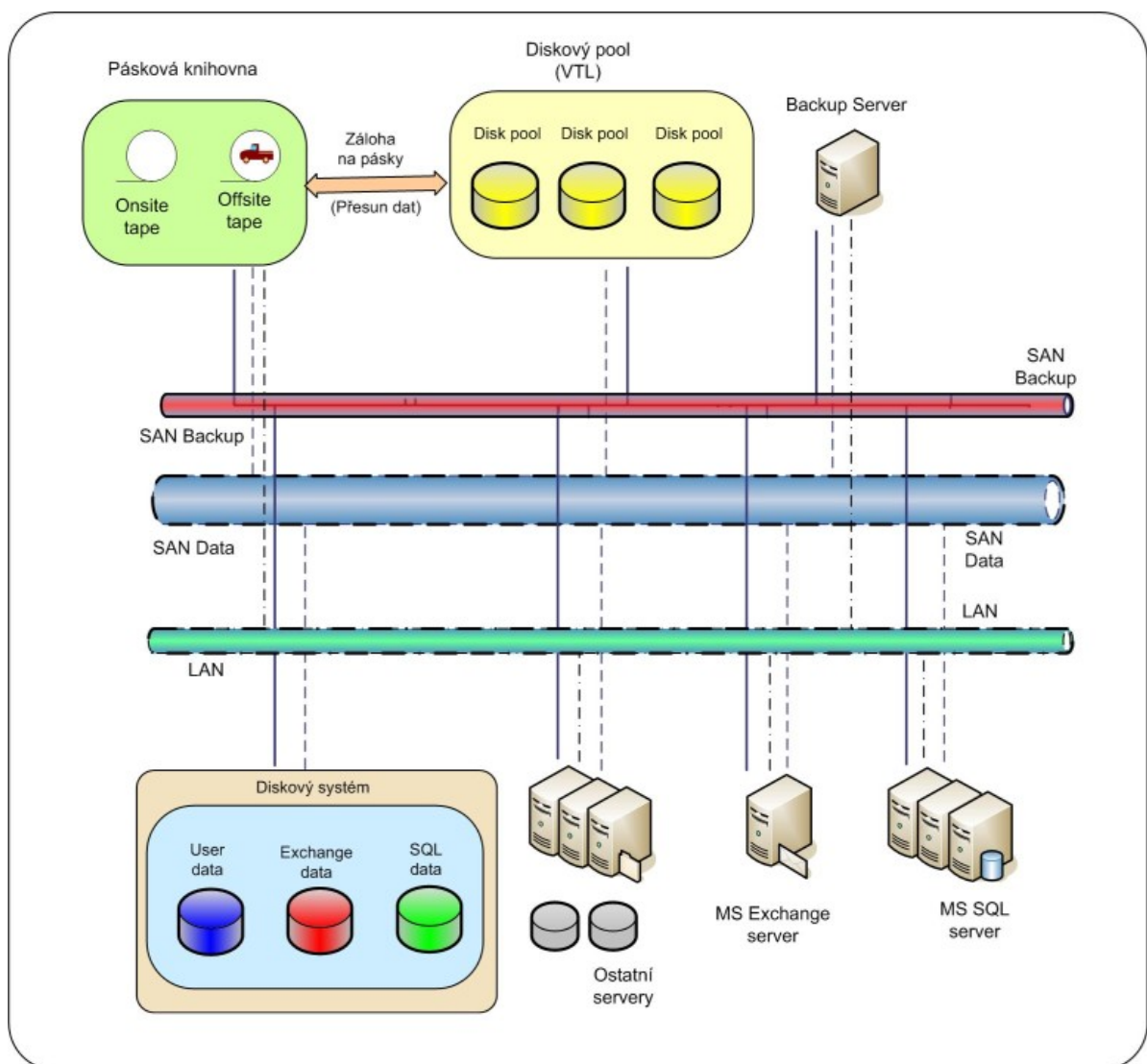
Jakmile tedy máme návrh hotový a odsouhlasený, lze přistoupit k samotné realizaci. Postup, kterým byla provedena samotná realizace, si popíšeme v této kapitole.

Na začátku jsme museli rozhodnout, do jakých prostor datové centrum umístíme, aby byly splněny veškeré požadavky na provoz a zabezpečení. Provedli jsme ověření, zda je v původním objektu s již provozovanými technologiemi dostatečně dimenzované elektrické vedení, jistící prvky, kapacita připojení na datovou síť a v neposlední řadě dostatečná kapacita klimatizační jednotky. Naštěstí původní místnost s technologiemi byla po této stránce, až na pár drobností, dostatečně dimenzovaná. Pouze se nechalo rozšířit elektrické vedení o další okruhy a jejich jištění. S tímto bodem tedy nebyl žádný větší problém a mohlo se začít s výstavbou. Tedy alespoň výstavbou části datového centra s diskovými poli, protože, jak již bylo popsáno v návrhu, ještě bylo potřeba vyřešit otázku umístění dislokovaného stanoviště pro zálohovací robotickou knihovnu. Tento úkol naštěstí také nebyl příliš náročný, protože se využila místnost datového rozvaděče

ve vedlejší budově, která byla kompletně vybavená pro podobné účely. Jediné, co bylo potřeba doplnit, byla konektivita typu SAN pomocí optického kabelu přímo z datového centra pro rychlý přenos zálohovaných dat.

Jako první, po zajištění vyhovujících prostor, byly sestaveny veškeré hardwarové komponenty a byla odzkoušena jejich funkcionalita. Všechny součásti byly v pořádku, tudíž byla na řadě konfigurace a nastavení všech zařízení podle vypracovaného návrhu.

Způsob zapojení všech součástí systému je názorně nakreslen na následujícím obrázku, který následně popíši.



Obrázek 18 Schéma zapojení nového způsobu zálohování

Jak je z obrázku patrné, byly vytvořeny dvě oddělené sítě typu SAN a celá lokalita je pak ve směru k uživatelům připojena k celopodnikové síti Ethernet LAN. Dvě oddělené

sítě SAN pro rychlý přenos dat mají své opodstatnění. Jedna (SAN Data) totiž slouží k přenosu právě využívaných dat mezi diskovým systémem a produkčními servery a druhá (SAN Backup) je vyčleněna čistě pro potřeby zálohování. Výhodou je, že se tyto sítě vzájemně neovlivňují ani při právě probíhajícím zálohování a tudíž se šířka pásma u sítě pro přenos produkčních dat nezmenší a uživatelé tím nejsou nijak ovlivněni. Zálohovaná data jsou přesouvána po SAN pro backup z úložiště do klienta a z něj do zálohovací mechaniky (VTL, disk pool). Této technologii se říká LAN free backup. Síť LAN je využívána z pohledu zálohování pouze pro výměnu META dat mezi backup klientem a backup serverem. Tato META data obsahují informace o probíhajícím či proběhlém zálohování. Mimo jiné nesou informace, odkud a kam se která data zálohovala, zda proběhlo vše v pořádku, čas začátku a konce zálohování atd. Tyto informace se ukládají na zálohovacím serveru a na základě těchto informací potom backup server může vyhodnotit, zda jsou všechna zálohovaná data v pořádku a v případě potřeby obnovy ví, odkud má potřebná data obnovovat. Jednotlivé kroky jsou logovány a logy následně vyhodnocovány. Zároveň byla nastavena automatická upozornění pomocí emailu, pokud se některé operace nezdaří nebo pokud dojde k nějaké jiné chybové události.

Oproti původnímu zapojení se toho příliš nezměnilo. Samozřejmě se přešlo na nové technologie, ale princip zůstal zachován. Vlastně největší změna nastala právě ve způsobu zálohování. Na novém diskovém systému se totiž pro navrženou novou metodu zálohování D2D2T vytvořil speciální diskový oddíl (Disk pool) a nakonfiguroval se jako virtuální pásková robotická knihovna (VTL), který slouží výhradně pro zálohování a na okolí se jeví jako fyzická zálohovací knihovna.

Hlavní části:

- Backup Server
 - Zálohovací software – server
 - OS: Linux
 - NetVault: Backup Server ® (verze 8.10)
 - Copyright © Bakbone Software Inc.
- Klienti
 - Zálohovací software – agent

- OS: Windows, Linux
- NetVault: Backup Client ® (verze 8.10)
- Copyright © Bakbone Software Inc.
- Typy agentů:
 - Filesystem
 - Exchange
 - SQL
 - Informix
- Disková pole
 - USP VM (virtual storage controller s možností přidání internal disk storage)
 - Disky: 33 x FC 300GB
 - AMS 2300
 - Disky: 15 x SAS 600GB (15K RPM)
 - Disky: 60 x SATA 2TB (7200 RPM)
- Pásková jednotka
 - actiLib 4U, 48 slotů, 2 mechaniky LTO4
 - Výrobce: actiData
 - Umístěná v jiné budově

5.3 Zálohovací prostředí v provozu

V okamžiku, kdy máme všechny technologie zapojené a zprovozněné, zkrátka připravené na plný provoz, je na řadě instalace zálohovacího softwaru a uvedení zálohování do provozu. V této kapitole popíšeme způsob instalace a prostředí zálohovacího nástroje.

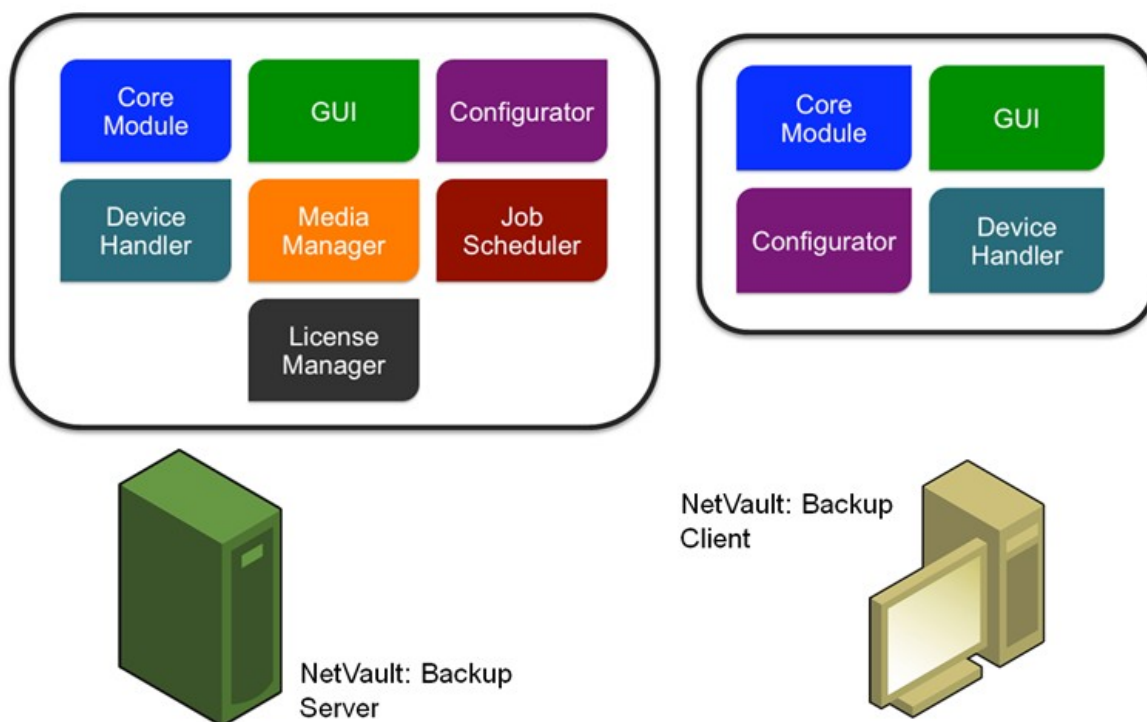
5.3.1 Instalace a GUI

Jak již bylo zmíněno v předcházejících kapitolách, vybrán byl zálohovací software od firmy Bakbone NetVault.

Minimální systémové požadavky na zálohovací server:

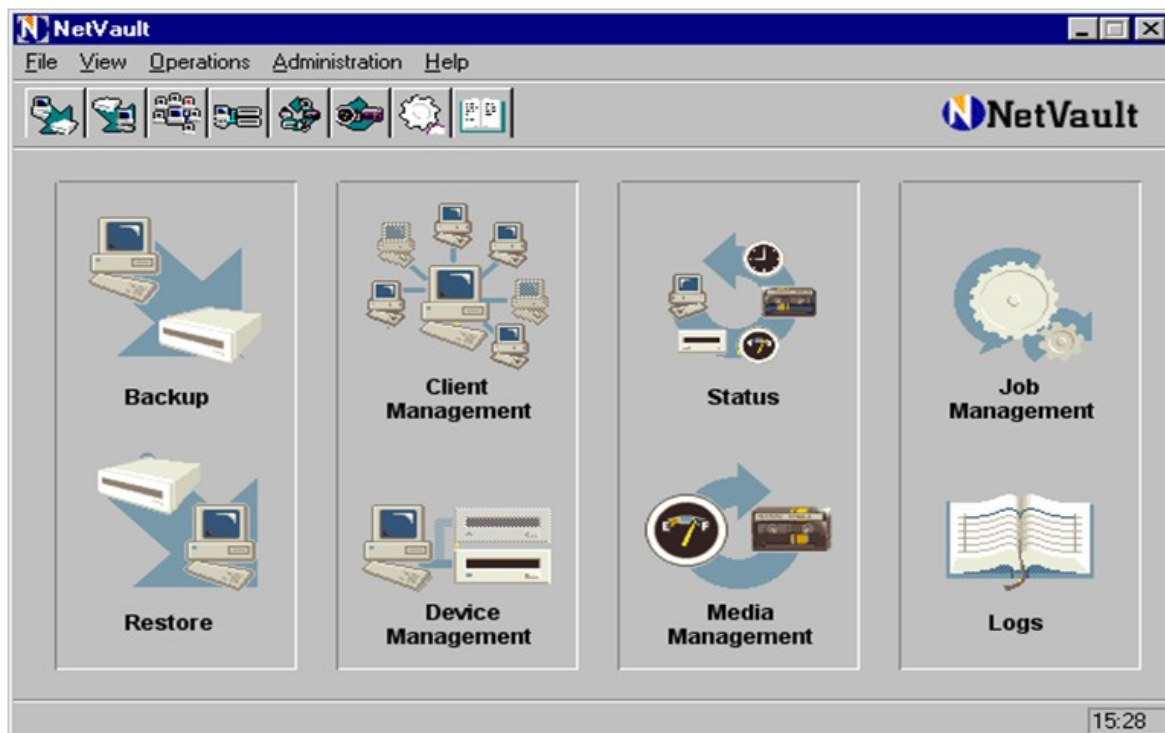
- Podporovaný operační systém: Windows, Linux, Solaris
- Rozlišení: 800 x 600 obrazovka a 256 barev nebo lepší
- Paměť: minimálně 128 MB, doporučovaných je minimálně 1-2GB
- Potřebné místo na disku: to je závislé na několika faktorech. Pro výpočet potřebného místa je využíván vzorec, kde se uvažuje s počtem zálohovaných souborů, dobou jejich držení, počtem zálohovaných serverů a režii backup serveru.

Instalace zálohovacího SW se dělí na část serverovou NetVault: Backup Server a na část klientskou NetVault: Backup Client. Na následujícím obrázku je názorně vidět, které moduly se instalují a kam.



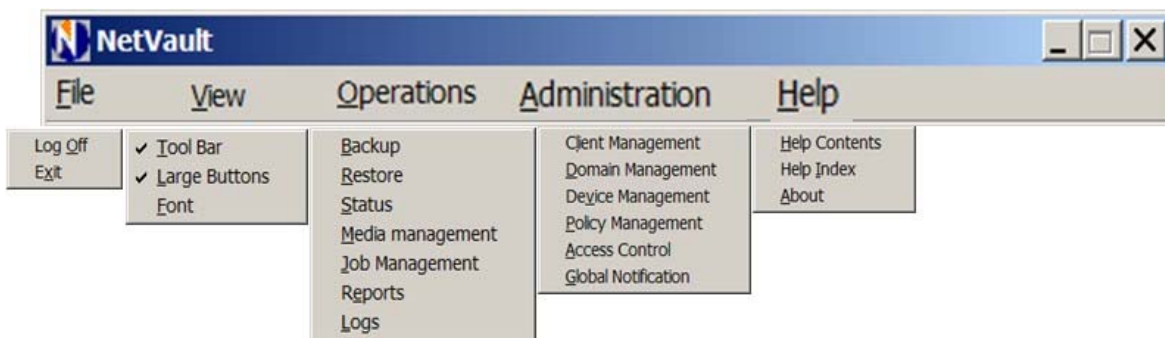
Obrázek 19 Rozdělení instalovaných komponent

Průběh instalace není ničím výjimečný, po otevření spustitelného instalačního souboru se spustí instalátor a ten nainstaluje požadované součásti. Po nainstalování lze spustit grafické uživatelské rozhraní, ze kterého se ovládá veškeré nastavení procesů zálohování.



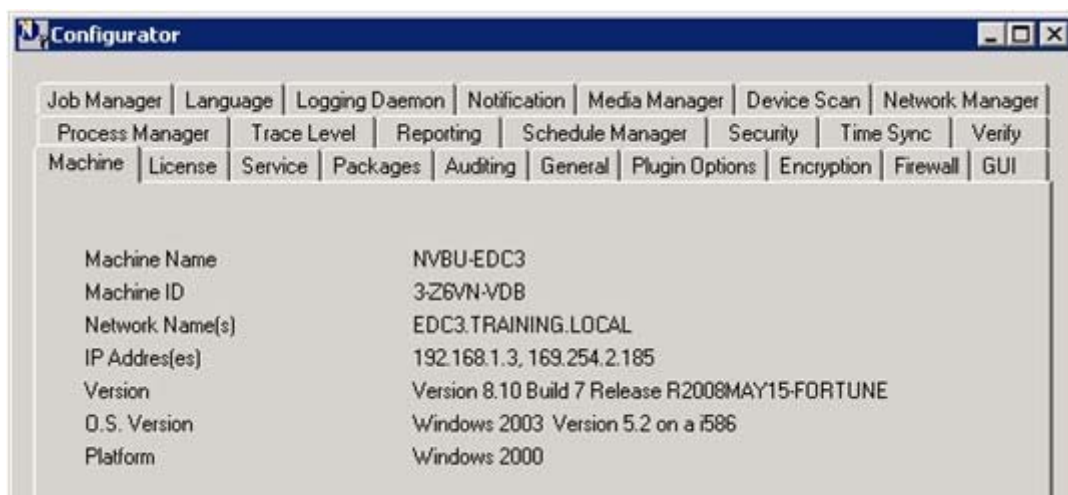
Obrázek 20 Grafické rozhraní konzole NetVault: Backup Serveru

Hlavní menu obsahuje mimo základní funkce grafického menu další rozšíření.



Obrázek 21 NetVault Menu Options

V grafickém okně konfiguratoru je ze seznamu záložek vidět, co vše je možné nastavovat v rámci chování zálohovacího serveru NetVault: Backup. Administrátor odtud může spouštět a zastavovat služby NetVaultu a provádět přizpůsobení nastavení SW NetVault: Backup.



Obrázek 22 NetVault Configurator

5.3.2 Nastavení zálohovacích jednotek

V tomto kroku byly nastaveny zálohovací jednotky pro zálohování. Pomocí softwaru NetVault, konkrétně jeho části pro správu jednotek Device Management, se vytvořila virtuální knihovna, v ní virtuální jednotky a provedlo se jejich napojení na systém. Zároveň se zde provedlo napojení fyzické dislokované zálohovací knihovny.

5.3.3 Nastavení zálohovacích médií

Dále byla potřeba rozdělit podle potřeby zálohovací média a to jak virtuální, tak fyzické pásy pro jednotlivé typy záloh a pro určité typy zařízení. Zde se vycházelo ze zkušeností ze stávajícího systému pro zálohování, jen se přidala dodatečná kapacita pro možnost růstu objemu dat. Toto nastavení se provádí v NetVaultu v položce pro správu médií Media Management.

5.3.4 Správa klientů

Neméně důležitým bodem bylo připojení a nastavení klientů pro zálohování. Tento bod byl proveden ve dvou fázích. V první fázi se na všechny klienty, které je potřeba zálohovat, musela nainstalovat klientská část ovládacího software pro zálohování tzv. agent NetVault: Backup Client. Jakmile byla tato fáze hotová, přistoupilo se k druhé části napojení a nastavení klientů a to na zálohovacím serveru, konkrétně v aplikaci NetVault: Backup Server v části pro konfiguraci a správu klientů Client Management.

5.3.5 Vytvoření zálohy (procesu)

Když už jsme měli připojené klienty a zálohovací zařízení a připravená zálohovací média, mohlo se přistoupit ke kroku, ve kterém se podle následující tabulky vytvořily zálohovací procesy tzv. joby. Tyto joby se nastavují a spravují v části Job Management a vytvářejí se v části pro vytváření záloh Backup.

Tabulka 11 Přehled nastavení zálohovacích procesů

Co se zálohuje	typ zálohy	četnost	kdy	čas
File System serverů (disk C a System State)	FULL	1x/týden	Sobota	3:00
	DIFF	-		
	INCR	-		
File System - sdílené síťové úložiště	FULL	1x/týden	Sobota	20:00
	DIFF	6x/týden	Sobota - Čtvrtek	18:00
	INCR	3x/denně	Pondělí - Pátek	9:00 11:00 13:00
File System - síťové úložiště uživatelů	FULL	1x/týden	Pátek	20:10
	DIFF	6x/týden	Sobota - Čtvrtek	18:00
	INCR	3x/denně	Pondělí - Pátek	9:00 11:00 13:00
File System - cestovní profily uživatelů	FULL	1x/týden	Sobota	8:00
	DIFF	6x/týden	Neděle - Pátek	21:00
	INCR	-		
File System - ostatní (diskové pole)	FULL	1x/týden	Sobota	10:00
	DIFF	-		
	INCR	6x/týden	Neděle - Pátek	20:00
Databáze - MS SQL	FULL	1x/týden	Sobota	5:00
	DIFF	6x/týden	Neděle - Pátek	5:00
	INCR	-		
Databáze - MS SQL master	FULL	1x/denně	Pondělí - Neděle	5:00
	DIFF	-		
	INCR	-		
Databáze - SQL Express	FULL	1x/týden	Sobota	4:00
	DIFF	6x/týden	Neděle - Pátek	4:00
	INCR	-		
Databáze - SQL Express master	FULL	1x/denně	Pondělí - Neděle	4:00
	DIFF	-		
	INCR	-		

Ve sloupci „typ zálohy“ byly v tabulce použity zkratky FULL pro plnou zálohu, DIFF pro diferenciální zálohu a INCR pro inkrementální typ zálohy. Všechny tyto zálohovací procesy byly nastaveny standardně přímo v zálohovacím softwaru NetVault: Backup až na dvě výjimky. Tento software totiž nemá možnost nativně nastavit vícenásobné vykonání jedno zálohovacího procesu v průběhu jednoho dne. Jak je vidět v tabulce u inkrementálního zálohování souborového systému síťových úložišť jsme takový požadavek měly, proto bylo nutné nastavit spouštění tohoto procesu odlišným způsobem. V nastavení jobu se namísto časového plánu jeho provádění nastavilo spouštění pomocí tak zvaného triggru (vyvolání). Aby byl tento trigger spouštěn v časech, ve kterých bylo navrženo, byl vytvořen skript, který zálohovací proces spouští a zároveň se vytvořily systémové události pomocí plánovače úloh systému (event manager), ve kterých se aktivuje vytvořený skript v předdefinovaných časech.

5.3.6 Nastavení způsobu zálohování

Byly nastaveny dva způsoby zálohování a to následujícím způsobem:

1. Po síti SAN se data přenáší do SVTL, z SVTL je pak prováděn přesun (migrace) do VTL. Ihned po dokončení tohoto procesu je provedena duplikace zálohovaných dat na dislokovanou fyzickou pásku opět po síti SAN.

Tímto způsobem se zálohuje:

- Poštovní server (Exchange server)
 - Databázové servery (SQL server, Informix)
 - Souborové servery
2. Po síti LAN do VTL a následně probíhá duplikace na dislokovanou fyzickou pásku po síti SAN.

Tímto způsobem se zálohují všechny ostatní servery.

5.3.7 Logy a stav zálohovacích procesů

Podobně jako je tomu u operačních systémů i u systému pro zálohování je dobré mít přehled o veškerém dění a stavu jak celého zálohovacího systému, tak o jednotlivých prováděných operacích. Z tohoto důvodu je po celou dobu běhu NetVault: Backup serveru veškerá činnost zaznamenávána a ukládána do logů. Logy lze kdykoliv prohlížet

a kontrolovat v části Logs. V případě problémů se zálohováním s těchto logů obvykle velmi snadno a rychle zjistíme, kde nastala chyba a co je potřeba udělat pro nápravu.

5.3.8 Notifikace

Abychom nemusely logy stále sledovat a prohlížet, zda nedochází k nějakým problémům při zálohování, obsahuje NetVault nástroj pro automatické upozorňování nazvaný Global Notification. V něm si můžeme zvolit, na jaké události chceme být upozorňováni a jakým způsobem. V našem případě jsme notifikace nastavily podle následující tabulky.

Tabulka 12 Přehled nastavení notifikací

Třída notifikací	Událost notifikace	Způsob notifikace
BakBone Time	Server Time Inconsistency	E-mail
BakBone Time	Server Time Unknown	E-mail
BakBone Time	Server Time Unknown	Operator message
Device	Device forced offline	E-mail
Device	Device gone offline	E-mail
Device	Drive unavailable	E-mail
Device	Library gone offline	E-mail
Device	No suitable drive	E-mail
Device	No suitable drive	Operator message
Job	All Job Retries Failed	E-mail
Job	Job died	E-mail
Job	Job failed	E-mail
Job	Job Hold	E-mail
Job	Job stopped	E-mail
Licensing	License Exceeded	E-mail
Licensing	License Exceeded	Operator message
Licensing	License Expiring	E-mail
Licensing	License Expiring	Operator message
Log Daemon	Home drive becoming full	E-mail
Machines	Client Down	E-mail

5.3.9 Obnova – ověření správného provedení zálohování

Nejlepším způsobem ověření správného a bezproblémového fungování zálohování je obnova dat. V ideálním případě, kdy systémy dlouhodobě běží, jak mají, uživatelé si omylem neumazávají důležitá data, ani nenastane žádná jiná katastrofa či příčina ztráty

dat, by tento krok nemusel být nikdy zapotřebí. Ovšem všichni ze zkušenosti víme, že v oboru informačních technologií tomu tak není, proto je dobré mít jistotu, že všechny naše provedené a uchovávané zálohy jsou v pořádku. To ověříme poměrně snadno tím, že provedeme obnovu dat ze zálohy. A to ať už cíleně, když se ozve nějaký uživatel, že si něco omylem smazal nebo jen namátkově pro kontrolu. Tuto kontrolu je dobré provádět pravidelně, abychom zabránili mnohem větší katastrofě při zjištění, že nám vyhořela serverovna a zálohy jsou nečitelné. V naší organizaci máme nastavené pravidlo, že se tato kontrola provádí pravidelně minimálně jednou za měsíc. Obnova dat se provádí opět pomocí zálohovacího nástroje NetVault: Backup přes obnovovací nástroj Restore.

6 ZHODNOCENÍ PROJEKTU A SMĚR DALŠÍHO VÝVOJE

Zálohování je nasazeno, zprovozněno a otestováno. Na řadě je závěrečné hodnocení, zda byly splněny podmínky a cíle, které byly před zahájením práce na projektu zadány.

6.1 Zhodnocení

Datové centrum je již v plném provozu a jeho zálohování běží bez problémů. Všechny zadané povinné podmínky, které byly kladeny na řešení zálohování datového centra, se podařilo bezesbytku splnit. Dokonce jsme se posunuli ještě o pomyslný krůček dál, protože díky velké rychlosti provádění zálohování dat i případné jejich obnovy a zároveň malé spotřebě zdrojů potřebných pro provádění záložních kopií jsme mohli nastavit provádění zálohování inkrementální metodou nejdůležitějších částí dat několikrát denně, u nás konkrétně třikrát, a to navíc v pracovní době, což nám dává větší možnost výběru obnovy dat k určitému nejen datu, ale dokonce i času. To samozřejmě nejvíce vítají uživatelé systému, pro které je tato možnost velmi vítanou novinkou v zabezpečení jejich pracovních dat.

Neměl bych opomenout vysvětlení, z jakého důvodu se zde nezmiňuji o zálohování poštovního serveru MS Exchange, přestože byla tato část projektu jednou s podmínek pro realizaci. Jak již jsem v textu několikrát zmiňoval, mým úkolem bylo realizovat zálohování nového datového centra v rámci konkrétního organizačního celku. Celá organizace má ale těchto organizačních celků několik a v každé z částí byl do práce na celkovém projektu určen někdo jiný. To se samozřejmě neobešlo bez vzájemné spolupráce a konzultace, ale konkrétní věci závislé na dané části organizace si již každý řešil ve své režii. S celkovým upgradem bylo navíc rozhodnuto, že některé služby systémů budou pro úsporu prostředků centralizovány. To se právě dotklo mimo jiné i poštovních serverů, které byly dříve nasazeny na všech lokalitách. Tímto krokem se ale nasazení poštovního serveru MS Exchange omezilo pouze na dvě lokality, na jedno centrální umístění a jedno vedlejší. Ani jedno z nich, ale nebylo to, které jsem dostal k realizaci já. Protože však jeden z cílů byl i sjednocení systému zálohování v rámci celé organizace, tato podmínka přesto musela být splněna.

Zjednodušeně se dá říct, že stanovené požadavky a cíle na realizaci projektu zálohování a obnova dat datových center byly splněny.

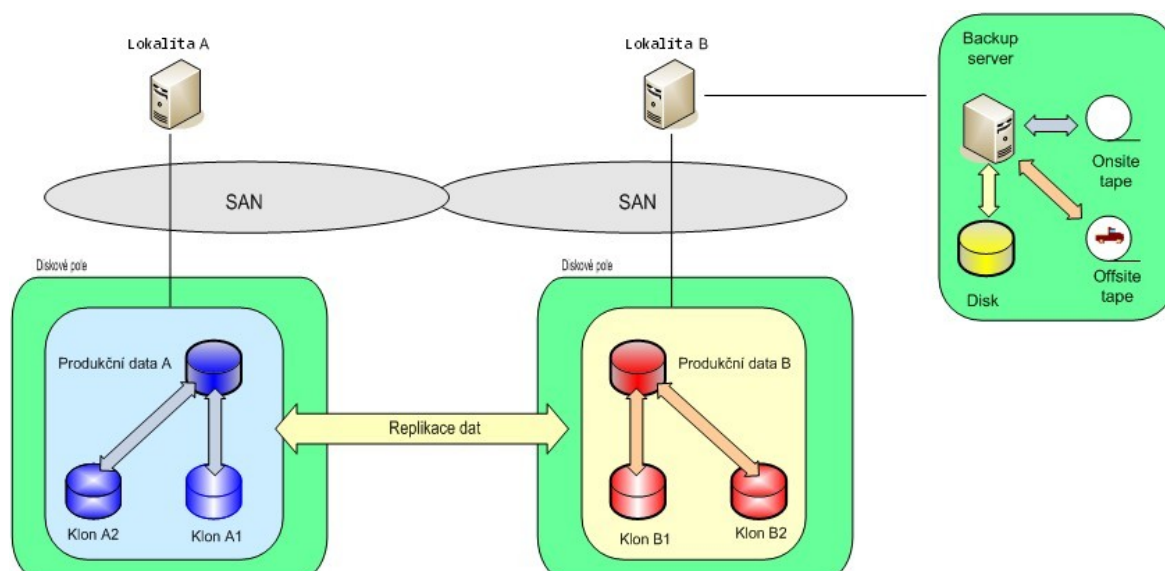
6.2 Další vývoj

Realizací tohoto projektu však veškerá práce nekončí. Tak jak se vyvíjí trh a technologie, je potřeba myslet i v této oblasti do budoucna. Již dnes padají podněty na vylepšení a ještě větší zabezpečení dat a technologií. Mým návrhem, kterým se již začínám zabývat, je vyřešení této problematiky pomocí clusterových technologií. Konkrétně by se mělo jednat o způsob zabezpečení dat pomocí Geoclusteru. Nasazení nových datových center s velkými kapacitami úložišť v naší organizaci k tomu přímo vybízejí. Z počátku by se mohlo přejít k realizaci tzv. offline clusteru a do vzdálenější budoucnosti i do online clusteru. Tyto technologie jsou ale poměrně nákladné, především druhá ze zmíněných.

Z nových technologií s vyšším stupněm zabezpečení již využíváme virtualizační techniky a lokální clusterová řešení pro stěžejní části systémů. To ale pokrývá především zvýšení dostupnosti aplikací nikoliv však dostupnosti dat, obzvláště při havárii celé lokality. Proto jsem navrhl řešení pomocí zmíněných geoclusterů, které podrobněji popisují v teoretické části méj diplomové práce.

Pro úplnost upřesním rozdíl mezi off-line a on-line geoclusterovým řešením na následujících obrázcích.

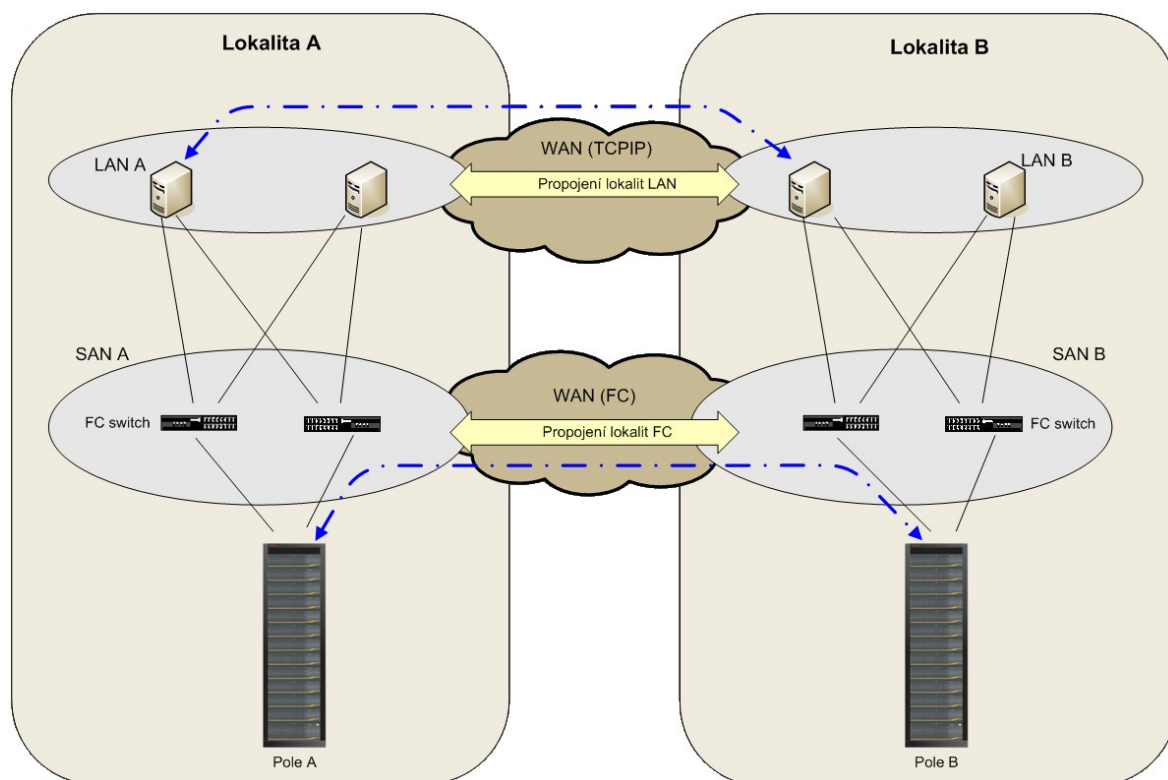
6.2.1 Offline geocluster



Obrázek 23 Schéma offline geoclusteru

Při použití offline geoclusterů běží lokality na sobě nezávisle, každá strana obsluhuje svou část klientů. Využívá se pouze vzájemné replikace dat na úrovni diskových polí, kdy strana A je zároveň záložní stranou strany B a naopak. Při havárii celé lokality lze obnovit data z protější strany.

6.2.2 Online geocluster



Obrázek 24 Schéma online geoclusteru

Online geocluster je v podstatě spojení funkcionalit lokálního clusteru a offline geoclusteru. To znamená, že se opět každá lokalita stará primárně o své klienty, ale zároveň plní roli clusteru pro lokalitu protější. Čili v případě výpadku jedné lokality ji téměř okamžitě nahrazuje lokalita vzdálená a provoz není v ideálním případě vůbec přerušeno. Toto řešení je jedno z nejlepších a nejbezpečnějších vůbec, ale také je velmi nákladné. Zejména je zapotřebí mít lokality vzájemně propojeny vysokorychlostní datovou sítí na velkou vzdálenost.

Můj předpoklad je, že k realizaci tohoto řešení v budoucnu dojde, vše ale závisí od vývoje trhu, ekonomiky a technologií.

ZÁVĚR

Cílem této diplomové práce bylo navrhnout a realizovat projekt zálohování a obnovy dat datových center. Všechny body zadání byly zpracovány. V souladu se zadáním práce byl proveden teoretický přehled problematiky zálohování a obnovy dat a byl vytvořen přehled v současnosti používaných řešení. Dále byla provedena analýza současného stavu zálohování dat v rámci příslušného organizačního celku. Na základě těchto poznatků byl proveden návrh způsobu zálohování a obnovy dat těchto zdrojů. Podle vypracovaného návrhu byla provedena realizace řešení zálohování a byly zde popsány použité technologie. Zálohování bylo v konkrétním organizačním celku uvedeno do provozu a byla otestována jeho funkčnost. V samotném závěru práce bylo provedeno závěrečné hodnocení a byly navrženy možné směry budoucího vývoje.

Navržený a realizovaný projekt zálohování a obnovy dat datových center splnil všechny kritéria a požadavky, které na něj byly kladeny. Pro administrátory, kteří budou nasazený způsob využívat, je příjemné, že se nemusí učit novému softwarovému aplikačnímu vybavení pro zálohování, protože produkt NetVault: Backup server, kterým byl projekt realizován, již do jisté míry znají z minulého způsobu zálohování a obnovy dat. Projekt je přínosem i pro koncové uživatele, protože je o jejich data lépe postaráno a navíc mají možnost obnovy dat ve více časových úsecích. To znamená, že se mohou podle výběru vrátit k několika předchozím verzím rozpracovaného souboru dat. Další nespornou výhodou je připravenost řešení na další možný vývoj ve zvyšování zabezpečení dat. Tento krok lze provést formou vzájemného propojení lokalit datových center a upravením nebo přidáním směru zálohování pomocí jednotného zálohovacího softwaru.

Přínos této práce z mého pohledu má dva rozměry. Jednak se mi podařilo navrhnout a realizovat poměrně složitý projekt, který navíc slouží široké klientele a je s ním, podle ohlasů, spokojenost. Druhou věcí, kterou mi práce dala, je získání velkého množství nových informací z prostředí informačních technologií, které mohu dále využít k další práci.

ZÁVĚR V ANGLIČTINĚ

The aim of this thesis was to design and implement a project of backup and restoring data in data centers. In accordance with the assignment of the dissertation work a theoretical overview of the problem of data backing-up and data restoring was made and a survey of currently used solutions was created. Furthermore, an analysis of the current state of backup of the data across the whole organization was prepared. Based on these findings, a suggestion of the ways how to backup and restore the data of these resources was made. According to the draft prepared, the implementation of backup solutions was realized and the used technologies were described. Backup strategies were put into operation in a particular organizational unit and their functionality was tested. The final evaluation was made in the conclusion and the possible ways of future development were suggested.

The designed and implemented project of data backup and restore of the data in data centers has met all criteria and requirements that were placed on it. For administrators who will be using the suggested ways of backing-up and restoring of data it can be very comfortable that they will not be forced to learn how to use new software equipment, because the product NetVault: Backup Server, which was carried out, is to some extent already known from the previous method of data backup and their restore. The project is also beneficial for final users, because their data are better cared for and on top of that they can also have the option to restore data in multiple time periods. This means that they can return to the several previous unfinished versions of the data set and they can choose the most suitable one. Another advantage is that the solutions are prepared for possible future development in improving data security. This step can be performed by the interconnection of the locations of data centers and adding or modifying the direction of backup using united backup software.

In my opinion, the contribution of this work has two dimensions. First, I was able to design and implement a fairly complex project, which also serves a wide clientele, who are satisfied with it, by the response. The second thing, which the work contributed to me, is the fact that I obtained a large amount of new information from information technology surroundings, which can also be used for my further practice.

SEZNAM POUŽITÉ LITERATURY

- [1] LEBER, Jody. Windows NT: zálohování a obnova dat : příručka pro správce systémů. Praha : Computer Press, 1998. 282 s. ISBN 8072261231.
- [2] LEIXNER, Miroslav. PC - zálohování a archivace dat : Nestůjte za dveřmi. Praha : Grada Publishing, a.s., 1993. 394 s. ISBN 8085424738.
- [3] KASTNER, Aleš. Zálohování a archivace. Praha : GComp, 1997. 128 s. ISBN 8085649586.
- [4] BASL, Josef. Podnikové informační systémy: podnik v informační společnosti. Praha : Grada Publishing a.s., 2008. 283 s. ISBN 8024722798.
- [5] PRESTON, W. Backup and recovery. Sebastopol, CA: O'Reilly, c2007. 729 s. ISBN 978-059-6102-463.
- [6] DE GUISE, Preston. Enterprise systems backup and recovery: a corporate insurance policy. Boca Raton: CRC Press, c2009. 306 s. ISBN 14-200-7639-6.
- [7] NELSON, Steven. Pro data backup and recovery. New York: Springer Science Business Media, 2011. 280 s. ISBN 14-302-2662-5.
- [8] ENGLEWOOD, Cliffs. Data storage management: a systems approach. NJ: Prentice Hall, 1998. ISBN 978-013-8901-875.
- [9] POVOLNÝ, Jan. Zálohujeme, zálohujete, zálohují. *PCWorld* [online]. 29.07.2008 [cit. 2012-04-12]. Dostupné z: <http://pcworld.cz/software/zalohujeme-zalohujete-zalohuji-3893>
- [10] Orientační ceník záchrany dat. *Záchrana dat a obnova dat od roku 1991* [online]. 2011 [cit. 2012-04-12]. Dostupné z: <http://www.zachrana-dat.cz/cz/zachrana-dat-orientacni-cenik.html>
- [11] Slovník pojmů. *ZÁLOHOVÁNÍ.NET* [online]. 2010 [cit. 2012-04-15]. Dostupné z: <http://www.zalohovani.net/inpage/pojem/Zalohovani.Odvirovani-PC> [online]. 2010 [cit. 2012-04-19]. Dostupné z: <http://www.odvirovani.cz/zalohovani.html>
- [12] Zálohování. *Odvirování PC* [online]. 2010 [cit. 2012-04-19]. Dostupné z: <http://www.odvirovani.cz/zalohovani.html>

- [13] Zálohování a archivace. *Mendelova univerzita v Brně* [online]. 2003 [cit. 2012-04-29]. Dostupné z: <https://akela.mendelu.cz/~lidak/bif/obrtelova.doc>
- [14] LTO, DLT, SLR - co je lepší?. *STORAGE.CZ. Systémy úschovy dat s bezpečnou a efektivní správou.* [online]. 2011 [cit. 2012-04-19]. Dostupné z: <http://www.storage.cz/775-lto-dlt-slr-co-je-lepsi>
- [15] Solutions. *Apex Microsystems* [online]. 2009 [cit. 2012-04-19]. Dostupné z: http://www.apexmicrosystems.com/?page_id=518
- [16] Zálohování a archivace. *XANADU* [online]. 2012 [cit. 2012-04-20]. Dostupné z: <http://www.xanadu.cz/cs/it-produkty/serverova-reseni/zalohovani-a-archivace/>
- [17] KOTRLE, Miroslav. D2D2T. *Data v péči* [online]. 18.06.2009, roč. 2009, č. 14 [cit. 2012-04-20]. Dostupné z: <http://www.datavpeci.cz/webdvp.nsf/articles/4AA3FD7846A982D0C12575CF002F6632>
- [18] ALGER, Douglas. *Build the best data center facility for your business.* London: Pearson Education [distributor], c2005, 374 s. ISBN 15-870-5182-6.
- [19] Definice a rotace záloh. *STORAGE.CZ. Systémy úschovy dat s bezpečnou a efektivní správou* [online]. 2011 [cit. 2012-04-23]. Dostupné z: <http://www.storage.cz/511-definice-a-rotace-zaloh>
- [20] OPATŘIL, Michal. Když se zálohuje špatně. *Data v péči* [online]. 07.04.2008, roč. 2008, č. 9 [cit. 2012-04-23]. Dostupné z: <http://www.datavpeci.cz/webdvp.nsf/articles/31A58968737C9134C125750C0034D831>
- [21] Ztráta dat. *ORCANA solutions* [online]. 2008 [cit. 2012-04-28]. Dostupné z: <http://www.orcana.cz/datalost.html>
- [22] KNETTIG, Tomáš. Aktuální pojem Dostupnost. *LOGIN* [online]. 2009, roč. 4, č. 3 [cit. 2012-04-29]. Dostupné z: http://www.kancelarskestroje.cz/uploads/assets/stories/pdf/clanek_ha_dr.pdf
- [23] Double-Take - Availability. *QTD spol s r.o.* [online]. 2011 [cit. 2012-04-29]. Dostupné z: http://www.qtd.cz/double-take_availability.php

- [24] Tandberg SLR Data Cartridge. *Tandberg Data International* [online]. 2007 [cit. 2012-03-25]. Dostupné z: http://www.tandbergdata.com/support/tandberg_slr5_data_cartridge_specification
- [25] Počítačové zálohovací pásky. *Fujifilm Česká republika* [online]. 2010 [cit. 2012-04-29]. Dostupné z: <http://www.cz.fujifilm.cz/cs/produkty/media/pocitacove%20pasky/lto.ep/>
- [26] Sony announces availability of its fifth generation tape drive and media in celebrating 10 years of AIT. *Sony Europe* [online]. 2006 [cit. 2012-04-29]. Dostupné z: <http://www.sony-europe.com/res/attachment/file/79/1159199085479.pdf>
- [27] HDS: Hitachi Universal Storage Platform (USP) VM - Advanced Storage Services Platform. *Hitachi Data Systems (HDS) - Data Storage Products Services Solutions* [online]. 2012 [cit. 2012-05-14]. Dostupné z: <http://www.hds.com/products/storage-systems/universal-storage-platform-vm.html>
- [28] HDS: Hitachi Adaptable Modular Storage (AMS) 2000 Family - AMS 2100 | 2300 | 2500. *HDS: Hitachi Universal Storage Platform (USP) VM - Advanced Storage Services Platform* [online]. 2012 [cit. 2012-05-14]. Dostupné z: <http://www.hds.com/products/storage-systems/adaptable-modular-storage-2000-family/>
- [29] Actidata actiLib Library: Products. ACTIDATA. *Actidata - We care about data* [online]. 2010 [cit. 2012-05-14]. Dostupné z: <http://www.actidata.com/en/products/actidata-actilib-library.html>

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

ARJ	Komprimační algoritmus.
ATA	(Advanced Technology Attachment) Paralelní sběrnice pro připojování zařízení k uchovávání dat.
Blu-ray	(Blue ray) Optický disk 3. generace s výrazně vyšší hustotou záznamu.
CD	(Compact Disc) Optický disk.
DVD	(Digital Video Disc) Optický disk 2. generace s větší kapacitou.
ERM	(Extended Runtime Module) Rozšiřující modul pro UPS.
GUI	(Graphical User Interface) Grafické uživatelské rozhraní.
HD DVD	(High Definition DVD) Optický disk 3. generace s výrazně vyšší hustotou záznamu.
HW	(Hardware) Pevné vybavení počítače.
KVM	(Keyboard, Video, Mouse) Zařízení umožňující ovládat více počítačů z jedné klávesnice, monitoru a myši.
LAN	(Local Area Network) Počítačová síť pokrývající malé geografické území.
MP3	Formát ztrátové komprese zvukových souborů.
MSA	(Modular Smart Array) Modulární diskové pole.
OS	(Operating System) Operační systém počítače.
PDU	(Power Distribution Unit) Napájecí panely.
RACK	Standardizovaný systém umožňující přehlednou montáž a propojování různých elektrických a elektronických zařízení.
RAID	(Redundant Array of Inexpensive/Independent Disks) Vícenásobné diskové pole laciných/nezávislých disků.
RAR	Komprimační algoritmus.
SAS	(Serial Attached SCSI) Sériově zapojené SCSI rozhraní.
SATA	(Serial ATA) Sériově zapojené ATA rozhraní.

SCSI	(Small Computer System Interface) Standardní rozhraní pro výměnu dat mezi externími nebo interními počítačovými zařízeními a počítačovou sběrnici.
SQL	(Structured Query Language) Standardizovaný dotazovací jazyk používaný pro práci s daty v relačních databázích.
SSD	(Solid-State Drive) Typ pevného disku neobsahující pohyblivé mechanické části.
SW	(Software) Aplikační a programové vybavení počítače.
TCP/IP	Sada protokolů pro komunikaci v počítačové síti.
UPS	(Uninterruptible Power Supply) Zdroj nepřerušovaného napájení.
USB	(Univerzal Serial Bus) Univerzální sériové rozhraní počítače.
ZIP	Komprimační algoritmus; název zálohovacího zařízení a komprimačního software.

SEZNAM OBRÁZKŮ

<i>Obrázek 1 Příčiny ztráty dat [21]</i>	13
<i>Obrázek 2 Princip plné zálohy</i>	19
<i>Obrázek 3 Princip diferenciální zálohy</i>	19
<i>Obrázek 4 Princip inkrementální zálohy</i>	20
<i>Obrázek 5 Schéma zapojení VTL</i>	26
<i>Obrázek 6 Schéma zapojení sVTL</i>	27
<i>Obrázek 7 Schéma připojení DAS k počítačům [15]</i>	28
<i>Obrázek 8 Schéma připojení NAS k síti [15]</i>	29
<i>Obrázek 9 Schéma zapojení SAN sítě [15]</i>	30
<i>Obrázek 10 Schéma zálohování typu D2T [16]</i>	31
<i>Obrázek 11 Schéma zálohování typu D2D [16]</i>	31
<i>Obrázek 12 Schéma zálohování typu D2D2T [16]</i>	32
<i>Obrázek 13 Schéma fungování clusteru [23]</i>	37
<i>Obrázek 14 Schéma zapojení zálohování ve stávajícím stavu</i>	43
<i>Obrázek 15 Hitachi USP-VM [27]</i>	47
<i>Obrázek 16 Hitachi AMS 2300 [28]</i>	48
<i>Obrázek 17 actiLib Library LTO Series 4U [29]</i>	49
<i>Obrázek 18 Schéma zapojení nového způsobu zálohování</i>	51
<i>Obrázek 19 Rozdělení instalovaných komponent</i>	54
<i>Obrázek 20 Grafické rozhraní konzole NetVault: Backup Serveru</i>	55
<i>Obrázek 21 NetVault Menu Options</i>	55
<i>Obrázek 22 NetVault Configurator</i>	56
<i>Obrázek 23 Schéma offline geoclusteru</i>	62
<i>Obrázek 24 Schéma online geoclusteru</i>	63

SEZNAM TABULEK

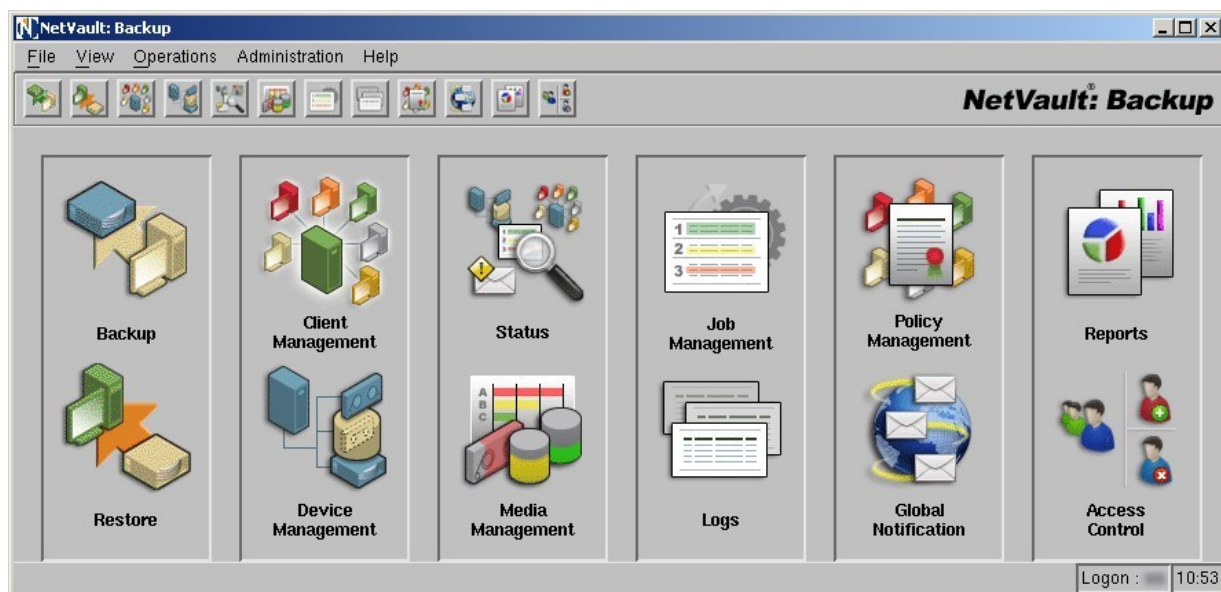
<i>Tabulka 1 Orientační ceny záchrany dat z pevného disku [10]</i>	14
<i>Tabulka 2 Plán záchrany dat z diskového pole RAID [10]</i>	14
<i>Tabulka 3 Orientační ceny záchrany dat z Flash disku a paměťové karty [10]</i>	15
<i>Tabulka 4 Kapacity a rychlosti DAT/DDS pásek [25]</i>	22
<i>Tabulka 5 Kapacity a rychlosti AIT pásek [26]</i>	23
<i>Tabulka 6 Kapacity a rychlosti SLR pásek [24]</i>	23
<i>Tabulka 7 Kapacity a rychlosti DLT pásek [25]</i>	24
<i>Tabulka 8 Kapacity a rychlosti SDLT pásek [25]</i>	24
<i>Tabulka 9 Kapacity a rychlosti LTO pásek [25]</i>	25
<i>Tabulka 10 Třídy vysoké dostupnosti [22]</i>	35
<i>Tabulka 11 Přehled nastavení zálohovacích procesů</i>	57
<i>Tabulka 12 Přehled nastavení notifikací</i>	59

SEZNAM PŘÍLOH

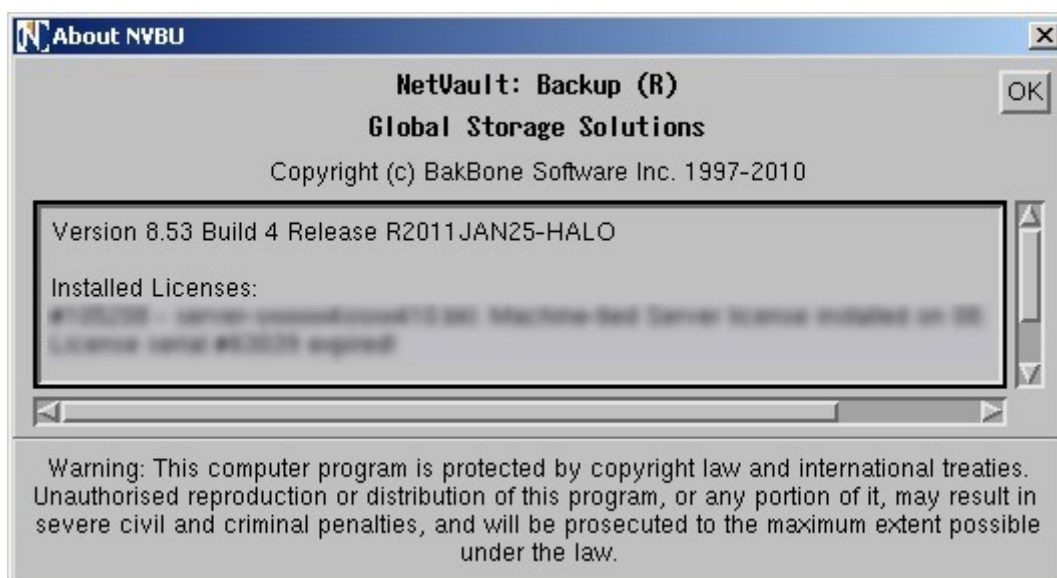
Příloha P I: Nové uživatelské rozhraní NetVault

PŘÍLOHA P I: NOVÉ UŽIVATELSKÉ ROZHRANÍ NETVAULT

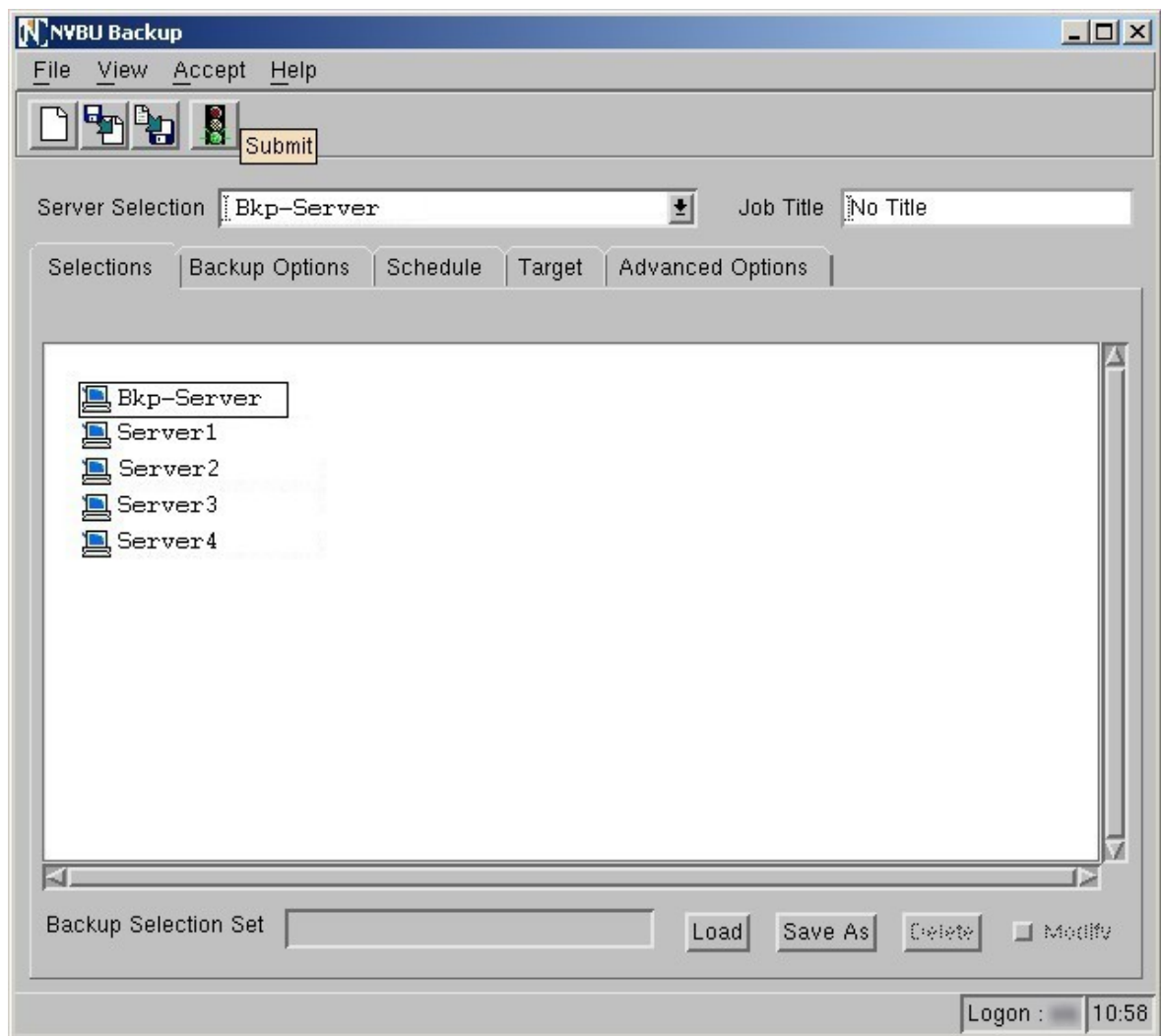
V průběhu dokončování mé diplomové práce byla vydána a v zápatí i nainstalována v naší organizaci nová verze zálohovacího software NetVault: Backup (verze 8.53). Některé části byly přepracovány, jiné zůstaly téměř nezměněny. Pro ukázkou přikládám kopie obrazovek vybraných částí nové verze. (Citlivé informace na snímcích jsou, z pochopitelných důvodů, rozmazány.)



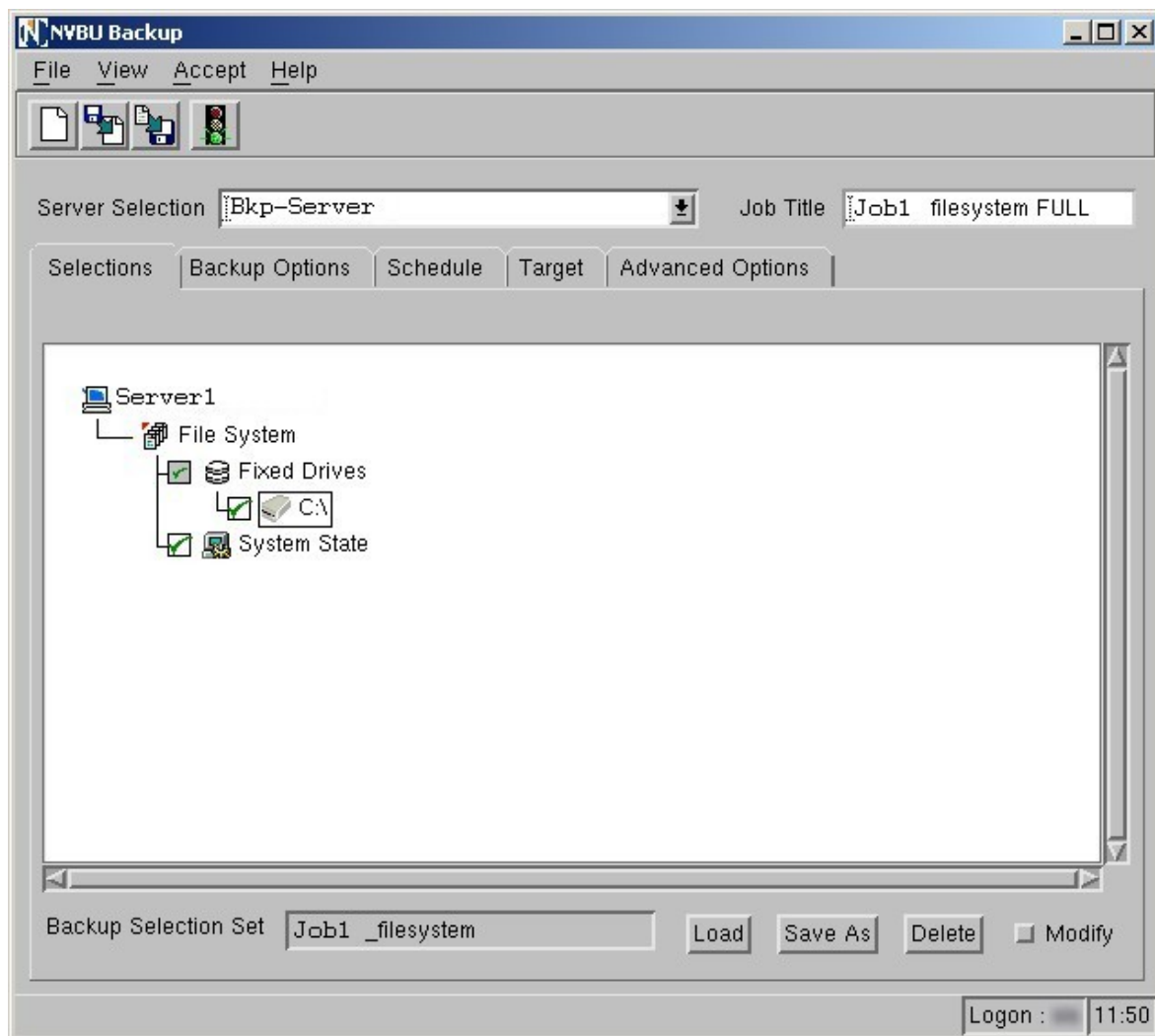
Obrázek P1 Nové uživatelské rozhraní NetVault: Backup



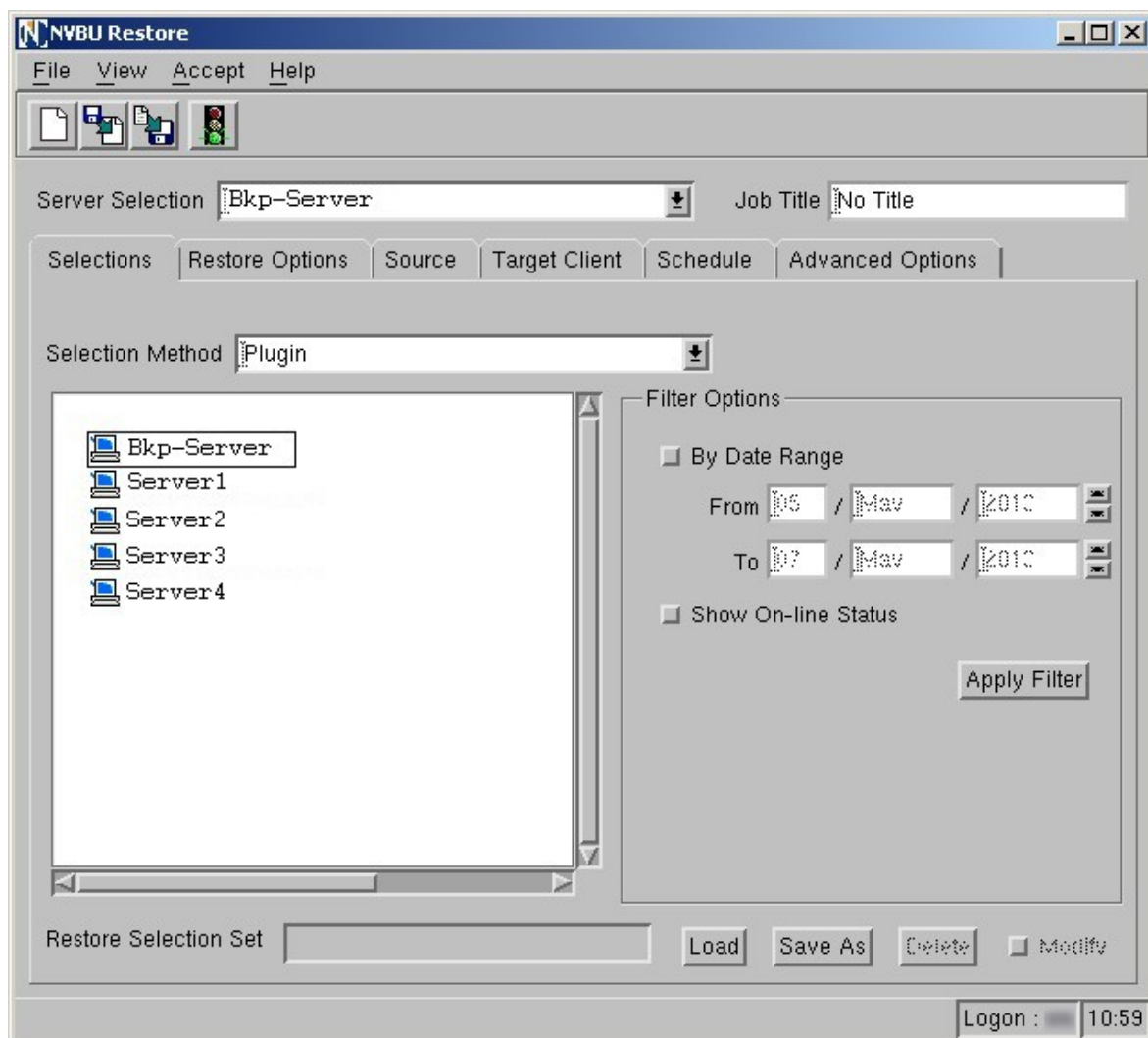
Obrázek P2 Informace o produktu NetVault: Backup



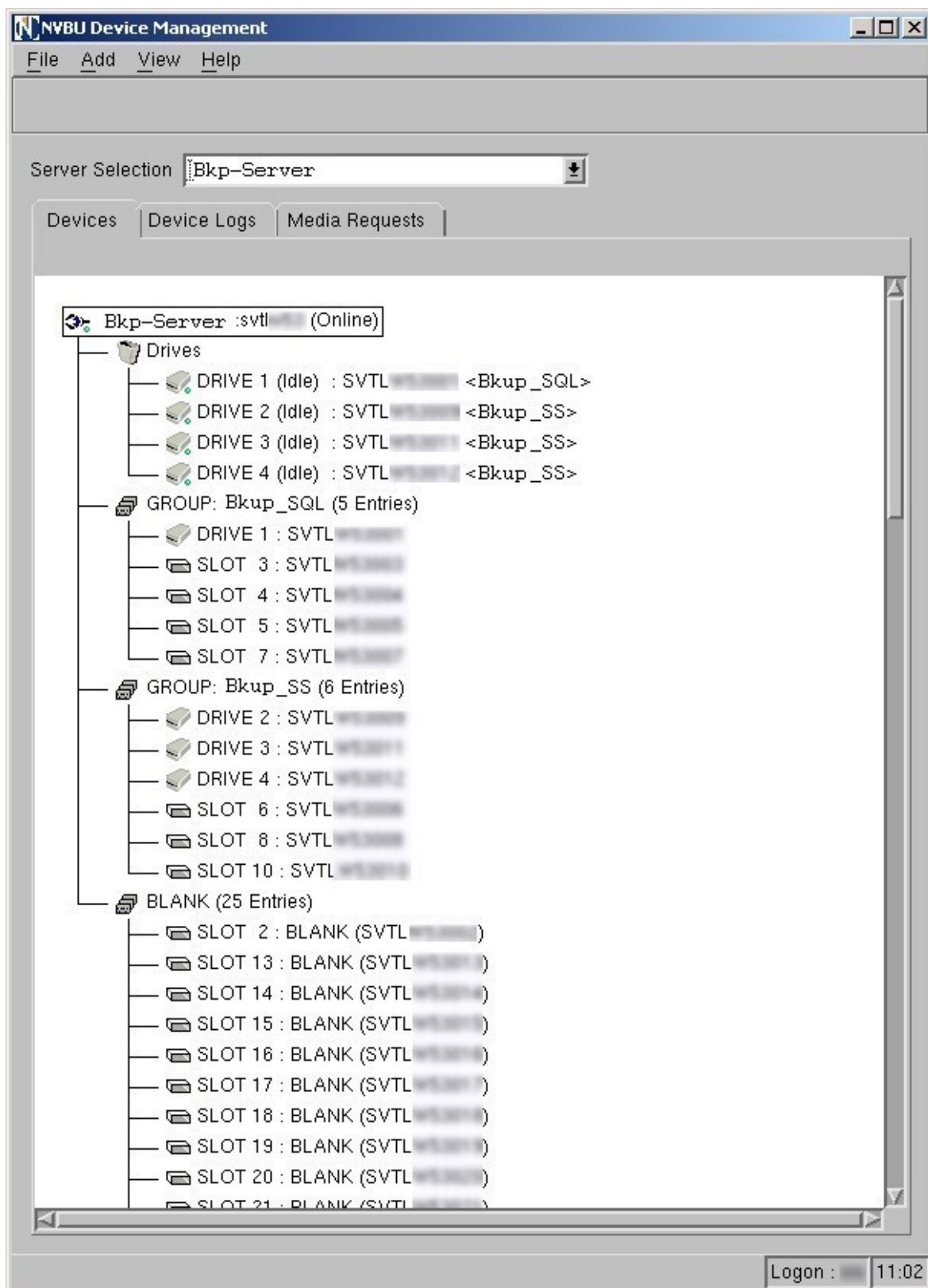
Obrázek P3 NetVault: Backup – nastavení zálohování



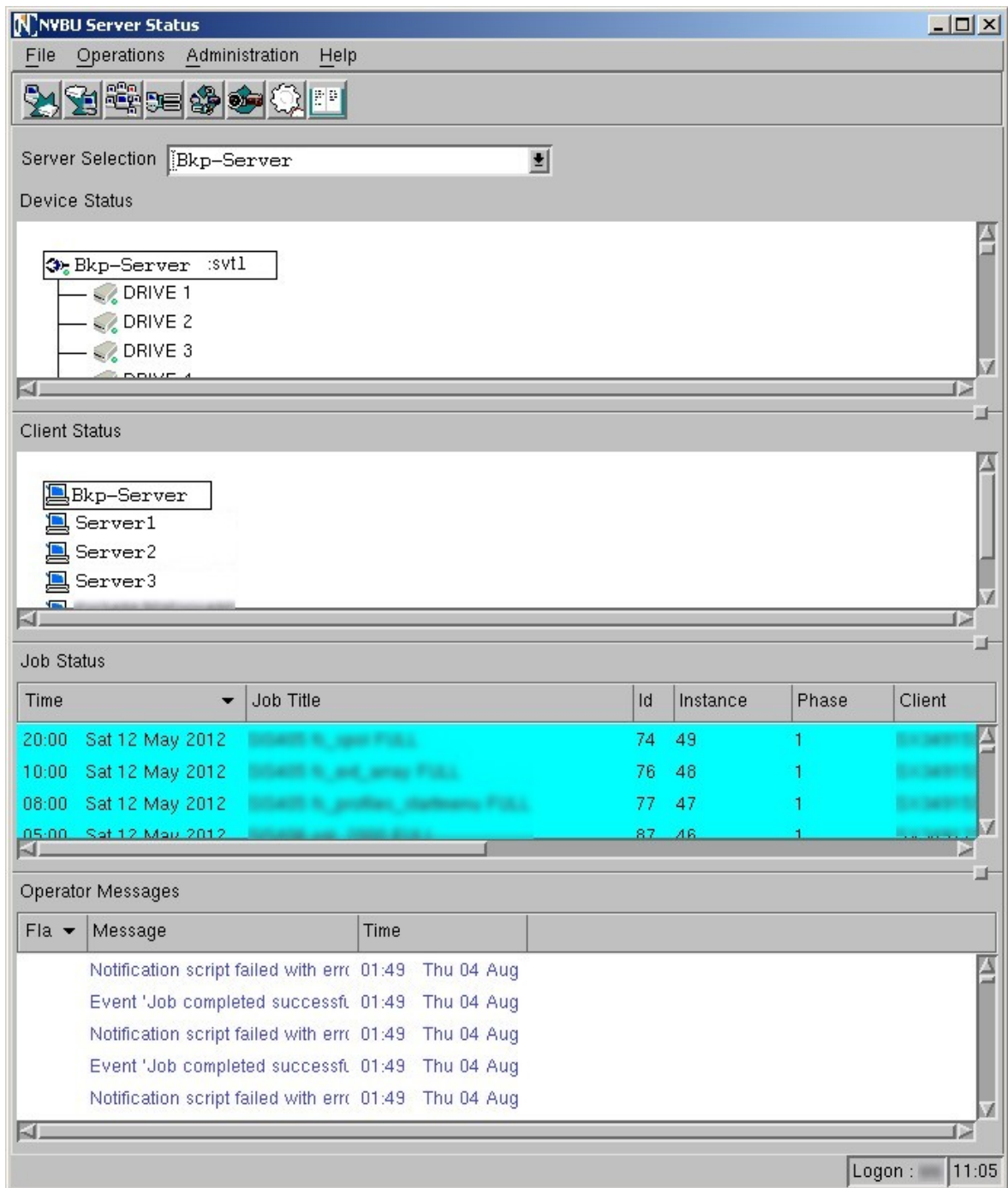
Obrázek P4 NetVault: Backup – nastavení zálohování vybraných částí



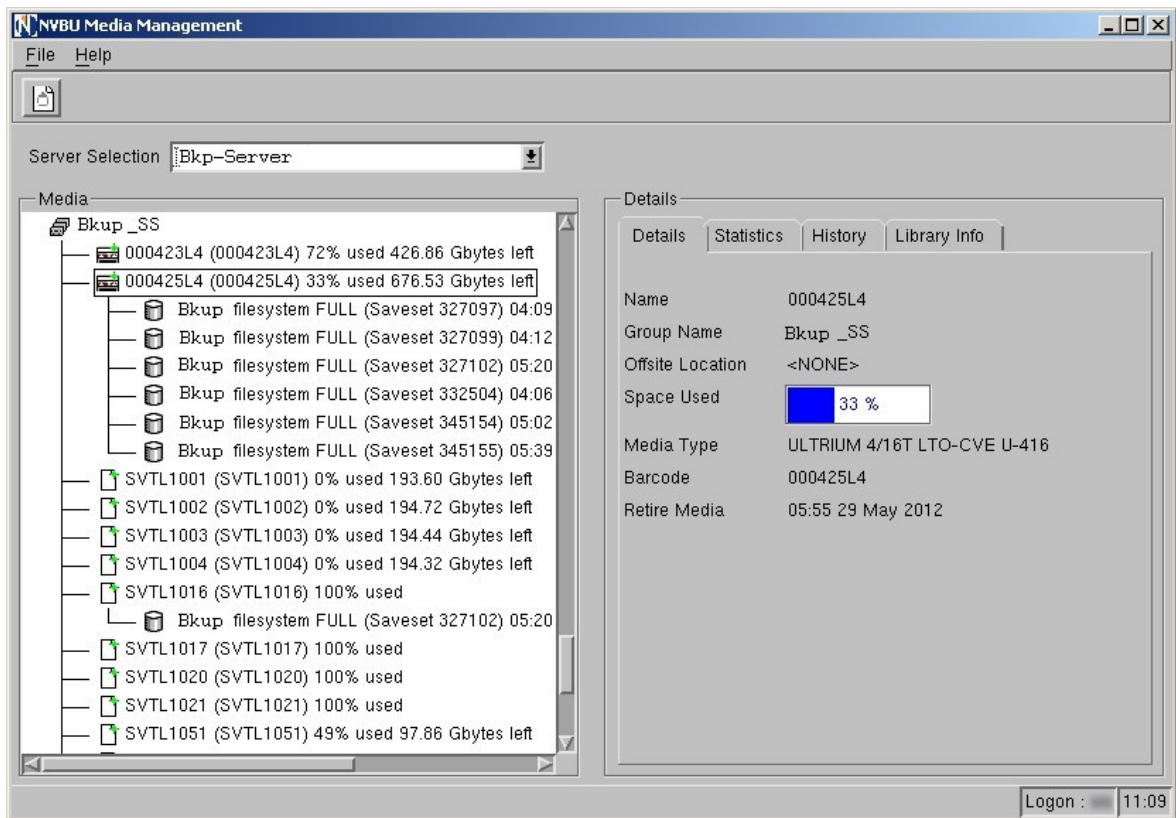
Obrázek P5 NetVault: Backup – nastavení výběru pro obnovu



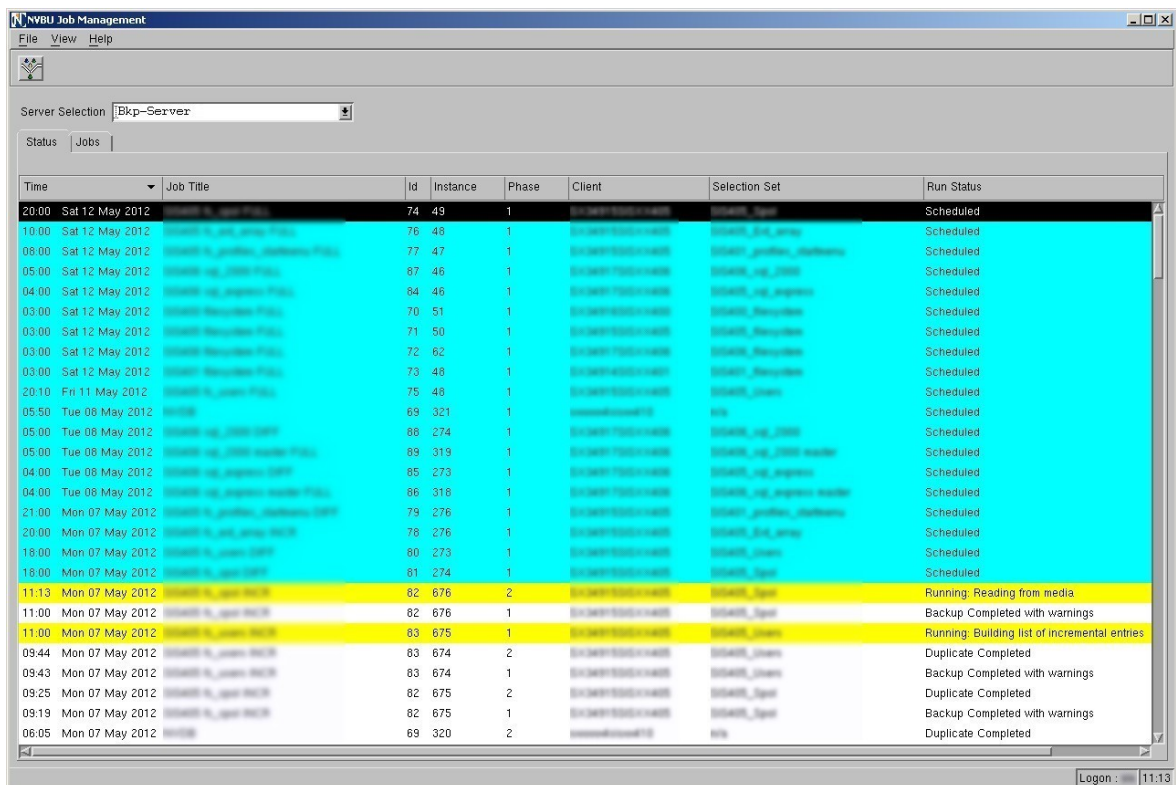
Obrázek P6 NetVault: Backup – správa zálohovacích jednotek



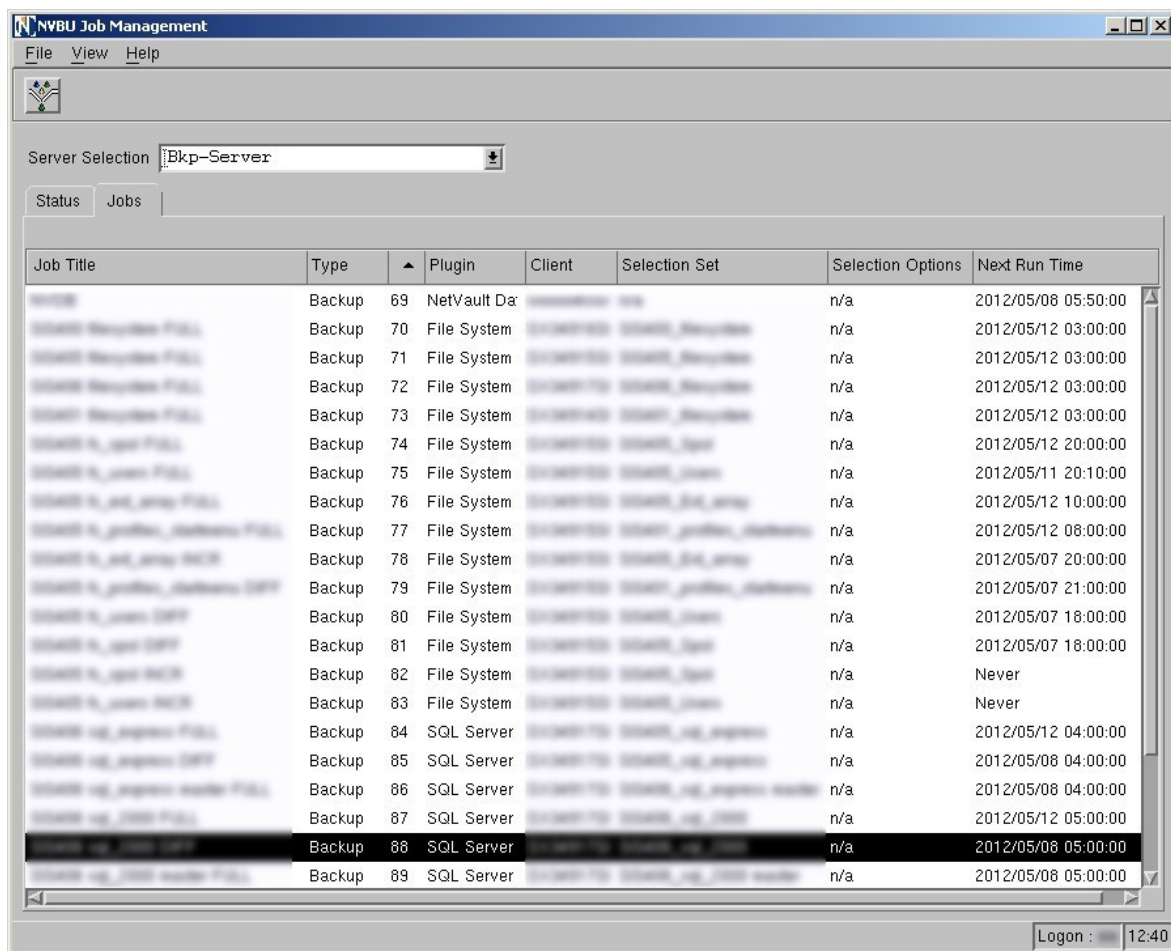
Obrázek P7 NetVault: Backup – okno pro přehled stavu zálohovacího serveru



Obrázek P8 NetVault: Backup – správa zálohovacích medií



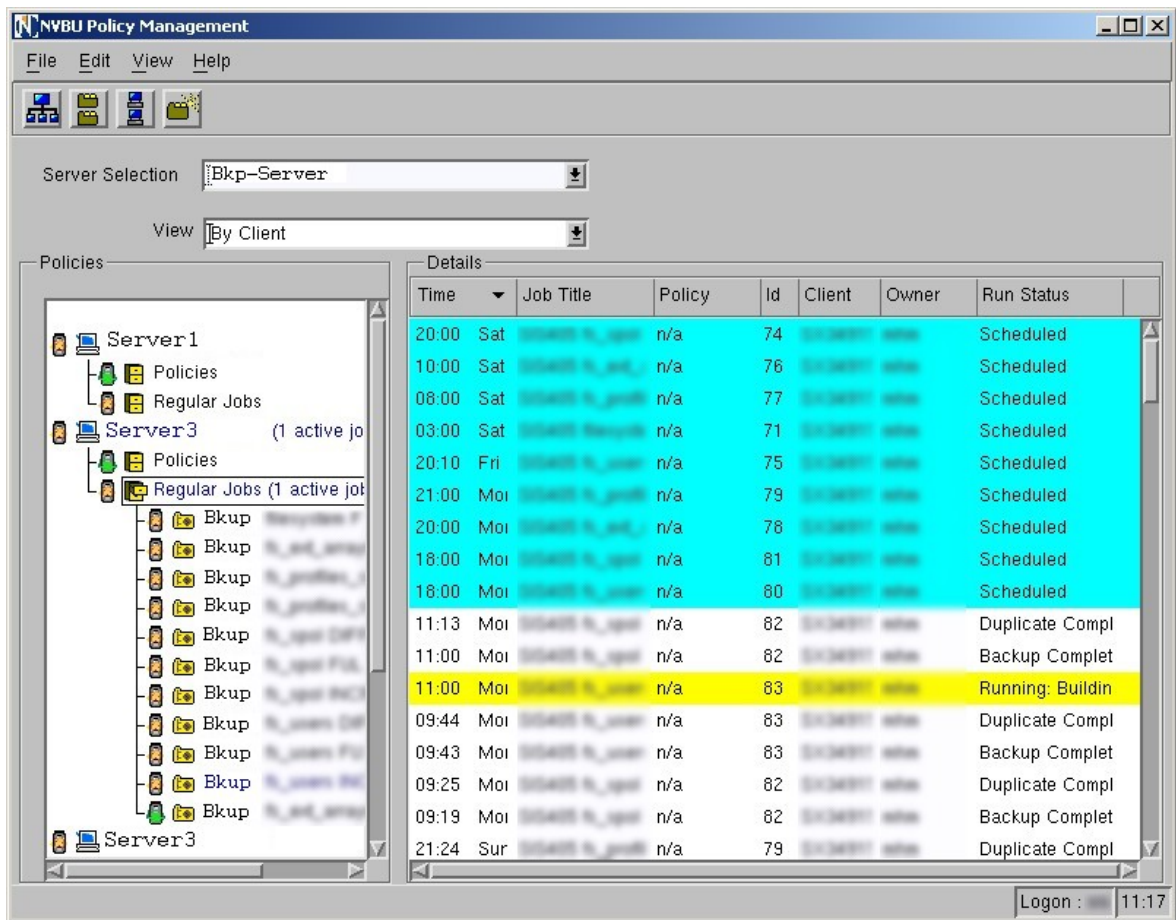
Obrázek P9 NetVault: Backup – okno pro přehled stavu zálohovacích procesů



Obrázek P10 NetVault: Backup – okno pro správu zálohovacích procesů

Warning Level	Date / Time	Job Id	Class	Client	Log Message
● Job Message	2012/05/07 05:56:32	69	Media	...	Check Media Manager database raw file structure
● Job Message	2012/05/07 05:56:36	69	Media	...	Check Media Manager database table integrity
● Job Message	2012/05/07 05:56:41	69	Media	...	Check Media Manager database inter-table references and dependencies
● Job Message	2012/05/07 05:57:15	69	Media	...	Media Manager database has successfully passed all checks
● Job Message	2012/05/07 05:57:15	69	Jobs	...	Saveset 355038 (instance 320)
ⓘ Job Message	2012/05/07 05:57:16	69	Media	...	(...) Drive 'DRIVE 3:...' has completed its transfer
ⓘ Job Message	2012/05/07 05:57:26	69	Schedule	...	Scheduler database successfully compacted
ⓘ Job Message	2012/05/07 05:57:39	69	Media	...	Media Manager database successfully compacted
⚠ Warning	2012/05/07 05:57:39	69	Jobs	...	Job Status: Backup Completed with warnings
● Job Message	2012/05/07 05:57:39	69	Jobs	...	Finished job 69, phase 1 (instance 320)
● Job Message	2012/05/07 05:57:40	69	Jobs	...	Starting job 69 '...' (Phase 2, Instance 320) for UID 0
● Job Message	2012/05/07 05:57:40	69	Data Plugin	...	Copying data from Job '...' (Id 69, Instance 320, Phase 1), Saveset 355038
ⓘ Job Message	2012/05/07 05:57:40	69	Media	...	(...:actlib_4u) Job has been assigned media '000432L4' <G_FILESYSTEM> in 'DRIVE 2:...
ⓘ Job Message	2012/05/07 06:01:09	69	Media	...	(...:actlib_4u) Media in 'DRIVE 2:...' assigned to job ready for data transfer
ⓘ Job Message	2012/05/07 06:01:09	69	Media	...	(...) Job has been assigned media 'VTL...' <G_FILESYSTEM> in 'DRIVE 3:...
● Job Message	2012/05/07 06:01:09	69	Media	...	(...) Media in 'DRIVE 3:...' assigned to job ready for data transfer
● Job Message	2012/05/07 06:01:26	69	Media	...	(...) Drive 'DRIVE 3:...' has completed its transfer
ⓘ Job Message	2012/05/07 06:01:26	69	Media	...	(...) Job has been assigned media 'VTL...' <G_FILESYSTEM> in 'DRIVE 3:...
● Job Message	2012/05/07 06:01:27	69	Media	...	(...) Media in 'DRIVE 3:...' assigned to job ready for data transfer
● Job Message	2012/05/07 06:05:14	69	Media	...	(...) Drive 'DRIVE 3:...' has completed its transfer
● Job Message	2012/05/07 06:05:32	69	Data Plugin	...	Job completed
● Job Message	2012/05/07 06:05:32	69	Jobs	...	Job Status: Duplicate Completed
ⓘ Job Message	2012/05/07 06:05:32	69	Media	...	(...:actlib_4u) Drive 'DRIVE 2:...' has completed its transfer
● Job Message	2012/05/07 06:05:32	69	Jobs	...	Finished job 69, phase 2 (instance 320)
● Job Message	2012/05/07 09:00:02	N/A	Schedule	...	Job 82 '...' activated by trigger '...'
● Job Message	2012/05/07 09:00:02	82	Jobs	...	Starting job 82 '...' (Phase 1, Instance 675) for UID 0
● Job Message	2012/05/07 09:00:02	N/A	Schedule	...	Job 83 '...' activated by trigger '...'
● Job Message	2012/05/07 09:00:02	83	Jobs	...	Starting job 83 '...' (Phase 1, Instance 674) for UID 0
● Job Message	2012/05/07 09:00:03	83	Data Plugin	...	File System 7.5.17 Backup
● Job Message	2012/05/07 09:00:03	82	Data Plugin	...	File System 7.5.17 Backup

Obrázek P11 NetVault: Backup – okno logů jednotlivých operací



Obrázek P12 NetVault: Backup – okno pro správu a nastavení bezpečnostních politik