

Porady: Kopie zapasowe

PRZEZORNY ZAWSZE UBEZPIECZONY

fotograf: www.photocase.de

Dane z komputerów tracimy zawsze w złym momencie, lecz mając właściwą strategię wykonywania kopii zapasowych, szybko przywrócimy stracone pliki na twardym dysku.

MARC ANDRÉ SELIG

Utrata danych może mieć wiele różnych przyczyn. Wpisanie polecenia `rm *` w niewłaściwym katalogu może spowodować usunięcie setek plików. Całkowita utrata danych następuje wskutek awarii dysku twardego. Zaś w razie złamania zabezpieczeń systemu użytkownicy nie mogą już uznawać pozostawionych przez intruza danych za wiarygodne. Administrator nie ma wyboru, musi przywrócić system z bezpiecznej kopii.

Ponieważ przyczyny i okoliczności utraty danych mogą być niesłychanie różne, z upływem czasu powstało wiele rozmaitych rozwiązań. Wszystkie mają swoje wady i zalety. W Warsztacie administratora w tym miesiącu opiszemy typowe narzędzia i techniki wykonywania kopii zapasowych.

Opcje kopii zapasowych

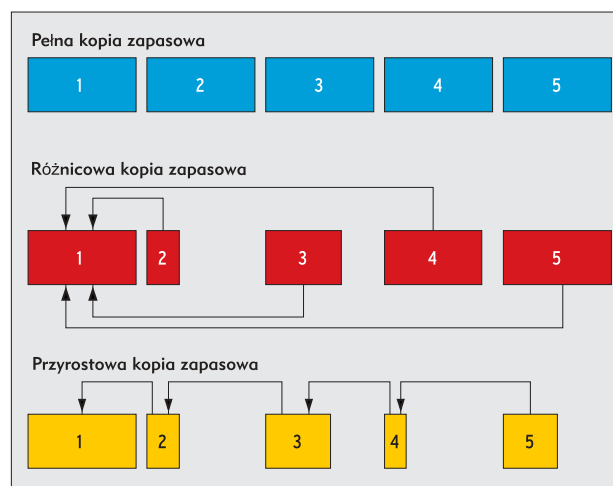
W Tabeli 1 porównałem różne nośniki kopii zapasowych. Niegdyś najpopularniejsze były taśmy magnetyczne. Są one do dziś typowe w sieciach z dużą ilością danych. Taśmy są dosyć tanie, mimo ich wysokiej pojemności, ale szybkość jest ich piętą achillesową. Kopie zapasowe wykonywane na taśmach z użyciem urządzeń typu jukebox są idealne do wykonywania automatycznych kopii zapasowych. Jednak napędy taśmowe są często za drogie dla małych biur i do domów.

Płyty CD, DVD, pamięć błyskowa oraz kopie na wewnętrznych i zewnętrznych dyskach twardech cieszą się obecnie większą popularnością. W większych środowiskach administratorzy mogą korzystać z systemu NAS (Network Attached Storage), aby zwiększyć pojemność dysków twardech.

Ile rodzajów nośników kopii zapasowych, tyle strategii wykonywania samych kopii. W większości wypadków administratorzy wybierają kopie przyrostowe, gdzie zapisywane są tylko zmiany, które nastąpiły od czasu wykonania ostatniej pełnej lub przyrostowej kopii zapasowej. Taka metoda pozwala zaoszczędzić miejsce na nośnikach kopii zapasowych, co zwiększa rentowność tego rozwiązania i przyczynia się do jego popularności.

Wielką wadą przyrostowych kopii zapasowych jest to, że przywrócenie utraconych lub uszkodzonych plików wymaga więcej czasu niż wykonanie

pełnej kopii. Poza tym administratorzy niekiedy muszą ręcznie zmieniać nośniki kopii zapasowych, jeśli nie mają rozwiązania wykorzystującego urządzenie typu jukebox (automatyczny zmieniacz). Trzeci wariant, różnicowa kopia zapasowa, zawsze zawiera zmiany od czasu ostatniej pełnej kopii zapasowej. Te trzy metody przedstawiono na Rysunku 1.



Rysunek 1: Pełna kopia zapasowa to kopia wszystkich plików. Kopia różnicowa to kopia wszystkich plików zmodyfikowanych od czasu wykonania ostatniej pełnej kopii. Przyrostowa kopia obejmuje wszystkie pliki zmodyfikowane od czasu wykonania ostatniej przyrostowej lub pełnej kopii.

W trybie off-line, online, podłączana w ruchu

Wybór metody tworzenia kopii zapasowej może zależeć od wymaganych sposobów przywracania danych. Jeśli plik bardzo potrzebny użytkownikowi jest zlokalizowany na taśmie w szafie, dostęp do pliku będzie wymagał interwencji człowieka. To może mieć zalety: w końcu agresorowi będzie trudno złamać zabezpieczenia danych na tej taśmie. Jednak przywracanie pliku będzie wymagać trochę czasu i pracy.

Natomiast kopie zapasowe dostępne w trybie online znajdują się na nośnikach, które obsługują automatyczny dostęp przez 24 godziny na dobę, 7 dni w tygodniu. Ta metoda oznacza oszczędność czasu, a często także pieniędzy.

Systemy mogą regularnie, a nawet stale tworzyć tak zwane kopie zapasowe podłączone w ruchu. Jednak kopie zapasowe tego rodzaju chronią jedynie przed uszkodzeniem sprzętu. Nie stanowią ochrony przed błędami użytkownika ani administratora, które zostaną przeniesione na nośnik kopii wkrótce po zapisaniu przez użytkownika lub administratora niewłaściwych danych. Z tego powodu administratorzy najczęściej nie uznają technik kopii zapasowych podłączanych w ruchu za coś, co mogłoby zastąpić konwencjonalne metody tworzenia kopii.

Formaty

Administratorzy nie są jednomyślni w ocenie, co jest lepsze: pojedyncze pliki, czy też bardziej skomplikowane archiwa, które zawierają strukturalny zestaw kopii zapasowych (lub wszystkie pliki zapasowe na nośniku kopii) wraz z metadanymi i sumą kontrolną.

Pojedyncze pliki łatwiej się przywraca, a jeśli na nośniku kopii występuje lokalna usterka, dotyczy zawsze pojedynczego pliku, podczas gdy zniszczone archiwum oznacza utratę wielu plików.

Pliki archiwów mają jednak funkcje, których nie mogą zapewnić kopie zapasowe pojedynczych plików. Oprócz plików zawierają na przykład dane o właścicielu, przywilejach dostępu oraz znacznikach czasowych. Można nawet wykonać kopie zapasowe urządzeń specjalnych z katalogu `/dev/`. Co więcej, taśmy magnetyczne nie są dobrze przystosowane do przechowywania ogromnych ilości indywidualnych plików. Znacznie lepiej sprawdzają się jako nośniki pojedynczych, wielkich plików.

Niektóre programy, m.in. `tar` i `cpio`, mają stanowić złoty środek. Jeśli plik w formacie `cpio` jest uszkodzony, uszkodzenia ograniczają się do plików przechowywanych w uszkodzonej lokalizacji. Program ponownie

synchronizuje archiwum w miejscu znacznika końca następnego pliku; kolejne pliki można odtworzyć bez problemu.

Rozważając wady i zalety pojedynczych plików i archiwów, należy brać pod uwagę także kompresję i szyfrowanie. Funkcja ponownej synchronizacji plików `cpio` jest dostępna tylko dla nieskompresowanych kopii zapasowych. Jeśli błąd odczytu uniemożliwia dekompresję archiwum, program `cpio` nie może nic wskórać.

Popularne narzędzie `gzip` kończy pracę w razie wystąpienia błędów odczytu, a zatem nie nadaje się do wykonywania kopii zapasowych. (Program `zcat` przynajmniej dekompresuje archiwum do miejsca występowania błędu). Jego konkurent, `bzip2`, kompresuje i dekompresuje pliki w blokach o maksymalnej wielkości 900 KB. W razie błędu odczytu można utracić tylko jeden blok; poprzednie i następne bloki prawdopodobnie nie będą uszkodzone.

Administratorzy mają podobne trudności z danymi szyfrowanymi. Większość szyfrów strumieniowych wykorzystywanych przez programy do wykonywania kopii zapasowych uniemożliwia jakikolwiek dostęp do archiwum w razie błędu. Pewnym wyjściem z tej sytuacji byłoby indywidualne kompresowanie i/lub szyfrowanie poszczególnych plików w archiwum. Narzędzie `afio` może zastąpić program `cpio`, gdyż może obsługiwać oddzielne szyfrowanie zarchiwizowanych plików.

Kopie zapasowe na dyskach CD

Rozwiązanie służące do tworzenia kopii zapasowych na taśmach, takie jak Amanda (patrz Ramka 1), dobrze skaluje się w dół, lecz jest chyba lepiej przystosowane do większych środowisk. Użytkownicy prywatni oraz małe firmy mogą zadowolić się prostą kopią na dyskach CD lub DVD. W porównaniu z taśmami magnetycznymi dyski CD i DVD są niezwykle tanie i mają większą żywotność.

Przykład: Taśmowa kopia zapasowa

Taśmy są popularnym i rozpowszechnionym nośnikiem kopii zapasowych. Taśmy często mają izolowane błędy odczytu, których można uniknąć dzięki wykorzystaniu bardziej zaawansowanych narzędzi programowych. Sprawę pogarsza to, że wiele sterowników jądra wymaga wstępnie sformatowanych bloków dla urządzeń taśmowych. Innymi słowy, nie każde urządzenie taśmowe nadaje się dla polecenia `tar cpf`.

Najprostszą metodą jest użycie gotowego systemu do wykonywania kopii, takiego jak Amanda [2], który może pobierać dane z (niemal) nieograniczonej liczby komputerów i zapisywać kopie na taśmie. Amanda obsługuje wiele systemów uniksowych. Istnieją nawet klienckie wersje dla systemu Microsoft Windows [3].

System opiera się na modelu klient-serwer. Na każdym komputerze, z którego mają być pobierane dane do tworzenia kopii zapasowych przez system Amanda, należy zainstalować klienta Amandy. Oczywiście, klient potrzebuje przywilejów odczytu do danych przeznaczonych dla serwera Amandy. Następnie serwer w protokole UDP transmituje zapytania do klientów, które odpowiadają, przysyłając kopie zapasowe przez protokół TCP. W systemie Amanda do tworzenia plików archiwów mogą służyć programy `dump` lub `tar`.

Amanda ma wyrafinowane metody planowania pojedynczych kopii zapasowych. Program serwera, na podstawie informacji o puli taśm oraz o skonfigurowanych odstępach między pełnymi kopiami, planuje zadania wykonania kopii pełnych i przyrostowych. Oznacza to, że każdy komputer ma wykonywaną kopię tak często, jak to tylko możliwe; przynajmniej tak często, jak określa to konfiguracja. Amanda następnie wypełnia luki na taśmach przyrostowymi kopiami zapasowymi.

Tabela 1: Możliwości wykonywania kopii zapasowych

	Trwałość	Niezawodność	Z zewnętrznej lokalizacji	Szybka dostępność
Taśma magnetyczna w trybie off-line	++	++	++	-
CD/DVD w trybie off-line	+	+	++	+
Dysk MO w trybie off-line	++	+	++	+
Dysk twardy (wewnętrzny) w trybie online (podłączany w ruchu)	-	-	-	++
Dysk twardy (zewnętrzny)	-	-	++	o w trybie online (podłączany w ruchu)

+ +: silny punkt, +: dotyczy, o: częściowo dotyczy - w zależności od nośnika, -: nie dotyczy

Na Listingu 1 znajduje się prosty skrypt do tworzenia kopii zapasowych, który wywołuje narzędzie *gpg* w celu zaszyfrowania danych kopii i przechowuje prostą sumę kontrolną MD5 w celu rozruchu. Jeśli dysk CD się zagubi, przynajmniej nie musimy się martwić nieautoryzowanym dostępem do danych. Łatwo zmodyfikować ten skrypt pod

kątem nośników pamięci błyskowej lub zewnętrznych dysków twardech.

Właściwe podejście

System kopii zapasowych jest tyle wart, co dane na nośniku. Zaś te dane to niekoniecznie to, co program do tworzenia kopii miał zapisać. Dlatego też warto regularnie

sprawdzać czytelność i adekwatność kopii zapasowych.

Należy też upewnić się, czy użytkownicy są w stanie przywracać swoje dane. Nie ma nic bardziej irytującego niż próba przywrócenia kopii zapasowej skonfigurowanej przez kogoś dawno temu, gdy okazuje się, że jest to niemożliwe, gdyż inżynier, autor konfiguracji tego systemu, już nie pracuje w firmie.

Przypadek całkowitej utraty danych prowadzi do całego szeregu dodatkowych problemów. Ponieważ sam system operacyjny może być niedostępny, przyda się system awaryjny. System awaryjny powinien uruchamiać się z dysku CD lub zewnętrznego dysku twardego, pozwalając administratorowi przywrócić pełen zbiór danych. Oczywiście, takiego rodzaju rozwiązanie wymaga starannego planu i dużej praktyki. ■

Listing 1: Prosty skrypt do tworzenia kopii zapasowych

```
#!/bin/sh
[ 'id -u' -eq 0 ] || (echo 'Tylko
użytkownik root może zapisywać
dyski CD/DVD!' && exit )

TODAY='date +%Y%m%d.%H%M'
MYKEY='0x598342d9'

umask 022
mkdir -p /tmp/root/backup-$TODAY

cd /
tar cf - etc home usr/local | \
gpg -v -homedir $HOME/.gnupg -e -r $MYKEY | \
tee > /tmp/root/backup-$TODAY/backup-$TODAY.tar.gpg | \
md5sum - b >/tmp/root/backup-$TODAY/backup-$TODAY.tar.gpg.md5

cd /tmp/root
mkisofs -r -pad -o backup.iso >
backup-$TODAY
cdrecord -v -eject -multi >
dev=0,0,0 -driveropts=burnproof >
-speed=24 -pad backup.iso

rm -rf backup-$TODAY backup.iso
```

INFO

[1] Afio: <http://directory.fsf.org/sysadmin/backup/afio.html>

[2] Amanda: <http://www.amanda.org>

[3] Klient systemu Amanda dla Windows: <http://sourceforge.net/projects/amanda-win32/>

Linux Magazine

Newsletter

Dowiedz się wcześniej,
co będzie w kolejnym numerze
Linux Magazine

<http://www.linux-magazine.pl/Newsletter>