**UNIVERZITA KOMENSKÉHO V BRATISLAVE**

**FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY**

**Zálohovanie súborového systému do obsahom adresovaného úložiska**

**Bratislava 2015  
Michal Molec**

**UNIVERZITA KOMENSKÉHO V BRATISLAVE**

**FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY**

**Zálohovanie súborového systému do obsahom adresovaného úložiska**

**Diplomová práca**

Študijný program: Aplikovaná informatika   
Študijný odbor: 2511 Aplikovaná informatika

Školiace pracovisko: Katedra aplikovanej informatiky

Školiteľ: Mgr. Michal Molec, PhD.

**Bratislava 2015  
Michal Molec**

Naskenované zadanie záverečnej práce

**Čestné vyhlásenie**

Čestne vyhlasujem, že som diplomovú prácu "Zálohovanie súborové systému do obsahom adresovaného úložiska" vypracoval samostatne s použitím uvedenej literatúry a zdrojov dostupných na internete.

V Bratislave dňa 1.6.2015

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Michal Molec

**Poďakovanie**

Chcel by som sa poďakovať vedúcemu diplomovej práce Mgr. Jánovi Kľukovi, PhD. za cenné rady a usmernenie pri vypracovaní tejto práce.

**Abstrakt**

MOLEC, Michal: Zálohovanie súborového systému do obsahom adresovaného úložiska - Univerzita Komenského v Bratislave. Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej informatiky. Vedúci diplomovej práce: Mgr. Ján Kľuka, PhD. Bratislava, 2015.

Cieľom diplomovej práce je navrhnúť vhodné obsahom adresované úložisko dát (content-addressable storage) a implemetovať nad ním zálohovanie súborového systému unixového OS (napr. Linux), údržbu a prezeranie záloh a obnovu dát. Návrh samotného úložiska môže vychádzať z už existujúceho (napr. git), treba ho však upraviť s ohľadom na požadovaný účel. Očakáva sa kreatívne využitie (a prípadná úprava) vhodne zvolených techník kompresie (klasická bezstratová, rozdielová) a garbage collection, technológie FUSE, prípadne sledovania zmien v súborovom systéme. Implementácia by mala byť objektová, modulárna a rigorózne otestovaná (unit testing pokrývajúci funkcionalitu kritickú pre spoľahlivosť zálohovania). Preferovaným implementačným jazykom je python..

**Kľúčové slová:** Zálohovanie, FUSE, Rsync, Linux, Python

**Abstract**

MOLEC, Michal: File-system backup into content-addressable storage – Comenius University in Bratislava. Faculty of mathematics, physics and informatics, Department of aplied informatics. Dissertation supervisor: Mgr. Ján Kľuka, PhD. Bratislava, 2015.

The goal of this work is to design suitable content-addressable storage (CAS) and implement backup, backup maintenance and browsing, and data recovery of a Unix file system over the CAS. The design of the CAS may be based on existing ones (e.g., Git), it should, however, be adjusted for backup purposes. The student is expected to creatively choose, use, and adjust suitable methods of compression (standard lossless compression, delta compression) and garbage collection, the FUSE technology, and, possibly, the file-system event notification mechanism (inotify). The implementation should be object-oriented, modular, and rigorously tested (unit testing covering components critical to reliability of backup). The preferred implementation language is python.

**Keywords:** Backup, FUSE, Rsync, Linux, Python

**Obsah**

[Úvod 10](#_Toc405922246)

[1 Úvod do problematiky 11](#_Toc405922247)

[1 Návrh riešenia 11](#_Toc405922248)

[2 Implementácia 12](#_Toc405922249)

[Záver 10](#_Toc405922250)

[Bibliografia 11](#_Toc405922251)

[Prílohy 12](#_Toc405922252)

# Úvod

# Existujúce riešenia

V tejto kapitole budú uvedené existujúce riešenia zálohovania dát a ich základný princíp fungovania.

## Software

Rdiff-backup: <http://www.nongnu.org/rdiff-backup/>

Duplicity: <http://www.nongnu.org/rdiff-backup/>

Time Machine: <https://support.apple.com/en-us/HT201250>

AMANDA: http://www.amanda.org/

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | rdiff-backup | duplicity | Time Machine | AMANDA |
| Licencia | GPL | GPL | Apple Inc. | BSD |
| Windows / MacOS / Linux | ✓/✓/✓ | ✓/✓/✓ | 🗶/✓/🗶 | ✓/✓/✓ |
| GUI / príkazový riadok | ✓/✓ | ✓/✓ | ✓/✓ | 🗶/✓ |
| Programovací jazyk | Python | Python | ? | C, Perl |
| Verziovanie | opačné inkrementálne | inkrementálne | inkrementálne | Inkrementálne |

### Rdiff-backup

Rdiff-backup je zálohovací software pre plne automatické inkrementálne zálohovanie a spätné zrekonštruovanie zálohovaných dát. Obnovovanie dát môže prebiehať na konkrétne zvolenom súbore alebo celej zálohe. Implementovaný je v jazyku Python a pri svojom behu využíva program rsync (knižnica librsync).

Základné vlastnosti:

* Kopírovanie súborov lokálne alebo po sieti cez protokol SSH
* Zachovanie práv súborov a médií (zálohovanie na flash disk so súborovým systémom FAT)
* Obnovenie zmazaných súborov
* Štatistiky a záznamy o zálohách

Samotné zálohovanie prebieha tak, že na cieľovom adresári je vytvorená presná kópia zálohovaného adresára. Ak nastanú na zálohovaných súboroch zmeny, tak rdiff-backup pri ďalšom zálohovaní ukladá aj reverzné diffy k zmeneným súborom do špeciálneho podadresára. Tie môžu byť neskôr využité pri prípadnej obnove starších verzií súboru. Výhodou takéhoto prístupu je aktuálne verzie zálohovaných dát sú ihneď k dispozícii a stačí ich len prekopírovať.

Spolu s dátami sú k jednotlivým súborom udržiavané aj práva, vlastník a skupina, čas zmeny. Rdiff-backup dokáže zálohovať aj súbory zariadení, pipe, hardlink a symlink.

### Duplicity

Duplicity je software pre zálohovanie dát a funguje na princípe vytvárania kryptovaných (prípadne podpísaných) tar archívov. Implementovaný je v jazyku Python a pri svojom behu taktiež využíva program rsync (knižnica librsync).

Základné vlastnosti:

* Kopírovanie súborov lokálne alebo po sieti cez protokol SSH
* Kryptovanie a podpisovanie vytvorených tar archívov pomocou GnuPG
* Inkrementálne zálohovanie

Princíp zálohovania v Duplicity je nasledovný. Každá záloha je ukladaná do tar archívu. Duplicity poskytuje aj možnosť kryptovania a podpisovania zálohovaných dát pomocou GnuPG. Pre ich využitie je potrebné si najskôr vygenerovať samotný GPG kľúč, prípadne digitálny podpis.

Následne je ako prvá vytvorená počiatočná záloha, ktorá obsahuje všetky dáta (full backup). Pri ďalšej zálohe sa ukladajú len rozdiely (diffy) voči pôvodným súborom. Nevýhodou takéhoto prístupu môže byť čas, potrebný pre obnovenie starších záloh. Napríklad pri prístupe k najnovšej verzii zálohy je potrebné prejsť všetkými inkrementálnymi zálohami, ktoré boli vykonané.

Samotné Duplicity je nástroj ovládateľný cez príkazový riadok. Existuje k nemu však kvalitná grafická nadstavba s názvom Déjà Dup. Medzi jej hlavné výhody patrí napríklad plánovanie záloh, integrácia v Linuxovom grafickom rozhraní GNOME (Nautilus) alebo podpora pre cloudové zálohovanie. Vďaka integrácii v GNOME si v súborovom manažéri Nautilus dokážete jednoducho obnoviť akýkoľvek súbor bez toho, aby ste museli obnovovať celú zálohu.

### AMANDA (Advanced Maryland Automatic Network Disk Archiver)

AMANDA je zálohovací software, ktorý umožňuje zálohovanie dát po sieti LAN. Funguje na princípe jedného zálohovacieho servera, ktorý sa stará o zálohovanie viacerých klientov (Windows, Linux, Unix, BSD, Mac OS-X). Používa k tomu natívne nástroje ako dump a GNU tar. Dump je linuxový nástroj vyvinutý hlavne pre zálohovanie diskových oddielov. Kvôli spätnej kompatibilite ukladania záloh na pásky, sú jeho výstupom bloky pevných veľkostí. GNU tar je taktiež linuxový nástroj ktorý slúži kompresiu dát.

Základné vlastnosti:

* Inkrementálne aj úplne zálohovanie
* Podpora šifrovania dát
* Zálohovanie viacerých platforiem – využíva k tomu buď samba protokol alebo svoj vlastný, natívny klient s názvom amanda-client
* Open Source
* Dáta su obnoviteľné aj bez samotného softwaru Amanda

### Time Machine

Time Machine je vstavaná zálohovacia služba systému Mac OS-X, ktorá vytvára zálohy na externý disk alebo na bezdrôtové zariadenie Time Capsule (Apple), ktoré slúži ako úložisko pre tieto zálohy.

Základné vlastnosti:

* Podpora šifrovania dát
* Záloha celého systému, vrátane systémových súborov, aplikácií, používateľských účtov atď.
* Kompletný obraz systému pre daný čas – možná obnova
* Nezmenené súbory ukladá ako odkazy (hardlinky) k pôvodným súborom a ukladá len nové, prípadne zmenené súbory
* Možnosť vyhľadávania súborov a rýchly náhľad do súboru
* Udržiava hodinové zálohy za posledných 24 hodín, denné zálohy za posledný mesiac a týždenné zálohy

Time Machine má taktiež podporu pre zálohovanie na NAS zariadenia a servery. Záleží však na samotnej verzii softvéru. Kým staršie verzie dokázali spolupracovať s veľkým množstvom NAS serverov, najnovšie verzie vyžadujú server podporujúci AFP (Apple Filling Protocol). AFP je sieťový protokol poskytujúci súborové služby pre Mac OS X a Mac OS.

### Git

Git je decentralizovaný / distribuovaný systém riadenia revízií (DVCS). Neslúži však ako zálohovací software, ale ako software pre správu verzií. Bol vyvinutý pre vývoj Linuxového jadra. Väčšinou je však využívaný vývojármi pri práci na softwarových projektoch, ale dá sa využívať na akékoľvek dáta. Dá sa tak jednoducho zistiť, kedy sa daný súbor zmenil, ako sa zmenil a taktiež kým bol zmenený.

Základné vlastnosti:

* Git je distribuovaný systém, čo znamená, že nepotrebuje žiadny centrálny bod, s ktorým by musel udržiavať spojenie. Všetky základné operácie sú vykonávané lokálne.
* Kompletná história bez možnosti zmeny
* Kompatibilita s existujúcimi systémami a protokolmi: HTTP, FTP, SSH, rsync, alebo komunikácia priamo cez socket
* Garbage kolekcia
* Delta kompresia
* Viaceré vetvy vývoja

Pojem garbage kolekcia bude zadefinovaný a vysvetlený neskôr v kapitole 1.2.4. V Gite je využívaná garbage kolekcia hlavne za účelom exportovania súborov, ktoré už nie sú viac využívané a označí ich ako „stratené“ objekty (loose objects). Pri hľadaní týchto súborov garbage kolekcia mení samotnú štruktúru repozitára na efektívnejšiu a snaží sa tak redukovať diskový priestor a zvýšiť výkonnosť.

Delta kompresia (delta kódovanie) je mechanizmus ukladania prípadne prenášania dát vo forme rozdielov medzi jednotlivými verziami. Rozdiely sú ukladané do súborov, ktoré sú nazývané taktiež delty alebo diffy. Git si vždy ukladá celý obsah najnovších súborov a rozdiely medzi súbormi sú ukladané pomocou delta kompresie len ak sú označené ako „stratené“ súbory (loose objects). Tieto súbory sú následne komprimované do binárnych súborov nazývaných tiež balíčkové súbory (packfile). Vďaka tomuto prístupu dokáže Git šetriť diskový priestor a byť efektívnejší.

## Techniky

### Obsahom adresované úložisko (content addressed storage)

Obsahom adresované úložisko dát funguje na nasledovnom princípe. Na strane klienta, je buď aplikačným serverom alebo samotným klientom vytvorený objekt na základe jeho obsahu (content address). Ten je následne odoslaný do úložiska a je pre neho vypočítaná adresa (hash), podľa ktorého je neskôr daný objekt jednoznačne identifikovateľný. Vypočítaná adresa je následne odoslaná späť aplikácií, ktorá pre ďalšie referencie na daný objekt, pracuje už len s ňou.

Výhody a nevýhody:

* Najvhodnejšie na dátach, ktoré sa často nemenia (kvôli častému počítaniu / prepočítavaniu adries pre jednotlivé súbory)
* Rýchle vyhľadávanie pre obsah daného súboru
* Nikdy neexistuje viac ako jedna kópia daného súboru v úložisku (dva také isté súbory majú rovnakú adresu – content address)

### Rsync

Rsync algoritmus bol vyvinutý Andrewem Tridgellom za účelom zeefektívnenia prenosu dát za podmienky, že sa na cieľovom počítači už nachádza predchádzajúca verzia daných dát.

Algoritmus efektívne počíta rozdiely medzi súbormi pomocou signatúr. V algoritme sú použité dva typy signatúr, silná („strong checksum“) a slabá („weak checksum“ alebo „rolling checksum“) signatúra. Slabá signatúra musí byť veľmi rýchla a lacná na výpočet. Silná signatúra musí mať veľmi nízku pravdepodobnosť vzniku kolízie a je počítaná len vtedy, keď sa dve slabé signatúry zhodujú. Princíp fungovania rsync algoritmu je nasledovný. Predstavme si, že máme dva počítače A a B. V počítači A sa nachádza súbor, ktorý pozostáva ai bytov a v počítači B sa nachádza súbor, ktorý pozostáva z bi bytov. B rozdelí týchto bi bytov na N rovnakých blokov a vypočíta slabú aj silnú signatúru ku každému bloku. Následne sú tieto signatúry odoslané na počítač A. Tu sa vypočítajú pre každý byte offset i v ai slabé signatúry pre každý blok začínajúci na pozícii i. Takto získané signatúry sú porovnané s každou slabou signatúrou získanou z počítaču B. Ak nastane prípad, že sa dve signatúry rovnajú, A ďalej vypočíta aj silnú signatúru a porovná ju so silnou signatúrou z B, ktorá patrí danému bloku. Ak sa zhodujú aj tieto dve signatúry, tak A odošle správu B o mieste zhody a ktoré bloky sa zhodujú. Ak sa nezhodujú, tak sú odoslané len samotné byty. Napokon sú na počítači B použité tieto informácie na rekonštrukciu ai. K výpočtu oboch signatúr je používaný hashovací algoritmus 128 bit MD4.

Samotná rekonštrukcia súboru je jednou z najjednoduchších častí rsync algoritmu. Akonáhle sú na počítač B odoslané všetky informácie o zhodujúcich sa blokoch alebo samotné byty, tak rekonštrukcia súboru sa začne vykonávať. Pri rekonštrukcii súboru sa postupne zapisujú získané byty z počítača A alebo jednotlivé zhodujúce bloky bytov v súbore, nachádzajúcom sa na počítači B.

### FUSE

FUSE (Filesystem in Userspace) alebo tiež súborový systém v užívateľskom priestore umožňuje bežným uživateľom (Linux, Unix, FreeBSD, NetBSD, OpenSolaris, Mac OS-X) pristupovať k virtuálnym súborom systému. Taktiež umožňuje vytváranie súborových systémov z akýkoľvek dát a poskytovaných služieb.

Ovládač súborového systému je kernelový modul a je pevnou súčasťou samotného kernelu. Daný súborový systém môže obsluhovať len root a ten prideľuje práva k jednotlivým mountovacím bodom. Samotné použitie súborového systému funguje tak, že kernel dostane požiadavku od aplikácie, napríklad „vypíš adresár“, a ten prenechá ovládaču súborového systému. Ten následne požiadavku obslúži a spracuje odpoveď, ktorú kernel vráti späť aplikácii.

Samotný FUSE v kerneli poskytuje rozhranie ako obyčajný súborový systém. Nezaujíma sa však o obslúženie požiadavky a spracovanie odpovedi a namiesto toho, pomocou knižnice libfuse spustí bežnú používateľskú aplikáciu a tá vykoná všetko potrebné. Akonáhle aplikácia vráti odpoveď v požadovanej forme, FUSE ju prenechá kernelu a kernel ju vráti späť samotnej aplikácii.

Obrázok 1. Tok riadenia FUSE



Výhody:

* Celá logika súborového systému je zapísaná v programe prípadne skripte, ktorý je nezávislý na platforme či operačnom systéme
* Súborové systémy môže obsluhovať aj bežný uživateľ
* Stabilné API pre FUSE – nie je potrebné prepisovať programy

Ako môžeme vidieť na obrázku 1. FUSE a taktiež všetky iné súborové systémy využívajú VFS (virtual file system). VFS poskytuje pre súborové systémy abstraktné rozhranie pre komunikáciu s kernelom. FUSE musí preto pre prácu so súbormi implementovať dôležité metódy ako napríklad open, close, read, write alebo seek. Toto rozhranie je poskytované skupinou štandardou s názvom POSIX (Portable Operating System Interface).

Najznámejšie existujúce FUSE aplikácie sú napríklad sshfs, ntfs 3g, exfat-fuse. Sshfs (secure shell file system) je súborový systém založený na SSH FTP (file transfer protokol) protokole. Vďaka tomu, že väčšina SSH serverov podporuje tento protokol, tak samotné používanie sshfs je veľmi jednoduché. Stačí len vykonať pripojenie (mount) takéhoto súborového systému.

Ntfs 3g a exfat-fuse open source read – write ovládače (drivery). Ntfs 3g je implementáciou súborového systému NTFS od Microsoftu. Medzi jeho hlavné výhody patrí napríklad žurnalovanie, kompresia alebo podpora šifrovania.

Exfat-fuse je implementáciou súborového systému exFAT (Extended File Allocation Table), ktorý je vhodný hlavne pre flash disky. Využíva namiesto NTFS, ktorý je nevhodný pre takéto zariadenia, kvôli množstvu dátových štruktúr, ktoré obshauje.

Využitie FUSE je však skutočne široké a dá sa využiť napríklad aj ako súborový systém pre Wikipédiu (WikipediaFS), Gmail (GmailFS) alebo Dropbox (DropboxFS). WikipediaFS slúži ako súborový systém, pomocou ktorého môžu používatelia Wikipedie priamo editovať články, ako na svojom lokálnom disku. Hlavnou výhodou je teda editácia článkov vo vlastnom textovom editore.

GmailFS je súborový systém, pomocou ktorého dokážete využívať svoje Gmail konto ako úložisko dát. Táto funkcia však nie je priamo podporovaná Googlom.

### Reference counting garbage collection

Garbage kolekcia je mechanizmus, ktorý slúži na automatickú správu zdrojov v počítači. Najčastejšie sa však využíva na správu pamäte. Jeho úlohou je v tomto prípade získavať pamäť, ktorá už nie je využívaná prostriedkami (programy a procesy). Uľahčuje tak aj prácu programátorom, ktorí by vždy museli manuálne definovať, ktoré objekty môžu byť dealokované a uvoľniť tak pamäť systému.

Najdôležitejší problém, ktorý riešia všetky garbage kolektory, je priradenie objektov do dvoch skupín. Na objekty, ktoré sú dosiahnuteľné a na objekty, ktoré nie sú dosiahnuteľné.

Najjednoduchšou implementáciou garbage kolektora je pomocou metódy reference counting. Funguje na nasledovnom princípe. Ku každému objektu je asiciovaný počet referencií na daný objekt. Ak hodnota klesne na 0, objekt je následne označený ako nedosiahnuteľný a je alebo bude uvoľnený z pamäte. Táto metóda však nedokáže pracovať s cyklickými štruktúrami.

# Návrh riešenia

V tejto kapitole je popísaný návrh riešenia našej aplikácie.

# Implementácia

V tejto kapitole bude rozoberaná konkrétna implementácia navrhnutých riešení z predchádzajúcej kapitoly a pre lepšiu predstavivosť a pochopenie budú uvedené aj konkrétne príklady.

# Záver

# Bibliografia

[1] ANDREW, Tridgell. Efficient Algorithms for Sorting and Synchronization [Dátum Júl 1999] http://gan.anu.edu.au/~brent/pd/Tridgell-thesis.pdf

[2] SCOTT, Chacon. Pro Git [Dátum 8. Február 2010] <http://labs.kernelconcepts.de/downloads/books/Pro%20Git%20-%20Scott%20Chacon.pdf>

[3] CURTIS, Preston. Backup & Recovery (4. kapitola)

[Dátum Január 2007]

[4] <https://www.kernel.org/doc/Documentation/filesystems/fuse.txt>

[5] L. PETER Deutsch, DANIEL G. Bobrow. An efficient, incremental, automatic garbage

collector

[Dátum 9. September 1976]

<http://dl.acm.org/citation.cfm?id=360345>

[6] HENRY Lieberman, CARL Hewitt. A real-time garbage collector based on lifetimes of

objects

[Dátum 6. Jún 1983]

http://dl.acm.org/citation.cfm?id=358147

# Prílohy

1. CD so zdrojovými kódmi aplikácie a bakalárskou prácou vo formáte PDF