

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačních technologií



Bakalářská práce

Zálohování a archivace dat

Radek Bartošek

© 2015 ČZU v Praze

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Katedra informačních technologií

Provozně ekonomická fakulta

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Radek Bartošek

Informatika

Název práce

Zálohování a archivace dat

Název anglicky

Data backup and archivation

Cíle práce

Hlavním cílem práce je návrh efektivního řešení zálohy a archivace dat.

Dílčím cílem je popsat různé možnosti a způsoby pro zálohu a archivaci dat, poukázat na jejich rozdíly, klady a zápory.

Praktická část bude soustředěna na zálohování a archivaci dat v konkrétní firmě. Bude provedena analýza způsobů zálohy a archivace dat v této firmě.

Metodika

Metodika bakalářské práce je založena na shromáždění a analyzování odborných informací týkající se daného tématu. Praktická část se zabývá analýzou metod zálohování a archivace dat ve firmě. IT technik této firmy poskytne informace o aktuálních postupech zálohy a archivace dat. Na základě této analýzy bude navrženo efektivnější řešení.

Doporučený rozsah práce

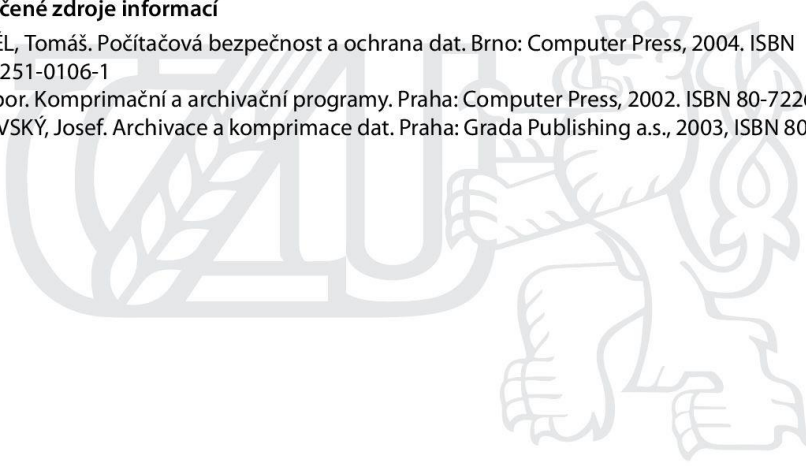
30-40 stran

Doporučené zdroje informací

DOSEDĚL, Tomáš. Počítačová bezpečnost a ochrana dat. Brno: Computer Press, 2004. ISBN 80-251-0106-1

KŘÍŽ, Libor. Komprimační a archivační programy. Praha: Computer Press, 2002. ISBN 80-7226-757-4

PECINOVSKÝ, Josef. Archivace a komprimace dat. Praha: Grada Publishing a.s., 2003, ISBN 80-247-0659-8



Předběžný termín obhajoby

2015/06 (červen)

Vedoucí práce

Ing. Martin Havránek, Ph.D.

Elektronicky schváleno dne 31. 10. 2014

Ing. Jiří Vaněk, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 11. 11. 2014

Ing. Martin Pelikán, Ph.D.

Děkan

V Praze dne 02. 03. 2015

Čestné prohlášení

Prohlašuji, že svou bakalářskou práci "Zálohování a archivace dat" jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu literatury na konci práce. Jako autor uvedené bakalářské práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 16.3.2015

Poděkování

Rád bych touto cestou poděkoval vedoucímu bakalářské práce Ing. Martinu Havránkovi za cenné rady a odborné vedení, které mi poskytl při zpracování bakalářské práce.

Zálohování a archivace dat

Data backup and archivation

Souhrn

Bakalářská práce pojednává o možnostech způsobu zálohování a archivace dat.

První část je zaměřena na souhrn teoretických znalostí v procesech zálohy a archivace dat, dále je poukázáno na jejich klady, zápory a rozdíly mezi nimi. Také jsou zmíněny možnosti zpracování dat před samotnou zálohou či archivací.

Druhá část je zaměřena na analýzu způsobu zálohy a archivace dat ve firmě s odbornou pomocí IT technika. Poté jsou vypsány nedostatky v procesech zálohy a archivace dat a navržena jejich možná řešení.

Summary

Bachelor's work deals about possibilities how to backup and archive data.

The first part is focused on summary of theoretical knowledges in the processes of data backup and archiving. It is also pointed on its proc and cons and differences between them. There are also mentioned processing possibilities prior to the backup and archiving of data.

The second part is focused on the analysis of the ways how to backup and archive data in a company, with professional help of IT technician. Then there are listed deficiencies in the processes of data backup and archive and suggested its possible solutions.

Klíčová slova: zálohování, archivace, komprimace, šifrování, data, uložení, obnova, rizika, server

Keywords: backup, archivation, compression, encryption, data, repository, restoration, risks, server

Obsah

1 Úvod.....	9
2 Cíl práce a metodika	10
2.1 Cíl práce	10
2.2 Metodika	10
3 Teoretická východiska	11
3.1 Zálohování	11
3.2 Archivace	11
3.3 Plánování zálohy dat	12
3.4 Příčiny ztráty dat	12
3.5 Základní kritéria pro výběr datového nosiče.....	13
3.6 Úložná média	15
3.6.1 Optická média	15
3.6.1.1 CD a DVD.....	15
3.6.1.3 Blu-ray	15
3.6.1.4 DTD	16
3.6.1.5 Budoucnost optických médií.....	16
3.6.2 Pevné disky HDD.....	16
3.6.3 Pevné disky SSD	18
3.6.4 Flash disky	18
3.6.5 Datové pásky	19
3.6.6 Cloud.....	19
3.7 Architektury síťového ukládání dat.....	20
3.7.1 DAS – Direct Attached Storage	20
3.7.2 NAS – Network Attached Storage	21
3.7.3 SAN – Storage Area Networks	22
3.8 Metody zálohování.....	23
3.8.1 Úplná záloha (Full Backup)	23
3.8.2 Přírůstová záloha (Incremental Backup).....	24
3.8.3 Rozdílová záloha (Diferencial Backup).....	24
3.8.4 RAID pole	24
3.9 Zálohovací strategie	29
3.9.1 Round Robin	29

3.9.2 Grandfather-Father-Son	30
3.9.3 Hanojské věže	30
3.10 Zpracování dat před zálohou	31
4 Vlastní zpracování.....	33
4.1 Specifikace	33
4.1.1 Hardware specifikace	34
4.1.2 Software specifikace	35
4.2 Politika zálohování firemních dat	37
4.3 Politika archivace firemních dat.....	38
4.4 Nedostatky v zabezpečení dat a jejich navrhovaná řešení	39
4.4.1 Zálohování	39
4.4.2 Archivace	42
4.5 Vyhodnocení	43
5 Závěr	44
6 Seznam použitých zdrojů.....	45
7 Seznam obrázků a tabulek.....	48

1 Úvod

V dnešní době je velice snadné přijít o svá data, a proto je nutné je chránit. Existuje mnoho způsobů jak data zabezpečit ať už pomocí hardwaru, tak i softwaru. Mnoho uživatelů si ani neuvědomuje hodnotu svých dat, nechtějí připustit jejich možnou ztrátu a proto data ani nezalohují. Datová úložiště mohou být plná vzpomínek, fotografií a videozáznamů, zachycujících důležité okamžiky v životě. Zanedbávat zálohování a archivaci může mít za následek ztrátu těchto velice cenných vzpomínek. Uživatel se většinou o ochranu svých dat začne zajímat až v okamžiku jejich první ztráty.

Aby bylo možné data zálohovat a archivovat, je nutné si nejprve ujasnit, co je pro uživatele důležité. Na tuto otázku se však nedá konkrétně odpovědět. To, co jeden uživatel považuje za důležité informace, jiný považuje jako zcela nepotřebné. Obecně tedy platí, že uživatel má zálohovat ta data, o které on sám nechce přijít.

Nebezpečí ztráty dat není hlavním problémem pro běžné uživatele, ale především pro firmy, které většinou uchovávají velké množství důležitých dat. Zálohování a archivace dat je důležitou formou ochrany dat. V případě výpadku, či havárie zařízení, na kterém jsou data uložena, by bez správného zálohování byla data ztracena. Tento scénář by mohl mít katastrofální následky a mohl by skončit až zkrachováním podniku.

2 Cíl práce a metodika

2.1 Cíl práce

Hlavním cílem je návrh efektivního řešení zálohy a archivace dat.

Dílčím cílem je popsat různé možnosti a způsoby pro zálohu a archivaci dat, poukázat na jejich rozdíly, klady a zápory.

Praktická část bude soustředěna na zálohování a archivaci dat v konkrétní firmě. Bude provedena analýza způsobů zálohy a archivace dat v této firmě.

2.2 Metodika

Metodika bakalářské práce je založena na shromáždění a analyzování odborných informací týkající se daného tématu. Praktická část se zabývá analýzou metod zálohování a archivace dat ve firmě. IT technik této firmy poskytne informace o aktuálních postupech zálohy a archivace dat. Na základě této analýzy bude navrženo efektivnější řešení.

3 Teoretická východiska

3.1 Zálohování

„Zálohování je mechanismus, při kterém jsou vybraná data (nemusí to být tedy všechna) ukládána na jiné médium. V případě zničení původního média jsou data obnovena ze zálohy.“^[1]

Záložní kopie dat se vytváří dle nastavené firemní politiky a pro každou organizaci existují jiná pravidla. Většina zálohovaných dat je především určena jen např. pro obnovení operačního systému, pro obnovu databáze nebo pro obnovu nedávno vytvořených dokumentů apod., není tedy určena pro soubory, které nemají nic společného se samotným chodem firmy. Zálohování dat se provádí z důvodu jejich ochrany před ztrátou, za účelem jejich obnovy po neočekávaně nastalé havárii. Pokud se médium, na které se data primárně ukládají, jakýmkoliv způsobem poškodí nebo dokonce zničí, budou data, která jsou na něm uložena nenávratně ztracena. Díky záloze, která se provádí na jiné médium než to poškozené, se může tento negativní jev omezit na naprosté minimum. Záloha dat se většinou provádí na dobu několika dnů, případně měsíců, aby se tak zajistila aktuálnost uložených dat. K zálohování je potřeba vlastnit výkonné záznamové zařízení, které musí mít rychlou odezvu a zajištěnou stálou konektivitu. Pro uložení dat na časový úsek, jako jsou měsíce nebo roky, se využívá archivace dat.[3]

3.2 Archivace

„Archivace dat znamená dlouhodobé uložení důležitých dat na záznamové médium.“^[2]

Smyslem archivace je uchování a zajištění dlouhodobé dostupnosti digitálních dat v časovém horizontu od několika měsíců až po desítky let. Data se ukládají na nosiče s dlouhou životností a dobrým bezztrátovým uchováním informace. Takovými nosiči bývají nejčastěji magnetické pásky a vybrané pevné disky. Archivace dat je znatelně levnější než zálohování, protože u zálohovacích médií jsou vývojáři nuceni držet krok s technologickým vývojem v oblasti rychlosti přístupu k datům a velikostí úložného prostoru, zatímco pro účely archivace postačí vlastnit starší a pomalejší úložná média.

[4]

3.3 Plánování zálohy dat

Před začátkem zálohování a archivace dat, je nutno vše naplánovat. Musí se určit, jaká data zálohovat, protože důležitým faktorem zálohování a obnovy dat, je čas. Není možné data používat a zároveň zálohovat, protože žádný systém nepovolí kopírování dat, se kterými se aktuálně pracuje, proto se zálohuje mimo pracovní dobu. Pro zálohu se volí data, která jsou důležitá a nenahraditelná.[5]

Důležitá data, která je potřeba zálohovat, se mohou nacházet jak na hlavním serveru, tak i na jednotlivých pracovních stanicích uživatelů. Je nutné určit, jestli se budou zálohovat jen vybrané adresáře nebo celá pracovní stanice. Je potřeba zálohovat jen data, které není možno obnovit z jiného zdroje. Na rozdíl od aplikací uživatele, které je možno snadno obnovit z jiných zdrojů, jsou data ztracena. [6]

Množství dat, které bude potřeba zálohovat je velmi důležitým aspektem pro výběr vhodného hardwaru. U menších společností je možné množství dat dopočítat na základě detailního průzkumu záložních klientů. Záložní klient je každá pracovní stanice a server určený k zálohování. Ve středních a velkých společnostech se určuje množství objemu dat odhadem, pomocí menšího vzorku záložních klientů. [7]

Kvalitní zálohování závisí na vhodně zvolené periodě zálohování. Existuje několik typů záloh. Nejčastěji používané typy jsou úplná, inkrementální a diferenciální záloha. Jednotlivé typy jsou podrobně popsány v kapitole 3.8.

Proces zálohování dat musí vždy obsluhovat člověk. V případě malého podniku to mohou být všichni uživatelé pracovních stanic. Ve velkých společnostech má tento proces na starost zodpovědná osoba, nejčastěji IT administrátor nebo dokonce celý tým. Ve velkých společnostech, ale není záloha dat z uživatelů zcela sňata. Každý uživatel musí pravidelně ukládat svá data přesně podle předpisů společnosti. [8]

3.4 Příčiny ztráty dat

Soubory mohou být nenávratně ztraceny z mnoha důvodů. V některých případech hraje roli při ztrátě dat člověk, jiným se nedá zabránit.

Nejčastější příčinou jsou hardwarové a systémové chyby. Bez ohledu na to, zda jsou data uložena na disku, optickém médiu nebo flash disku, dané zařízení může selhat a data jsou s největší pravděpodobností ztracena. V současné době existují společnosti, které zvládnou data z pevného disku zachránit, ale tato operace je finančně nákladná a s nejistým výsledkem.

Další příčinou mohou být počítačové viry. Žádný antivirový program není bezchybný. Antivirové centrum se dlouhodobě zabývá testováním antivirových programů. Nejúspěšnějším antivirem pro rok 2015 je NOD 32 od společnosti ESET s 97,8% úspěšnosti.[9]

Ať už se jedná o lidskou chybu nebo přírodní katastrofu, je důležité data vhodným způsobem zálohovat.

3.5 Základní kritéria pro výběr datového nosiče

Výběr správné technologie pro zálohování a archivaci dat je základní a důležitý krok. Při výběru je nutno zvážit mnoho okolností, aby vybraná technologie vyhovovala potřebám daného subjektu a data byla spolehlivě chráněna. Je třeba si stanovit základní kritéria pro výběr vhodného datového nosiče.

Úložná kapacita

Úložná kapacita je velmi důležité kritérium. Ideální je volit technologii s vyšší úložnou kapacitou, než mají aktuální data. V budoucnu je pravděpodobný nárůst dat. Pro možnost automatické zálohy dat je potřeba ukládat jen na jeden typ média. Pokud systém nedisponuje technologií automatické výměny, je nutná asistence člověka.

Životnost

Životnost úložného média je důležitým faktorem zejména pro archivaci dat, kde je potřeba dlouhodobé uchování.

Odolnost

Zde se hodnotí odolnost úložného média proti vnějším vlivům. Prach, slunce, poškrábání a podobně.

Náklady

Důležité kritérium při výběru vhodného média jsou pořizovací náklady. Většinou je udáváno jako cena za GB úložného prostoru.

Metoda přístupu k datům

Náhodná metoda umožňuje téměř okamžitý přístup k datům uloženým v jakékoliv části média.

Sekvenční metoda je typická pro pásková média. Přístup k datům probíhá od začátku neboli od poslední polohy čtecí hlavy.

Rychlost zápisu

Rychlost zápisu udává objem dat, která mohou být zálohována za určitý časový okamžik. Na rychlost nemá vliv jen dané médium, ale i rychlost mechaniky, která provádí zápis. Rychlost je většinou uvedena v MB/s.

Rychlost čtení

Důležité kritérium pro obnovu dat. Čím vyšší rychlost čtení, tím je rychlejší požadovaná obnova dat.

Možnost automatizace

Zálohovací systémy mohou být vybaveny možností automatizace neboli výměny jednotlivých médií. Efektivní záloha dat, která nevyžaduje asistenci člověka.

3.6 Úložná média

3.6.1 Optická média

Hlavní výhodou v zálohování a archivaci dat u optických médií je pořizovací cena a snadná dostupnost médií. Jejich velkou nevýhodou je náchylnost na poškození, nízká životnost a menší kapacita. V dnešní době se na optická média většinou nezalohuje. Začínají se spíše využívat polovodičové flash paměti a online zálohy.

3.6.1.1 CD a DVD

U CD je ukládání dat ve stopách na jedné dlouhé spirále, která začíná ve středu optického disku. Délka celé spirály je zhruba 6 km a má konstantní hustotu dat. Kapacita CD je 700 MB. Rychlost zápisu 8 MB/s a přístupová doba 120 ms. [10]

U DVD optických médií se data ukládají pod povrch do jedné nebo dvou vrstev ve formě spirály. Kapacita DVD záleží na počtu vrstev a stop. Jedna vrstva má 4,7 GB, které představuje minimum u DVD disků. Kapacita u dvouvrstvého zápisu na obě strany disku je 17,1 GB. Rychlost zápisu 22 MB/s a přístupová doba 120 ms. [11]

Optické disky CD a DVD jsou určeny jen pro jednorázové zálohování nebo archivaci. Média CD-RW a DVD-RW už umožňují opakovatelný zápis a lze je využít pro periodické zálohy.

Nevýhoda CD a DVD je ve velmi nízké životnosti. Výrobci udávají životnost v ideálních podmínkách kolem 100 let, ale tato doba je výrazně zkráceně mnoha aspekty. Použitím média, vystavení prachu a slunečnímu světlu. Skutečná životnost se nejčastěji pohybuje od 5 do 15 let.

3.6.1.3 Blu-ray

Blu-ray je třetí generace optických disků určených pro ukládání digitálních dat. Pojmenováno podle barvy laserového paprsku, který je modrofialový. V České

Republice jsou k dispozici jednovrstvá blu-ray s kapacitou 25 GB a dvouvrstvá s kapacitou 50 GB. Rychlost zápisu 18 MB/s a přístupová doba 180 ms.[12]

3.6.1.4 DTD

Optické médium se zaměřením na archivaci dat. Odlišuje se od ostatních optických disků v záznamové vrstvě, která lépe chrání před vlivy prostředí a nedochází tak přirozenému stárnutí záznamové vrstvy, jako u předešlých optických disků. Kapacita 4,7 GB. Přístupová doba 150 ms. [13]

3.6.1.5 Budoucnost optických médií

Optická média jsou v dnešní době pomalu na ústupu a to z důvodu možného mechanického poškození, nečitelnosti dat při vystavení světelnému záření a v neposlední řadě stále se zvyšujícím požadavkům na kapacitu úložného média. Nicméně společnost Fujifilm vyvinula médium s kapacitou 1 TB a v roce 2015 počítá s kapacitou až 15 TB. Jejich výhodou bude rychlejší přístupová doba k datům, nižší pořizovací cena než například u magnetických pásek a samozřejmě obrovská kapacita. [14]

3.6.2 Pevné disky HDD

Pevné disky poskytují největší komfort pro potřeby zálohování. Díky nízké ceně a vysoké kapacitě jsou pevné disky HDD velmi rozšířené. Nevýhodou je mechanická konstrukce, která je náchylnější na poškození, proto je možné častější selhání pevných disků. Ztrátě dat z důvodu mechanického selhání se může předejít zapojením více disků do pole RAID.

Moderní pevné disky dosahují kapacity 6 TB. Rychlost otáček je různá, ale mezi nejrozšířenější patří 5400 ot/min a 7200 ot/min.



Obrázek 1 Interní pevný disk HDD [24]

Pevný disk nemusí být jen v počítači. Pro zálohování dat je možné využít i externí disky, které jsou k počítači připojeny pomocí USB. Externí disky mohou být opatřeny například číselným PIN zabezpečením.



Obrázek 2 Externí pevný disk HDD s PIN zabezpečením [25]

3.6.3 Pevné disky SSD

SSD je zkratka pro Solid State Drive neboli polovodičový disk. Jedná se o alternativu klasických HDD pevných disků. SSD disk je založen na soustavě energeticky nezávislých flash pamětí, které jsou osazeny na destičce tištěného spoje. Hlavním rozdílem mezi HDD a SSD je absence mechanických součástí u SSD, které jsou složeny pouze z elektronických součástek. Při práci spotřebují méně energie a dosahují vysokých rychlostí díky velmi nízkým přístupovým dobám. Jejich velkou nevýhodou je nižší kapacita než HDD a vysoká cena. Problémem může být také nižší životnost. Jsou limitovány maximálním počtem zápisů do stejného místa až na 100 000 zápisů. [15]

Moderní SSD dosahují kapacity 1 TB. Rychlost čtení 550 MB/s, rychlost zápisu 500 MB/s.



Obrázek 3 SSD disk [14]

3.6.4 Flash disky

Flash disk je paměťové zařízení vybavené pamětí typu flash, která umožňuje uchování dat i po odpojení napájení. Data se do flash disku nahrávají přes sběrnici USB. Jedná se o kompaktní přenosová média, která disponují vysokou kapacitou a jsou energeticky nenáročná.

Moderní flash disky s rozhraním USB 3.0 v České Republice dosahují kapacity 1 TB. Rychlost čtení 240 MB/s a rychlost zápisu 160 MB/s. Průměrná životnost 10 let. [16]



Obrázek 4 Flash disk [16]

3.6.5 Datové pásky

Datové pásky patří již dlouhou dobu mezi nepoužívanější médium pro zálohu i archivaci dat. Od doby jejich vzniku prošly řadou inovací, díky kterým jsou spolehlivější, rychlejší a mají větší kapacitu. Jejich největší výhodou je možnost zálohy, respektive archivace velkého množství dat za nízkou cenu, a možnost automatizace výměny pásky pomocí páskového robota. Nevýhodou je vysoká pořizovací cena páskové knihovny, sekvenční přístup k datům a velká citlivost na magnetické pole. Dalším problémem může být zpětná kompatibilita páskových mechanik. [6]

V závislosti na formátu datové pásky se kapacita nejdostupnějších médií pohybuje mezi 1 - 3 TB, přenosová rychlost 24 -250 MB/s, doba přístupu 45-72s, životnost 30 let

Vývoj datových pásek neustále pokračuje a v roce 2014 společnost Sony vyvinula datovou pásku o kapacitě 185 TB.

3.6.6 Cloud

Cloud je online úložiště dat. Jedná se o datová centra, která umožňují pronájem datového prostoru, kde se dají data ukládat přes internet. Primárním účelem této služby nebývá zálohování dat, ale spíše jejich dostupnost v jakýkoliv okamžik. Data jsou

šifrována, ale klíč má i provozovatel, proto je vhodné zde ukládat již šifrovaná data.
[17]

Na rozdíl od zahraničí v České Republice není tato služba společnostmi moc využívána, ale její popularita roste každým rokem. Společnosti se nejvíce obávají odcizení dat, přestože datová centra mají vysoké zabezpečení, většinou silnější než mají samotné společnosti.

Nejznámějšími službami jsou například Dropbox, IDrive, SkyDrive nebo Google Drive.

3.7 Architektury síťového ukládání dat

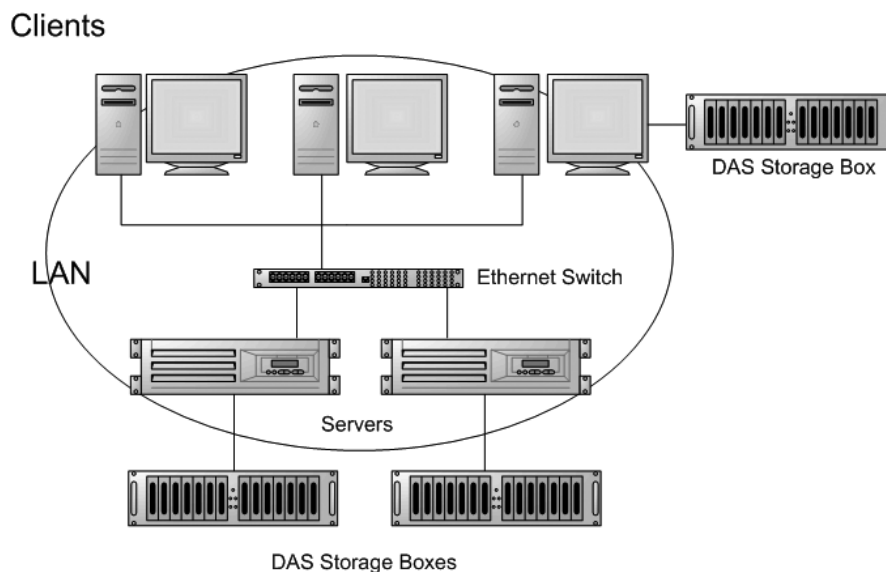
Rozlišují se tři základní architektury řešení síťového ukládání dat. Architektury se liší dle způsobu připojení úložných zařízení na podnikovou síť, která většinou odpovídá velikosti podniku, respektive podnikové síti a jeho nárokům na úložná řešení.

3.7.1 DAS – Direct Attached Storage

Jedná se o úložná zařízení přímo připojená k serveru. Přesněji se jedná o připojení jedno a více disků k serveru, serverům nebo samostatnému počítači pro ukládání dat. Architektura DAS poskytuje celou škálu možností řízení dat. Řešení může být velmi jednoduché, například připojení druhého pevného disku do stolního počítače. Přidaný disk se připojuje na server. [18] [19]

U architektury DAS se využívají rozhraní EIDE (ATA), USB a SCSI.

Direct Attached Storage



Obrázek 5 Direct Attached Storage [18]

3.7.2 NAS – Network Attached Storage

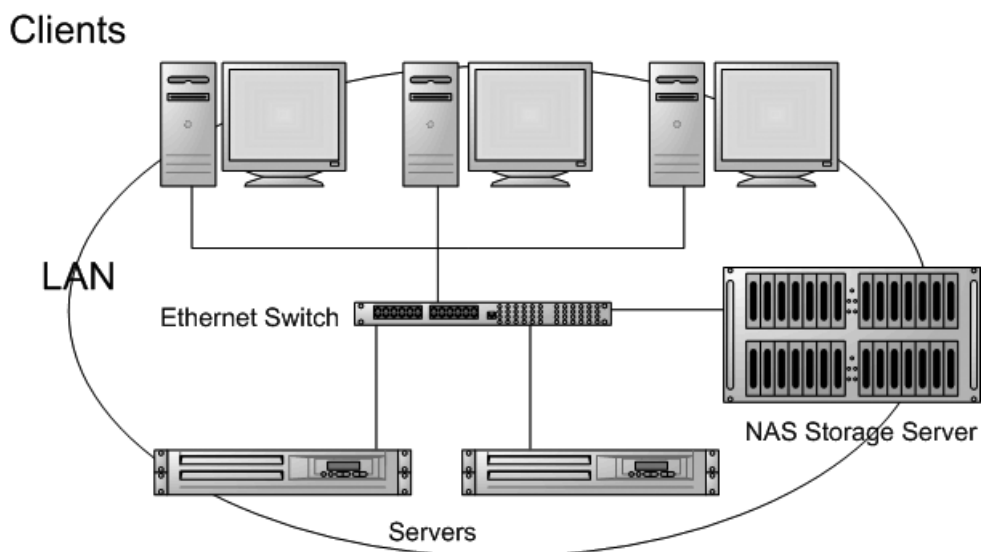
NAS je datové úložiště připojené k místní síti LAN a představuje kompletní úložný systém.

NAS je složený z jednoho či více pevných disků, které je možno slučovat do větších datových struktur. Jedná se o specializované servery, které jsou určeny pro připojení úložných zařízení k síti. Jedná se například o páskové jednotky nebo diskové soubory.

Výhodou NAS je nižší pořizovací cena, jednodušší implementace a management než u architektury SAN.

NAS servery podporují primárně NFS (Network File System) a CIFS (Common Internet File System) protokoly. Mezi další podporované protokoly například patří FTP (File Transfer Protokol) nebo SCP (Secure CoPy). NAS server se k síti připojí pomocí Ethernet a podporuje různé protokoly pro přenos a různé operační systémy. [18] [19]

Network Attached Storage



Obrázek 6 Network Attached Storage [18]

3.7.3 SAN – Storage Area Networks

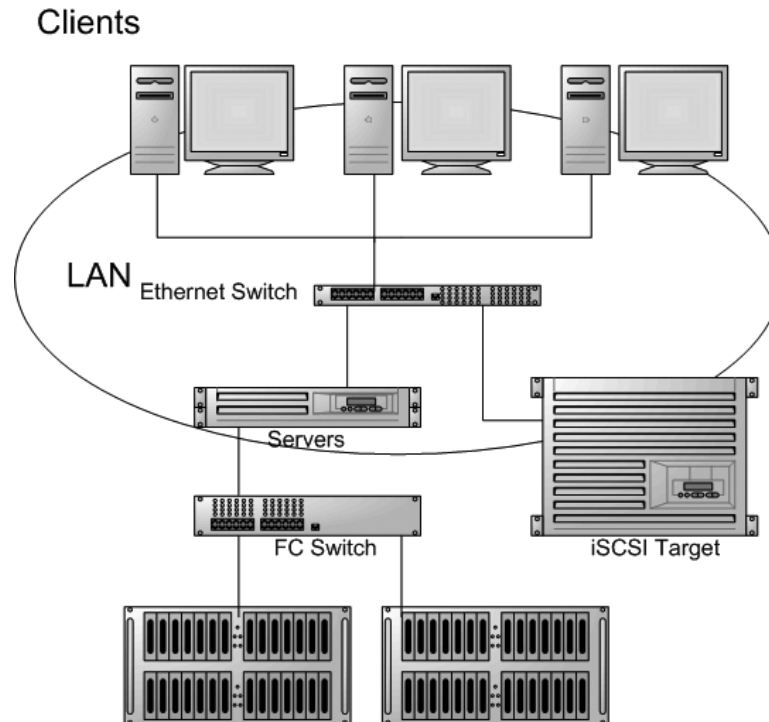
Jedná se o síť nezávislou na podnikovém intranetu, určenou k propojení úložných zařízení. SAN je optimalizovaná na ukládání, sběr a výběr bloků. Využívá technologie Fibre channel (komunikační rozhraní používané pro vysokorychlostní přenos dat v síti) nebo řešení orientovaná na IP (FCIP, iSCSI, IFCP) protokol. Standardním řešením pro SAN je síť obsahující minimálně 2 servery s přístupem k sadě disků, propojeno minimálně jedním Fibre Channel switchem. Pro větší bezpečnost se využívají switche většinou dva. SAN umožňuje geografické rozdělení úložných systémů skrze Fiber Channel konektivitu.

Na rozdíl od řešení NAS, řešení SAN prezentuje celou síť, která propojuje veškerá zařízení, která jsou určena k ukládání dat na servery. Výhodou SAN je vysoká výkonost, spolehlivost, možnost řízení a možnost téměř nekonečného růstu, což je ideální do podnikových sítí. [18] [19]

Nevýhodou SAN řešení je cena. Vysoká cena je zejména kvůli využití Fibre Channel technologií.

SAN a NAS se nenahrazují. Obě řešení se mohou vzájemně doplňovat.

Storage Area Network



Obrázek 7 Storage Area Network [18]

3.8 Metody zálohování

Existuje několik metod pro zálohování dat, proto se musí zvolit správná. Jedním z hlavních kritérií je velikost dat. V případě menšího množství dat lze použít úplnou zálohu a v případě většího množství může být vhodnější využít přírůstkovou nebo rozdílovou zálohu dat.

3.8.1 Úplná záloha (Full Backup)

Metoda úplného zálohování dat je základní metodou zálohování a je výchozí metodou pro všechny následující. Při úplném zálohování dat se ukládají kopie všech dat určených pro zálohování. Úplná záloha je ideální metodou, protože je nejvíce komplexní a soběstačná.

Úplná záloha zabere mnoho času, proto se často nevyužívá a je omezena na určitou periodu. Většinou v řádu týdnů a měsíců. Úplná záloha obsahuje kompletní data, proto nám poskytuje možnost rychle obnovit všechny zálohované soubory.

Výhodou je, že všechna data jsou uložena na jednom místě v jediném souboru a je jednodušší správa úložného média.

Nevýhodou je pomalé zálohování, vysoké požadavky na úložný prostor a bezpečnost firmy v případě ztráty nebo odcizení úložného média. [20]

3.8.2 Přírůstková záloha (Incremental Backup)

Přírůstková záloha je záloha dat, při které se ukládají pouze kopie souborů, které přibyly nebo byly změněny od poslední provedené zálohy. To znamená, že na začátku procesu se vytvoří úplná záloha a každá další ukládá jen změny proti záloze předešlé.

Přírůstkové zálohování poskytuje rychlejší způsob zálohování dat než opakování úplné zálohy. Výhodou je úspora času a úložného prostoru potřebného pro tuto metodu zálohování.

Hlavní nevýhodou této metody je, že pokud dojde ke ztrátě nebo poškození jakékoliv části dat v jednom z přírůstků, nelze obnovit data ani z dalších přírůstků. Proto je potřeba po určitém čase znovu využít úplné zálohy dat. [20]

3.8.3 Rozdílová záloha (Differential Backup)

Rozdílová záloha obsahuje všechny soubory, které se změnily od poslední úplné zálohy. Pokud dojde k poškození jedné rozdílové zálohy, nemá to vliv na žádnou jinou rozdílovou zálohu. Výhodou je kratší čas obnovení zálohy než u zálohy úplné.

Nevýhodou jsou vysoké požadavky na úložný prostor. V případě časté rozdílové zálohy, může její velikost být dokonce větší než velikost úplné zálohy dat. [20]

3.8.4 RAID pole

RAID (Redundant Array of Independent / Inexpensive Disks) je metoda postavená na seskupování více pevných disků do jedné logické jednotky. Tato logická jednotka se nazývá pole. RAID pole v operačním systému vypadá jako jeden souvislý disk. Mezi

jeho hlavní výhody patří vysoká spolehlivost a rychlost přístupu k datům. Moderní RAID pole v dnešní době využívají technologie HOT-SWAP, která umožňuje výměnu pevných disků za běhu systému. Nevýhodou je uložení originálních a zálohovaných dat na stejném místě.

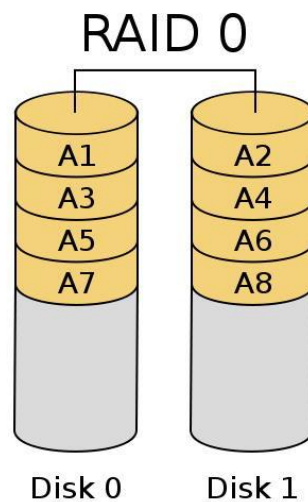
Existuje několik typů RAID polí. Mezi základní patří pole RAID 0, 1, 2, 3, 4, 5, 6. Dále existují nestandardní typy, které vzniknou kombinací jednotlivých RAID polí. Například RAID 0+1, 10, 50, 100 a další. V praxi nejpoužívanější jsou RAID 0, 1, 5, 6 a kombinace RAID 0+1, 10. [21]

RAID 0

Tento typ pole RAID využívá dvě metody ukládání dat na disky. Jedná se o metodu zřetězení a prokládání.

Metoda zřetězení ukládá data nejdříve na první disk v poli a po jeho zaplnění se data začnou ukládat na druhý disk a tento proces se opakuje při zaplnění každého dalšího disku. Tato metoda využívá maximální kapacitu všech disků a celková kapacita se rovná součtu kapacit jednotlivých disků. Hlavní nevýhodou je ztráta možnosti plnohodnotné obnovy dat při poruše na jednom z disků.

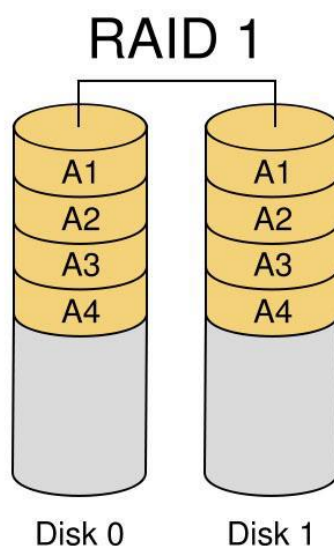
Metoda prokládání, ukládá data na disky cyklicky. Jednotlivé soubory jsou rozděleny na bloky a ty se ukládají střídavě na všechny disky. Díky ukládání na více disků najednou je zápis i čtení rychlejší. V případě poruchy jednoho disku je velké riziko ztráty plnohodnotné obnovy dat. [21]



Obrázek 8 Raid 0 [20]

RAID 1

Metoda RAID 1 je metoda zrcadlení. Jedná se o jednoduchý způsob zálohy dat. Všechna data jsou současně zapisována na dva disky. V případě poruchy jednoho disku, existuje kopie na jiném disku. Výhodou je rychlost čtení a nízká odezva. Nevýhodou je pomalý zápis a potřeba dvojnásobné kapacity než u RAID 0. [21]

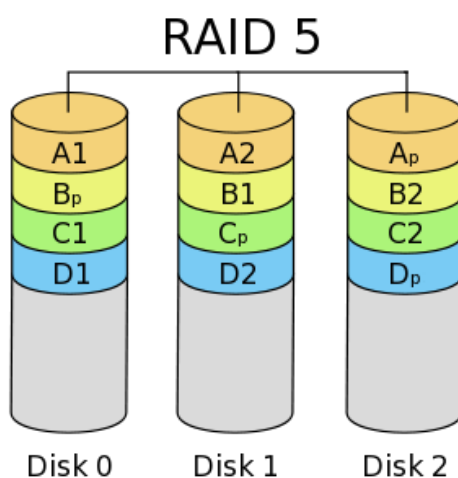


Obrázek 9 Raid 1 [20]

RAID 5

RAID 5 využívá metodu prokládání dat a metodu parity. Jedná se o zapojení více disků, na která jsou data ukládána po blocích. Při ukládání dat jsou zároveň vytvářena data paritní. Paritní data jsou ukládána střídavě na všechny disky. Minimální počet disků pro RAID 5 je roven třem. [21]

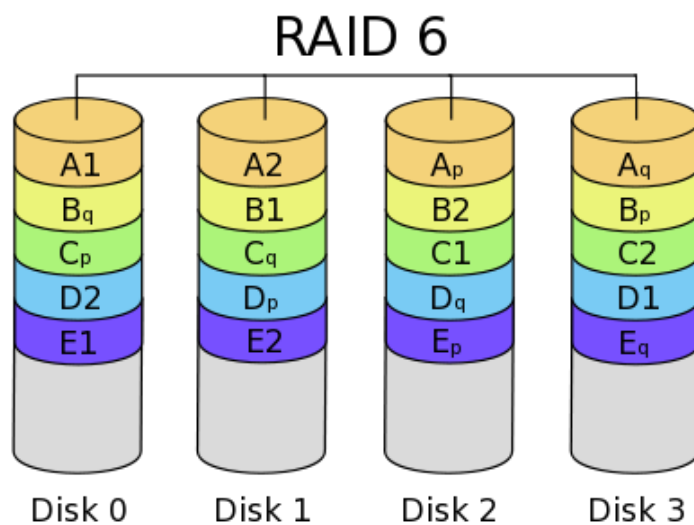
Hlavní výhodou jsou redundantní informace uložené na různých discích. Díky metodě prokládání je rychlé čtení dat z disků, ale kvůli využití parity je zde pomalý zápis.



Obrázek 10 Raid 5 [20]

RAID 6

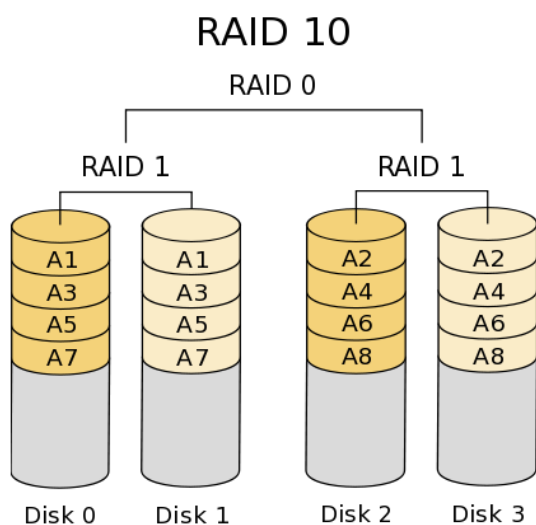
RAID 6 na rozdíl od RAID 5 vytváří paritní kopii na dvou discích. Využití více paritních disků, dodává poli větší odolnost proti ztrátě dat. Rychlost čtení je stejná jako u RAID 5, ale kvůli dvojité paritě má pomalejší zápis dat. RAID 6 potřebuje minimálně 4 disky.



Obrázek 11 Raid 6 [20]

RAID 10

V praxi často používaná metoda je kombinace RAID 1 a RAID 0. Jedná se o řešení spojením několika diskových polí RAID. U dat se nejdříve provede zrcadlení jako u RAID 1 a poté se vloží do diskového pole typu RAID 0. Tato kombinace zajišťuje vysokou přenosovou rychlost dat oproti vytvoření parity disku.



Obrázek 12 Raid 10 [20]

3.9 Zálohovací strategie

Zálohování dat s použitím zálohovacích médií vyžaduje určitou strategii. Jednotlivá záložní média se mohou po provedení zálohy archivovat a vytvořit kartotéku záloh z libovolného data. Tato metoda není z ekonomického hlediska zcela ideální a vyžaduje velký prostor na uskladnění neustále se zvyšujícího množství zálohovacích médií. Vhodnější variantou je zvolení strategie rotace záložních médií. Při rotaci dochází k postupnému střídání a rovnoměrnému opotřebení všech záložních médií. Jednotlivé strategie se liší celkovým počtem potřebných médií, jejich rozdělením do záložních sad a délkou historie uložených dat.[22]

3.9.1 Round Robin

Round Robin je asi nejpoužívanější strategie zálohování dat. Dochází k jednoduché rotaci sad záložních médií. Každá sada je určena pro určitou časovou periodu a při každé další periodě jsou data uložená na zálohovacím médiu přepsána. Například první sada je určena na první týden, druhá na druhý týden atd. a každý měsíc se tyto média přepisují. Díky této jednoduché strategii dochází k rovnoměrnému opotřebení jednotlivých sad záložních médií. Rychlost opotřebení médií záleží na délce jednotlivých period. Při krátkých periodách dochází k rychlejšímu opotřebení a je vhodné zálohovací média častěji nahrazovat za nové, při delších periodách bude opotřebení záložních médií pomalejší. V případě havárie jsou k dispozici záložní kopie po dobu délky jednotlivé periody záloh.[20]

<i>Media set</i>	<i>Pondělí</i>	<i>Úterý</i>	<i>Středa</i>	<i>Čtvrtek</i>	<i>Pátek</i>	<i>Sobota</i>	<i>Neděle</i>
<i>A</i>	X					-	-
<i>B</i>		X				-	-
<i>C</i>			X			-	-
<i>D</i>				X		-	-
<i>E</i>					X	-	-

Tabulka 1 Round Robin [20]

3.9.2 Grandfather-Father-Son

Grandfather-Father-Son je strategie, která využívá tři sad záložních médií. Sada Son pokrývá jednotlivé dny v týdnu. Na konci každého týdne se provede týdenní záloha sady Father, která obsahuje média pro jednotlivé týdny a na konci měsíce se provede záloha na sadu zálohovacích médií Grandfather. U zálohovacích médií typy Grandfather se obvykle nastavuje z důvodu bezpečnosti doba expirace dat. To znamená, že data během této doby nemohou být přepsána. Strategie Grandfather-Father-Son poskytuje zálohy za jednotlivé dny v týdnu, týdenní zálohy za poslední měsíc a měsíční zálohy za poslední rok.[20]

	<i>Pondělí</i>	<i>Úterý</i>	<i>Středa</i>	<i>Čtvrtek</i>	<i>Pátek</i>	<i>Sobota</i>	<i>Neděle</i>
<i>Týden 1</i>	Son 1a	Son 1b	Son 1c	Son 1d	Father 1	X	X
<i>Týden 2</i>	Son 1a	Son 1b	Son 1c	Son 1d	Father 2	X	X
<i>Týden 3</i>	Son 1a	Son 1b	Son 1c	Son 1d	Father 3	X	X
<i>Týden 4</i>	Son 1a	Son 1b	Son 1c	Son 1d	Grandfather	X	X

Tabulka 2 Grandfather-Father-Son [20]

3.9.3 Hanojské věže

Strategie zálohování dat Hanojské věže má původ v logické čínské hře. Strategie Hanojské věže využívá pět sad záložních médií. Na záložní médium z první sady se provádí záloha každý druhý den. Na záložní médium z druhé řady se provádí záloha každý čtvrtý den. Na záložní médium ze třetí sady každý osmý den. Záložní média ze čtvrté a páté sady jsou používána střídavě každý šestnáctý den. Tato strategie umožní získání větší historie zálohovaných dat použitím nové sady médií. Častěji používané sady zálohovacích médií obsahují novější kopie dat a méně používané sady záložních médií obsahují starší verze souborů. Vzhledem k náročnosti na obsluhu této strategie je lépe využitelná při použití automatizované knihovny médií.[20]

Den	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Media set	A		A		A		A		A		A		A		A	
		B				B				B				B		
				C								C				
								D								
																E

Návrat k prvnímu dni

Tabulka 3 Hanojské věže [20]

3.10 Zpracování dat před zálohou

S digitálními daty se mohou provádět různé úpravy, které vedou k zamezení nebo v některých případech dokonce podpoření jejich kopírování a mazání.

Duplikace dat

Principem duplikace dat je vytvoření dvou záloh na dvou různých médiích a na různých místech. Zvyšuje se tím zabezpečení dat proti jejich ztrátě a zvyšuje se rychlost při jejich obnovení po katastrofě. Duplikace sebou nese nevýhodu v potřebě vlastnit mnoho úložného prostoru. Pro odstranění nadbytečných dat se může využít metoda zvaná deduplikace.

Deduplikace dat

Deduplikace dat je metoda, která snižuje požadavky kladené na uložení odstraněním redundantních dat, takže určitý datový obsah zabírá prostor na disku pouze jednou. Pokud jsou nalezena duplicitní data, nahradí je ukazatel na již uložený soubor údajů a vlastní data jsou zahozena, neboli deduplikována. Deduplikaci dat provádíme na úrovni bloků, nikoli na úrovni souborů. Deduplikace zmenšuje objem ukládaných dat až

o 95%. Jedná se o poměrně novou technologii, která v budoucnu bude velmi potřeba, protože objem ukládaných dat ročně roste téměř exponenciálně.

Redundance dat

Redundance dat je označení pro nadbytečný výskyt stejných hodnot v celém objemu dat. Výhodou redundance je to, že může sloužit jako prostředek ke zvyšování spolehlivosti a odolnosti proti chybám. Její velkou nevýhodou je nadbytečnost dat, které zvyšují potřebu velké úložné kapacity.

Komprese dat

Komprese dat je metoda, která je založena na odstraňování nadbytečnosti. Tato operace je velmi důležitá, zejména s ohledem na velikost datových médií. Při kompresi dat dojde ke zmenšení jejich velikosti.

Proces komprese umožňuje na datový nosič uložit více informací, díky zmenšení velikosti dat, dále šetří čas při kopírování a přenosu dat přes počítačovou síť. Pomocí komprese lze zálohovaná či archivační data setřídít podle druhu, například grafika, dokumenty za určité období, a tím je více zpřehlednit. Velkou výhodou je možnost rozdělení výsledného archivu na několik částí, u kterých lze předefinovat velikost. Zejména vhodné pro datové nosiče s menší kapacitou. [23]

Šifrování dat

Před zálohou či archivací je vhodné data zašifrovat a tím zabránit jejich možnému zneužití. Šifrování patří mezi nejjednodušší, ale zároveň nejúčinnější nástroj pro ochranu dat. Šifrování je zvláště důležité při přenosu dat pomocí počítačové sítě, protože přenos může být odposloucháván.

Jedná se o převod dat do nečitelné podoby pomocí šifrovacího algoritmu a šifrovacího klíče. Šifrovací algoritmy jsou zpravidla veřejnosti známé, proto je důležité zvolit vhodný šifrovací klíč, pomocí kterého jsou data šifrována a bez něhož nelze data zpět dešifrovat.

4 Vlastní zpracování

Praktická část bakalářské práce je zaměřena na způsob zálohování a archivace dat ve firmě, která se zaměřuje na poskytování komplexních řešení pro efektivní řízení organizací veřejné správy. Odborné podklady poskytl IT technik analyzované společnosti Otto Bartošek. Pokud budou během analýzy zjištěny nedostatky v procesech zálohy a archivace dat, budou navrženy způsoby jejich vyřešení.

4.1 Specifikace

Analyzovaná firma se zaměřuje na poskytování komplexních řešení pro efektivní řízení organizací veřejné správy. Jedná se o pomoc jednotlivým veřejnoprávním organizacím a jejich skupinám (veřejnoprávním korporacím) zlepšit řízení, dosáhnout úspor, eliminovat rizika a zvýšit kvalitu služeb.

Poskytují takzvané komplexní řešení, což je široké spektrum služeb:

- Pre-audit
- Audit
- Analýza a návrh opatření
- Studie proveditelnosti
- Nastavení vnitřního prostředí organizace
- Softwarová podpora
- Konzultace a metodické vedení
- Školení
- Dlouhodobá podpora

Vzhledem k širokému spektru poskytovaných služeb, pracuje analyzovaná firma s velkým množstvím dat, která je potřeba ukládat, zálohovat a archivovat pro případnou ztrátu či pozdější použití.

Analyzovaná firma svoje data ukládá na dva servery, z nichž každý disponuje čtyřmi sloty pro pevné disky. Pro potřeby zálohování je aktuálně využíváno šest pevných disků. Servery pracují na operačním systému Ubuntu 14.04.1 LTS server. Pro potřeby zálohování se využívá softwarový program Bacula verze 5.2.6. Pro některé

systemy se kromě programu Bacula využívají i jiné programy. Jedná se o Windows Backup, SQL Backup a zálohování v TFS/SP2010.

4.1.1 Hardware specifikace

Analyzovaná firma využívá pro potřeby zálohování dva servery. V serverech je celkem zapojeno 6 pevných disků. Každý disk má určeno, k čemu slouží, mají různé rychlosti otáček a kapacitu. Většina disků bývá zaplněna zhruba z 40% jejich kapacity.

Parametry Server 1

- Výrobce: Hewlett Packard
- Procesor: Intel Pentium Dual-Core G6950, 2,80 GHz
- Operační paměť: 2 GB RAM DDR3 1333 MHz
- Operační systém: Ubuntu 14.04.1 LTS server
- Počet slotů pro pevné disky: 4
- Počet zapojených disků: 4
- Zálohovací software: Bacula 5.2.6
- Do sítě připojen přes přepěťovou ochranu
- Funkce: DHCP, DNS

Parametry Server 2

- Výrobce: Vlastní výroba
- Procesor: Core 2 Duo E4600 2,40 GHz
- Operační paměť: 512 MB RAM
- Operační systém: Ubuntu 14.04.1 LTS server
- Počet slotů pro pevné disky: 4
- Počet zapojených disků: 2
- Zálohovací software: Bacula 5.2.6
- Do sítě připojen přes přepěťovou ochranu
- Funkce: DHCP, DNS

DHCP

DHCP, neboli Dynamic Host Configuration Protocol, je síťový protokol založený na síťové architektuře klient – server, zajišťující přidělování pracovním stanicím IP adresu, masku, bránu a další parametry, umožňující jejich připojení do sítě.

DNS

Počítače mezi sebou komunikují přes IP adresy. Jelikož jsou IP pro člověka špatně zapamatovatelné, byl zpracován hierarchický systém doménových jmen neboli Domain Name System, který překládá IP adresy na doménová jména. Uživatelům tedy stačí si zapamatovat doménová jména serverů, což je mnohem snadnější.

Parametry pevných disků

Pevné disky			
Počet	2	3	1
Výrobce	Seagate	Western Digital	Western Digital
Kapacita	500 GB	2TB	1TB
Rozhraní	SATA II	SATA II	SATA II
Velikost	3,5“	3,5“	3,5“
Rychlost otáček	7200 ot/min	5400 ot/min	5400 ot/min

Tabulka 4 Parametry pevných disků [autor]

4.1.2 Software specifikace

Jak již bylo zmíněno, analyzovaná firma využívá pro zálohování dat softwarový program Bacula 5.2.6.

Jedna z hlavních výhod programu Bacula je automatizace zálohování v případě využití vícediskového systému. Bacula si jednotlivé diskové oddíly rozdělí a sám

označí. Jakmile začne docházet místo na oddílu kam zálohuje, administrátora systému automaticky kontaktuje emailem či pomocí konzole a začne zálohovat na další určený volný diskový oddíl. Bacula také umožňuje vytvořit spouštěcí (bootovací) soubor, který obsahuje původní nastavení čistého systému, v případě jeho selhání, tak umožní velice rychlou instalaci s daty a nastavením původního systému.

Mezi další výhody programu Bacula patří podpora databází MySQL. Pro soubory, které jsou zálohovány, lze vytvářet podpisové soubory SHA-1 a MD5 spolu s možností spouštění skriptů před a po zálohování nebo zálohování všech dat na jednom místě. Analyzovaná firma, ale možnost šifrování pomocí Bacula nevyužívá. Komunikace mezi jednotlivými komponentami probíhá za pomoci TCP/IP protokolu, lze použít i OpenSSH. V případě inkrementální zálohy, kterou analyzovaná firma využívá, systém nehlídá nové soubory dle poslední zálohy, ale za pomoci metadat informujících o času vytvoření či upravení souboru. V případě přidání souboru se starším datem vytvoření, než je poslední záloha, nebude Bacula brát tento soubor jako nový a tedy nutný zálohovat. Bacula dále zasílá automatické zprávy na email a to ať již o chybách či úspěšném provedení zálohy dat. Administrátor systému provádí kontrolu úspěšnosti zálohy jen podle této automatické zprávy, která přijde na email.

Bacula se skládá z několika komponent:

Bacula Director

Director v překladu znamená ředitel, což je výstižné slovo pro tuto komponentu, neboť se stará o řízení zálohování, sledování závislostí, ověřovací a obnovovací operace a také podávání zpráv o zálohování. Pomocí tohoto ředitele lze také zasílat dotazy Bacula Storage, či Bacula File démonu na klientských serverech.

Bacula Console

Bacula Console je komponenta, která umožňuje administrátorovi systému komunikaci s Bacula Directorem. Komponenta je dostupná v několika verzích v závislosti na uživatelském rozhraní. Analyzovaná firma využívá verzi textové

konzole, která je nejzákladnějším ovládacím prvkem, ale nabízí kompletní škálu možností, s kterými lze v programu Bacula pracovat.

Bacula File

Bacula File je démon, který se především stará o přesun dat na vybraného storage démona a to včetně metadat, tyto metadata obsahují informace o velikosti souboru, přístupových právech a datum vytvoření nebo poslední úpravy těchto dat. Musí být nainstalován na každém systému, který se zálohuje. Komunikuje také s Bacula Directorem ohledně informací jaký klient je právě zálohován a jaký storage démon je právě využíváný.

Bacula Storage

Bacula Storage démon slouží k interakci s médii, na která se zálohuje. Komunikuje tedy s diskovými oddíly, stará se o jejich připojování a odpojování.

Bacula Catalog

Funkci Catalog si lze prakticky představit jako databázový server. Hlavní funkcí je indexování souborů v databázi, která se vytváří z názvů a metadat jednotlivých zálohovaných dat. Díky indexování umožňuje administrátorovi přehledně prohlížet zálohy a soubory v nich uložené. Udržuje také záznamy o všech diskových oddílech a běžících procesech.

4.2 Politika zálohování firemních dat

Zálohování je velice důležitý proces, zejména ve firemní sféře. Až na pár výjimek se v analyzované firmě všechny systémy zálohují každý den. Využita je zálohovací strategie na způsob Grandfather-Father-Son. Každý den se provádí inkrementální zálohy, které se uchovávají po dobu jednoho týdne a po té se provede jejich recyklace. Na konci týdne se provede plná záloha, která se uchovává po dobu sedmi dní a poté se recykluje. Plná záloha v posledním týdnu daného měsíce se uloží na

disk Western Digital o velikosti 2TB, který je určený k dlouhodobému uchování. K záloze dochází pravidelně každý den v noci ve 3:00.

Pro analyzovaný podnik jsou velmi důležitá veškerá data uložená v databázích. Databáze jsou neustále aktualizované a doplňované o nové údaje a proto je nutné je velmi často zálohovat. Databáze se zálohují podle základní strategie firmy.

Důležitými daty, která je potřeba zálohovat jsou data z File serveru. Na File serveru jsou uchovávány veškeré informace o projektech pro zákazníky, na kterých se aktuálně pracuje. Jedná se také o kapacitně největší zálohovaná data. Disk, na který se tyto data zálohují, se po zaplnění mění jiný a uchovává se pro případné pozdější využití. K zaplnění disku dochází za 2-3 měsíce. Vzhledem k velké kapacitě dat se na tento disk provádí inkrementální záloha jen jednou týdně a plná záloha jednou za měsíc. Záloha se provádí na jeden z disků Western Digital o velikosti 2TB.

Dále dochází pravidelně k záloze stavu operačního systému, což je také velice důležitý proces. Stav operačního systému se zálohuje na dva disky Seagate o velikosti 500 GB. Tyto disky jsou zapojeny do pole RAID 1. To znamená, že záloha se provádí na oba disky současně a v případě poškození jednoho z disků jsou všechna data uložena na druhém disku. Díky této metodě je možné snadno obnovit předchozí stav systému. Stav operačního systému se zálohuje podle základní strategie firmy.

Dalším postupem v plánu zálohy dat je zkopírování plné zálohy dat všech zálohovaných systémů z posledního týdne v měsíci na jeden disk Western Digital o velikosti 2TB. Po zaplnění se disk mění za jiný a uchová pro případné pozdější využití. Tyto zálohy jsou uchovávány, než dojde k zaplnění druhého disku. Většinou po dobu 2-3 měsíců, po té dojde k jeho recyklaci.

4.3 Politika archivace firemních dat

Archivace dat je velmi důležitý proces. V analyzované firmě se provádí kompletní archivace všech důležitých dat jednou ročně. Tyto data se ukládají na zvláštní pevné disky, které se bezpečně uchovávají v trezoru v sídle analyzované firmy. Pro data určená k archivaci je využita komprese. Pro kompresi využívá analyzovaná firma

program 7-Zip. Program 7-Zip je využit i pro šifrování archivovaných dat pomocí symetrické šifry AES-256.

4.4 Nedostatky v zabezpečení dat a jejich navrhovaná řešení

Při analýze procesu zálohy a archivace firemních dat bylo zjištěno několik nedostatků.

4.4.1 Zálohování

Velkým nedostatkem analyzované firmy v procesu zálohování je ukládání všech zálohovaných dat pouze v sídle firmy. Analyzovaná firma se tímto vystavuje riziku ztráty zálohovaných dat v důsledku havárie nebo nečekané přírodní katastrofy. Tento problém se dá vyřešit více způsoby.

Prvním řešením by mohlo být vytvoření kopie pevného disku, na kterém jsou uchovávány měsíční zálohy. Tento disk by se mohl uložit mimo hlavní sídlo analyzované firmy. Vzhledem k tomu, že se tento disk po zaplnění mění za jiný a uchovává 2-3 měsíce a po té dojde k recyklaci, by tato kopie mohla fungovat na stejném principu. Ideální by bylo uchovávání této kopie po celý rok, než dojde k archivaci dat.

Druhým řešením by mohlo být využití Cloud úložišť. Hlavní výhodou využití vzdáleného datového úložiště je možnost zabezpečení dat bez nutnosti investice do dalšího hardware a software, data budou uložena vzdáleně od zálohovaného počítače a přístupné z jakékoliv pracovní stanice. Základní verze služby s omezenou kapacitou jsou dostupné zcela zdarma, čehož se dá využít pro otestování systému a pro porovnání jednotlivých poskytovatelů. Při využívání vzdáleného datového úložiště, ale dochází k nebezpečí odposlechu citlivých údajů během přenosu dat. Proto je vhodné data zašifrovat již před samotným odesláním.

Na vzdálená datová úložiště je lepší zálohovat pouze důležitá a kapacitně ne příliš velká data. Hlavním důvodem je celkem vysoká cena za větší úložný prostor. Pro

ukládání na tento typ úložišť mohou být vhodné například databáze. Plná záloha databází v analyzované firmě dosahuje 4 GB.

Pro srovnání jsou využita dvě velmi využívaná úložiště DropBox a IDrive.

Název	DropBoX
Free kapacita	2
Pro cena za 1 TB	9,99 €/měsíc
Business cena za GB	-
Business cena za uživatele	12 €/měsíc
Šifrování	AES-256, SSL

Tabulka 5 DropBox specifikace [autor]

Dropbox pro službu Business nenabízí přímo cenu za kapacitu, ale cenu za uživatele. V základu za 12€/měsíc dostane přístup 5 uživatelů a za každého dalšího uživatele se cena zvyšuje o dalších 12€. Kapacita je 1TB za uživatele.[26]

Název	IDrive
Free kapacita	5
Personal cena za 1 TB	44,62\$/rok
Business cena za 1,25 TB	374,62\$/rok, 49,95\$/měsíc
Business cena za uživatele	-
Šifrování	AES-256

Tabulka 6 IDrive specifikace [autor]

IDrive si na rozdíl od DropBoxu u Business služby účtuje cenu za kapacitu. Nabízí několik plánů, které je možno platit ročně nebo měsíčně. Účet pro přístup k datům může vytvořit administrátor a není omezen jejich počet.[27]

Poskytovatel IDrive nabízí větší kapacitu pronajatého úložného prostoru zdarma a služba Personal je cenově výhodnější než u konkurenčního DropBoxu Pro. Stanovení výhodnosti Business služby od jednotlivých poskytovatelů je závislé na počtu uživatelů, kteří potřebují mít k zálohovaným datům na datovém úložišti přístup, bohužel tyto

informace nejsou k dispozici, a proto nelze přesně určit, který poskytovatel by byl se službou Business pro analyzovanou firmu výhodnější.

Pro aktuální potřeby analyzované firmy, alespoň ze začátku není nutná služba Business. Proto by zde byl vhodnější kandidát IDrive zejména pro jeho nižší cenu se službou Personal než nabízí DropBox se službou Pro.

V procesu zálohování bylo dále nalezeno několik menších nedostatků, které by mohli být vylepšeny pro vyšší efektivitu procesu zálohování.

Doba uchování zálohovaných dat

Menším nedostatkem procesu zálohování je doba uchování inkrementálních a plných záloh před jejich recyklací. Aktuálně se uchovávají inkrementální a plné zálohy jen 7 dní. Vzhledem k tomu, že inkrementální zálohy jsou citlivé na poškození a na pevných discích je dostatek volné kapacity, mohlo by být výhodnější uchovávat inkrementální a plné zálohy po dobu 14 dní místo jen 7. Vhodné by bylo uchovávat plné zálohy po dobu celého měsíce, ale zde může nastat problém nedostatečné kapacity pevných disků.

Kompresce

Dalším menším nedostatkem je nevyužívání komprese pro kapacitně vyšší zálohy. Analyzovaná firma využívá metodu komprese jen pro potřeby archivace, pomocí programu 7-Zip. Vhodné by bylo využívat metodu komprese dat pomocí stejného programu i pro zálohu File serveru a pro disk určený k uchování měsíčních záloh. Disky se zaplní za zhruba 2-3 měsíce. Efektivita komprese závisí na tom, jaká data jsou použita. Průměr efektivita komprese se pohybuje kolem 35%. Díky tomu by se mohli zálohy uchovávat delší dobu, přibližně o 1 měsíc. Nevýhodou metody komprese je vyšší čas pro uložení zálohy, ale vzhledem k tomu, že v obou případech se provádí plná záloha jen jednou měsíčně, nejedná se o veliký problém.

Zálohování databází

Dalším menším nedostatkem u analyzované firmy je proces zálohování databází. V analyzované firmě jsou databáze neustále aktualizovány a doplňovány o nové údaje. Vhodné by bylo databáze zálohovat 2x denně místo jen 1x denně. Aktuálně jsou databáze zálohované každý den v noci v 03:00 hodin. Vhodný čas je pro druhou zálohu v 13:05. Důvodem pro tento čas je odpolední přestávka na oběd v analyzované firmě. Databáze nejsou kapacitně příliš velké a jejich záloha zabere nejvýše pár minut. V tomto čase se navíc vzhledem k odpolední přestávce s databázemi nepracuje, proto by měl být tento čas vhodný k jejich záloze.

Šifrování

Posledním menším nedostatkem je absence šifrování v procesu zálohování dat. Analyzovaná firma využívá šifrování jen u archivace dat. Vzniká zde vysoké nebezpečí zneužití citlivých dat neoprávněnou osobou. Důležitá data by tedy měla být zašifrována.

Program Bacula, který analyzovaná firma využívá pro zálohování, umožňuje vytvářet podpisové soubory pomocí algoritmů SHA-1 a MD5. Bylo by vhodné začít tuto službu využívat. Možné je i využití freewarových programů, kterých internet nabízí celou řadu.

4.4.2 Archivace

Archivace dat je důležitou součástí ochrany dat. V analyzované firmě se uchovávají archivovaná data na pevném disku v trezoru v sídle analyzované firmy. Analyzovaná firma se zde však vystavuje riziku ztráty veškerých dat v případě nečekané přírodní katastrofy či například požáru.

Bylo by vhodné uchovávat, alespoň jednu kopii archivovaných dat mimo hlavní sídlo společnosti a tím se zabezpečit proti nečekaným událostem a přírodním katastrofám. Nejsnadnějším řešením by pro analyzovanou firmu bylo vytvoření druhé kopie archivovaných dat na další pevný disk, který by se zabezpečil na místě mimo

hlavní sídlo analyzované firmy, ideálně vzdáleném i několik kilometrů v případě přírodní katastrofy.

4.5 Vyhodnocení

Analyzovaná firma používá pro zálohu svých systémů poměrně dobré metody a technologie. Program Bacula je velmi komplexní, kvalitní a dobře nastavený software pro automatizované zálohování dat.

Zálohování operačního systému pomocí pole RAID 1 je velice efektivní. V případě poškození jednoho z disků může být ihned nahrazen bez jakékoliv ztráty.

Zálohování a uchovávání kopie měsíčních záloh na jednom z disků vhodně doplňuje základní zálohovací strategii analyzované firmy a pomáhá k ochraně dat před jejich ztrátou.

I přes některé větší nedostatky disponuje analyzovaná firma schopnými administrátory, kteří zavedli vhodný a poměrně efektivní způsob zálohy a archivace dat. Největším problémem je absence uchovávání zálohovaných a archivovaných dat mimo hlavní sídlo analyzované firmy. Tento problém by bylo vhodné napravit.

5 Závěr

Záloha a archivace dat je v dnešní době nutností. Ve světě existuje mnoho důvodů, kvůli kterým může uživatel přijít o cenné informace a záleží teda jen na každém uživateli, jak velké úsilí do ochrany svých dat vloží.

Hlavním cílem práce byl návrh efektivního řešení zálohy a archivace dat. Prvním dílčím cílem bylo popsat různé možnosti a způsoby pro zálohu a archivaci dat, poukázat na jejich rozdíly, klady a zápory. Druhým cílem bylo analyzovat způsob zálohy a archivace dat v konkrétní firmě a v případě nedostatků navrhnout pomocí získaných teoretických znalostí jejich vyřešení.

Práce je rozdělena do dvou částí. První část shrnuje teoretické poznatky týkající se procesů zálohování a archivace dat a metody zpracování dat před samotnou zálohou či archivací. Druhá, praktická část, analyzuje firmu, která se zaměřuje na poskytování komplexních řešení pro efektivní řízení organizací veřejné správy, z hlediska zálohy a archivace podnikových dat. Jsou shrnuty zjištěné informace týkající se dané problematiky a poté jsou navrženy vhodné metody řešení objevených nedostatků. Je doporučeno ukládat zálohovaná a archivovaná data na místě mimo hlavní sídlo firmy, pro případ nečekané přírodní katastrofy. Dále by bylo vhodné zvýšit dobu uchování zálohovaných dat, zkrácení intervalu zálohy databází, komprese pro kapacitně vyšší zálohy a využít šifrovacích algoritmů, které nabízí využívaný software Bacula.

Veškeré vytyčené cíle byly tedy splněny.

6 Seznam použitých zdrojů

- [1] HUMPHRIES, Mark. *Data warehousing – návrh a implementace* Přel. M. Kocan. 1.vyd. Praha: Computer Press, 2001, 257 s. CD. ISBN 80-722-6560-1.
- [2] HARING D.: *Linuxové noviny*. České sdružení uživatelů operačního systému Linux. 2001, roč. 4 [cit. 2008-06-20]. Dostupný z WWW: <<http://www.linux.cz/noviny/2001-08/index.html>>. ISSN 1213-3655 .
- [3] DOSEDĚL, Tomáš. *Počítačová bezpečnost a ochrana dat*. Brno:Computer Press, 2004. ISBN 80-251-0106-1.
- [4] PECINOVSKÝ, Josef. *Archivace a komprimace dat*. Praha:Grada Publishing a.s., 2003, ISBN 80-247-0659-8.
- [5] VYMĚTAL, Dominik. *Informační systémy v podnicích - teorie a praxe projektování*. Praha: Grada Publishing, a.s., 2009. 142 s. ISBN 80-24730-46-4.
- [6] LEBER, Jody. *Windows NT: zálohování a obnova dat : příručka pro správce systémů*. Praha : Computer Press, 1998. 282 s. ISBN 80-72261-23-1.
- [7] RAIS, Karel. *Řízení rizik ve firmách a jiných organizacích*. Praha : Grada Publishing a.s., 2006. 296 s. ISBN 8024716674.
- [8] RICHTA, Karel. *Zásady a postupy zavádění podnikových informačních systémů: praktická příručka pro podnikové manažery*. Praha : Grada Publishing a.s., 2005. 187 s. ISBN 8024711036.
- [9] Srovnání antivirových programů, srovnání antivirů. *Antivirové centrum* [online]. 2015 [cit. 2015-01-25]. Dostupný z WWW: <<http://www.antivirovecentrum.cz/antiviry/srovnani.aspx>>
- [10] Compact Disc. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, [cit. 2014-10-22]. Dostupný z WWW: <http://en.wikipedia.org/wiki/Compact_disc>
- [11] DVD. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, [cit. 2014-10-22]. Dostupný z WWW: <<http://en.wikipedia.org/wiki/DVD>>
- [12] Blu-ray Disc. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, [cit. 2014-10-22]. Dostupný z WWW: <http://en.wikipedia.org/wiki/Blu-ray_Disc>

[13] DataTresorDisc [online]. datum vydání neznámé [cit. 2014-11-25]. Dostupný z WWW: <<http://www.datatresordisc.eu/technologie.html>>

[14] PAVLIS, Jakub. *Výrobci vidí budoucnost optických disků v zálohování a 4K* [online]. 2013 [cit. 2014-11-25]. Dostupný z WWW: <<http://notebook.cz/clanky/technologie/2013/budoucnost-optickych-disku-zalohovani-a-4k>>

[15] VAŘINA, Josef. *SSD disky – budoucnost?* [online]. 2009 [cit. 2014-11-25]. Dostupný z WWW: <<http://magazin.stahuj.centrum.cz/ssd-disky-budoucnost/>>

[16] Kingston DataTraveler HyperX Predator 1000G. *Alza* [online]. 2014 [cit. 2014-11-27]. Dostupný z WWW: <http://www.alza.cz/kingston-datatraveler-hyperx-predator-1000gb-d456739.htm>

[17] Coto je Cloud computing?. *Cloud Computing*[online]. datum vydání neznámé [cit. 2014-11-27]. Dostupný z WWW:<<http://www.cloudcomputing.cz/index.html>>

[18] DAS, SAN, NAS: *apexmicrosystems.com* [online]. [cit. 2014-11-28]. Dostupný z WWW: <http://www.apexmicrosystems.com/?page_id=518>

[19] DORIAN Cougias. *The Backup book: disaster recovery from desktop to data center*. 3. ed. Lecanto, FL: Schaser-Vartan Books, 2003. ISBN 09-729-0390-9.

[20] JUNEK, Pavel. *Zálohování a archivace dat v podnikovém prostředí* [online]. 2013 [cit. 2014-11-28]. Dostupný z WWW: <http://www.zalohovani.net/zalohovani-a-archivace-dat-v-podnikovem-prostredi-1-dil-zakladni-seznameni/>

[21] GUAY, Patrice. An overview of RAID technology. *IWeb* [online]. 2010 [cit. 2014-11-28]. Dostupné z: <http://blog.iweb.com/en/2010/05/an-overview-of-raid-technology/4283.html>

[22] LEIXNER, Miroslav. *PC - zálohování a archivace dat : Nestůjte za dveřmi*. Praha : Grada Publishing, a.s., 1993. 394 s. ISBN 80-85424-73-8.

[23] KRÍŽ, Libor. *Komprimační a archivační programy*. Praha : Computer Press, 2002. ISBN 80-7226-757-4.

[24] WD Green 600GB, *Alza* [online]. 2015 [cit. 2015-01-25]. Dostupný z WWW: <http://www.alza.cz/western-digital-green-6000gb-64mb-cache-d2141165.htm>

[25] Lenovo ThinkPad USB 3.0 Secure Hard Drive, *Alza* [online]. 2015 [cit. 2015-01-25]. Dostupný z WWW: <http://www.alza.cz/lenovo-thinkpad-usb-3-0-secure-hard-drive-1000gb-d562765.htm>

[26] Pricing. *DropBox* [online]. 2015 [cit. 2015-01-25]. Dostupný z WWW: <https://www.dropbox.com/business>

[27] *IDrive* [online]. 2015 [cit. 2015-01-25]. Dostupný z WWW: <https://www.idrive.com/online-backup-features>

7 Seznam obrázků a tabulek

Seznam obrázků

Obrázek 1 Interní pevný disk HDD [24].....	17
Obrázek 2 Externí pevný disk HDD s PIN zabezpečením [25]	17
Obrázek 3 SSD disk [14]	18
Obrázek 4 Flash disk [16]	19
Obrázek 5 Direct Attached Storage [18].....	21
Obrázek 6 Network Attached Storage [18].....	22
Obrázek 7 Storage Area Network [18].....	23
Obrázek 8 Raid 0 [20].....	26
Obrázek 9 Raid 1 [20].....	26
Obrázek 10 Raid 5 [20].....	27
Obrázek 11 Raid 6 [20].....	28
Obrázek 12 Raid 10 [20].....	28

Seznam tabulek

Tabulka 1 Round Robin [20].....	29
Tabulka 2 Grandfather-Father-Son [20].....	30
Tabulka 3 Hanojské věže [20].....	31
Tabulka 4 Parametry pevných disků [autor]	35
Tabulka 5 DropBox specifikace [autor].....	40
Tabulka 6 IDrive specifikace [autor]	40