Optymalizacja procesu archiwizacji dla administratorów i nie tylko

Opisuje Windows, Linux, UNIX i OS X

Archiwizacja i odzyskiwanie danych

HELION

O'REILLY®

W. Curtis Preston

Tytuł oryginału: Backup & Recovery

Tłumaczenie: Piotr Pilch (wstęp, rozdz. 1 – 16),
Marek Pętlicki (rozdz. 17 – 24)

ISBN: 978-83-246-5971-5

Wydawnictwo HELION
ul. Kościuszki 1c, 44-100 GLIWICE
tel. 032 231 22 19, 032 230 98 63
e-mail: *helion@helion.pl*
WWW: *http://helion.pl* (księgarnia internetowa, katalog książek)

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
http://helion.pl/user/opinie?odzyda_ebook
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Pliki z przykładami omawianymi w książce można znaleźć pod adresem:
ftp://ftp.helion.pl/przyklady/odzyda.zip

Printed in Poland.

Książka jest dedykowana
tym dzielnym kobietom i mężczyznom,
którzy poświęcili się w służbie ojczyźnie.

„Nikt nie ma większej miłości od tej,
gdy ktoś życie swoje oddaje
za przyjaciół swoich".

— Ewangelia św. Jana 15:13

Spis treści

Przedmowa

Mam nadzieję, że Czytelnik dowie się chociaż połowy tego, czego ja się dowiedziałem, pisząc tę książkę. Był to dość interesujący projekt, w ramach którego oryginalną książkę rozbudowaliśmy tak bardzo, że trzeba było zmienić jej tytuł. Siedem lat temu napisałem książkę *Unix Backup and Recovery*. Od tego czasu wiele się zmieniło, zarówno w branży, jak i w moim życiu. Największą zmianą w przemyśle komputerowym było rozpowszechnienie się w centrach danych komputerów z systemami Windows i Mac OS, a także serwerami Exchange i SQL Server (nigdy nie spotkałem się z serwerem Apple Xserve).

Dla mnie największą zmianą było bliższe zainteresowanie się aplikacjami archiwizującymi i odtwarzającymi niezaliczanymi do „tradycyjnych". Prawdą jest, że większość mojej kariery zawodowej spędziłem jako konsultant dla dużych firm, które wydawały na oprogramowanie i sprzęt archiwizujący ilość pieniędzy wystarczającą, żeby sfinansować niewielką armię. Cieszy mnie to, czym się zajmuję. Bardzo motywujące jest pokazanie firmie, jak może zaoszczędzić miliony dolarów rocznie, i jednocześnie sprawić, że procesy archiwizowania i odtwarzania danych będą szybsze i bardziej niezawodne (nawiasem mówiąc, jeśli Czytelnik potrzebuje pomocy w związku z używanym systemem archiwizującym, może wysłać wiadomość pod adres *curtis@backupcentral.com* — właśnie w ten sposób zarabiam na życie).

Sporo czasu zajęło mi też podróżowanie po świecie i rozmawianie z użytkownikami na temat radzenia sobie z archiwizowaniem i odtwarzaniem danych. Zawsze kierowano do mnie następujące pytania:

- Dostałem ofertę z firmy XYZ na oprogramowanie archiwizujące. Firma chce za nie XXXX złotych! Skąd mam niby zdobyć taką sumę pieniędzy?
- Nie mogłem sobie pozwolić na oprogramowanie archiwizujące firmy XYZ, dlatego nabyłem aplikację firmy ABC, która nie spełnia oczekiwań. Czy może Pan polecić coś lepszego?
- Żadne z komercyjnych narzędzi nie potrafi zarchiwizować mojej bazy danych MySQL lub PostgreSQL. Jak to można zrobić?
- W jaki sposób przeprowadzić proces przywracania od podstaw w przypadku systemu operacyjnego ABC?
- Czy są dostępne jakieś narzędzia open source, które umożliwią wykonanie tego typu rzeczy?

W czasie, gdy właściwie przygotowywałem się do pisania swojej kolejnej książki, dotyczącej metod wybierania komercyjnych aplikacji archiwizujących, ich instalowania i zarządzania nimi, stwierdziłem, że ta książka musiała powstać jako pierwsza. Kierowana jest do osób, które uznały, że komercyjne produkty archiwizujące nie spełniają wszystkich ich wymagań.

Być może Czytelnik ma niewielki sklep, który może przeznaczyć 10 000 złotych na przyzwoite oprogramowanie archiwizujące. Być może Czytelnik już korzysta z komercyjnej aplikacji archiwizującej, lecz nie zamierza wydać tysięcy złotych na agenta tworzącego kopię zapasową baz danych DB2 lub nie może znaleźć kogoś, kto będzie archiwizować bazy danych MySQL lub PostgreSQL. Książka ma na celu zaoferować opcje — *darmowe*.

Niemal wszystko, co omówiłem w książce, jest dołączone do systemu operacyjnego lub aplikacji bądź dostępne w ramach projektu open source (opisywane przeze mnie komercyjne produkty kosztują tylko 99 dolarów). Czytelnik może być zaskoczony tym, co może zrobić za darmo lub prawie darmo.

Chciałbym mieć taką książkę

Chciałem napisać książkę, która zapewni, że nikt już nie będzie musiał zaczynać czegoś od początku. Wierzę, że właśnie to udało się osiągnąć mnie i moim współpracownikom. W książce przedstawiłem każde narzędzie archiwizujące, którym chciałbym dysponować, gdy wkraczałem w branżę archiwizacyjną. Ponadto zawarłem w niej wszystkie lekcje i sztuczki, których dotąd się nauczyłem. Książka wyjaśnia, jak archiwizować i odtwarzać dane, począwszy od zwykłych stacji roboczych z systemem Linux, Windows lub Mac OS, a skończywszy na złożonych bazach danych, takich jak DB2, Oracle i Sybase. Niezależnie od tego, czy dostępny budżet ledwie starcza na pokrycie kosztu nośnika archiwizacyjnego, czy pozwala na zakup silosu większego od mieszkania, w książce Czytelnik znajdzie coś dla siebie. Niezależnie od tego, czy zadaniem Czytelnika jest stwierdzenie, jak zarchiwizować dane bez użycia komercyjnych narzędzi w środowisku, z jakim miałem do czynienia, gdy zaczynałem karierę, czy wybranie jednego z ponad 50 komercyjnych produktów archiwizujących, książka podpowie, jak postępować. Mając to na uwadze, pozwolę sobie wspomnieć o kilku istotnych rzeczach związanych z książką.

Liczy się tylko odtwarzanie danych

Mój przyjaciel Joe Fitzpatrick mówił mi coś takiego: „Nikogo nie obchodzi to, czy możesz zarchiwizować dane, lecz wyłącznie to, czy jesteś je w stanie odtworzyć". Ile Czytelnik przeczytał rozdziałów na temat archiwizowania, w których mniej niż 10% miejsca poświęcono procesowi odtwarzania? Ta książka jest inna. Starałem się z całych sił, żeby zagadnienie odtwarzania danych zostało poruszone w równym stopniu.

Produkty zmieniają się

Niektórzy mogą być zaskoczeni tym, że w części książki poświęconej komercyjnym aplikacjom archiwizującym nie podano żadnych ich nazw. Powodów jest kilka. Podstawowy jest taki, że produkty nieustannie się zmieniają. Byłoby niemożliwe aktualizowanie w książce informacji dotyczących ponad 50 produktów dostępnych dla samego systemu Unix. Książka zdezaktualizowałaby się już w momencie, w którym trafiłaby do księgarń. W książce objaśniono *pojęcia* związane z komercyjnymi aplikacjami archiwizującymi i odtwarzającymi dane. Dzięki temu Czytelnik może pojęcia te skonfrontować z deklaracjami składanymi obecnie przez producentów. Aktualne informacje na temat konkretnych produktów są dostępne pod adresem *http://www.backupcentral.com*.

Archiwizowanie baz danych nie jest takie trudne

Jeśli Czytelnik jest administratorem baz danych, może nie być zaznajomiony z poleceniami wymaganymi do zarchiwizowania bazy. Jeżeli jest się administratorem systemu, można nie mieć wiedzy na temat architektury platformy bazodanowej zarządzanej przez administratora baz danych. Obie kwestie zostaną szczegółowo omówione w książce. Prostym językiem objaśniam narzędzia archiwizujące, tak aby było to zrozumiałe dla administratora baz danych. Ponadto omawiam architekturę baz danych w taki sposób, żeby nie miał z nią problemów administrator systemu, nawet taki, który nigdy nie widział na oczy bazy danych.

Przywracanie komputera od podstaw nie jest takie trudne

Wcześniej lub później zdarzy się, że po tym, gdy utraci się dysk z systemem operacyjnym ważnego komputera i stanie przed koniecznością odtwarzania, cały dzień będzie trzeba poświęcić na *przywracanie komputera od podstaw*. Standardowa metoda odtwarzania opisana w dokumentacji wielu produktów archiwizujących polega na zainstalowaniu systemu operacyjnego w minimalnej wersji i przywróceniu danych. Przywracanie komputera od podstaw w ten sposób jest najgorszym możliwym wariantem. Jednym z licznych problemów będzie to, że nadpisze się część plików systemowych, gdy system zostanie załadowany z tego samego dysku, na którym próbuje się odtworzyć dane. W książce szczegółowo omówiono najlepsze metody przywracania komputera od podstaw w przypadku takich systemów operacyjnych, jak AIX, Solaris, HP-UX, Windows, Linux i Mac OS.

Struktura książki

Książkę podzielono na sześć części:

Część I

W pierwszej części książki zamieszczono wystarczającą ilość informacji, żeby zachęcić Czytelnika do rozszerzania wiedzy na temat archiwizowania i odtwarzania danych.

Rozdział 1. „Filozofia archiwizacji”
: W rozdziale zaprezentowałem swoją filozofię archiwizacji, w tym odpowiedź na pytanie, dlaczego powinno się tworzyć kopie zapasowe, a także wstępne informacje na temat metod archiwizowania.

Rozdział 2. „Archiwizowanie wszystkich danych”
: W rozdziale drobiazgowo omówiono zasadnicze składniki dobrego systemu archiwizowania i odtwarzania danych.

Część II

W tej części książki przedstawiono podstawowe dostępne narzędzia umożliwiające archiwizowanie komputera i kilka systemów archiwizujących open source ułatwiających zarządzanie kopiami zapasowymi.

Rozdział 3. „Podstawowe narzędzia do archiwizacji i odtwarzania”
W rozdziale zaprezentowano podstawowe narzędzia archiwizujące i odtwarzające, które Czytelnik prawdopodobnie znajdzie w takich systemach, jak Unix, Windows i Mac OS. Te narzędzia to m.in. *dump*, *tar*, *cpio*, *dd*, *ditto*, *ntbackup* i *rsync*.

Rozdział 4. „Amanda”
W rozdziale omówiono ciągle popularne oprogramowanie *Amanda* (*Advanced Maryland Disk Archiver*).

Rozdział 5. „BackupPC”
W rozdziale objaśniono system archiwizujący *BackupPC* bazujący wyłącznie na dyskach. Narzędzie to może archiwizować znacznie więcej niż tylko zawartość komputera PC.

Rozdział 6. „Bacula”
W rozdziale omówiono oprogramowanie *Bacula*, które nocą przenosi dane z centrum danych i korzysta z najważniejszych zasobów sprzętowych komputerów.

Rozdział 7. „Narzędzia open source służące do niemal ciągłej ochrony danych”
W rozdziale przedstawiono trzy produkty niemal ciągłej ochrony danych, czyli narzędzie *rsync* z migawkami, a także programy *rsnapshot* i *rdiff-backup*.

Część III

Jeśli możliwości darmowych narzędzi nie są wystarczające lub po prostu chciałoby się skorzystać z nowych technologii archiwizowania i odtwarzania, trzeba będzie przyjrzeć się komercyjnemu produktowi. Aby móc ocenić pełny zakres opcji archiwizacji i odtwarzania, należy mieć również wiedzę na temat najnowszego sprzętu dostępnego na rynku.

Rozdział 8. „Komercyjne narzędzia archiwizujące”
Rozdział pełni rolę przewodnika po setkach funkcji dostępnych w ponad 50 komercyjnych produktach archiwizujących znajdujących się obecnie w sprzedaży. Dzięki temu rozdziałowi będzie można podjąć w pełni świadomą decyzję zakupu.

Rozdział 9. „Urządzenia archiwizujące”
W rozdziale objaśniono wiele różnego typu dostępnych obecnie urządzeń archiwizujących i podano kryteria ułatwiające zdecydowanie, który rodzaj napędu archiwizującego będzie najbardziej odpowiedni.

Część IV

Odtwarzanie komputera od podstaw jest najszybszą metodą przywrócenia do życia komputera po wystąpieniu awarii, nawet wtedy, gdy napęd systemu operacyjnego został całkowicie zniszczony.

Rozdział 10. „Przywracanie od podstaw komputera z systemem Solaris”
W rozdziale omówiono produkt Flash Archive firmy Sun dołączony do systemu Solaris, będący odpowiednikiem narzędzia *mksysb* systemu AIX.

Rozdział 11. „Systemy Linux i Windows”
W rozdziale wyjaśniono kilka procedur i narzędzi, które mogą być zastosowane do przywracania od podstaw komputerów z systemami Linux i Windows. Omówiono też produkt open source o nazwie Ghost for Linux (G4L), który sporządza obrazy.

Rozdział 12. „Przywracanie od podstaw komputera z systemem HP-UX”
W rozdziale przedstawiono narzędzia *make_net_recovery* i *make_tape_recovery*, które obecnie są dołączane do systemu HP-UX, aby można było przywrócić od podstaw komputer z tym systemem.

Rozdział 13. „Przywracanie od podstaw komputera z systemem AIX”
W rozdziale omówiono narzędzie *mksysb* systemu AIX, które prawdopodobnie jest jednym z najstarszych i najlepiej znanych programów do przywracania komputera od podstaw.

Rozdział 14. „Przywracanie od podstaw komputera z systemem Mac OS X”
W rozdziale wyjaśniono, jak przeprowadzić proces przywracanie od podstaw komputera z systemem Mac OS X.

Część V

W tej części książki prostym językiem objaśniono dziedzinę stawiającą kilka z największych wyzwań związanych z archiwizowaniem i odtwarzaniem, z którymi może mieć do czynienia administrator systemu lub bazy danych. Dziedziną tą jest archiwizowanie i odtwarzanie baz danych.

Rozdział 15. „Archiwizowanie baz danych”
W rozdziale omówiono architekturę baz danych, kojarząc każdy jej element z odpowiednim terminem stosowanym w przypadku takich baz, jak DB2, Exchange, Informix, MySQL, Oracle, PostgreSQL, SQL Server i Sybase. Rozdział okaże się przydatny dla kogoś, kto jest administratorem systemu i boi się baz danych lub uczy się obsługi nowej bazy.

Rozdział 16. „Archiwizowanie i odtwarzanie bazy danych Oracle”
W rozdziale wyjaśniono, jak wykonywać „gorące” kopie zapasowe bazy danych Oracle przy użyciu narzędzia *rman* lub przeprowadzać proces archiwizacji zarządzany przez użytkownika.

Rozdział 17. „Archiwizowanie i odtwarzanie bazy danych Sybase”
W rozdziale pokazano, jak za pomocą serwera archiwizującego tworzyć kopie zapasowe bazy danych Sybase ASE.

Rozdział 18. „Archiwizowanie i odtwarzanie bazy danych IBM DB2”
W rozdziale omówiono archiwizowanie i odtwarzanie baz danych DB2.

Rozdział 19. „SQL Server”
W rozdziale wyjaśniono, jak archiwizować i odtwarzać bazy danych SQL Server.

Rozdział 20. „Exchange”
W rozdziale pokazano, w jaki sposób archiwizować i odtwarzać bazy danych serwerów Exchange przy użyciu wbudowanego dodatku narzędzia *ntbackup*.

Rozdział 21. „PostgreSQL”
W rozdziale przedstawiono procesy archiwizowania i odtwarzania baz danych PostgreSQL.

Rozdział 22. „MySQL”
W rozdziale dokonano przeglądu różnych opcji archiwizacji i odtwarzania dostępnych dla serwera bazodanowego MySQL.

Część VI

Informacje zamieszczone w tej części książki w żadnej mierze nie są mało istotne. Po prostu nie mogły znaleźć się w innym miejscu!

Rozdział 23. „Różności"
: W rozdziale omówiono oprogramowanie VMware, przedstawiono często poruszaną kwestię zrzutów aktywnych systemów plików, a nawet zamieszczono trochę poezji o archiwizowaniu danych.

Rozdział 24. „Ochrona danych"
: Rozdział zmusza do poważnych przemyśleń. Czytelnik znajdzie w nim odpowiedź na pytanie, dlaczego kopie zapasowe nie rozwiązują wszystkich problemów. Czytelnik powinien też zastanowić się nad innymi obszarami podlegającymi ochronie danych, takimi jak tworzenie archiwów, przywracanie po awarii i zabezpieczanie magazynów danych.

Co nowego można znaleźć w książce?

Aby poznać odpowiedź na to pytanie, należy przeczytać poprzedni podrozdział. W rzeczywistości w porównaniu z książką *Unix Backup & Recovery* nowy materiał to około 75% całości. Niektóre rozdziały oryginalnej książki zostały napisane całkowicie od nowa. Oto najważniejsze ze zmian:

Nowe podejście
: Książka odzwierciedla moje nowe podejście do archiwizowania danych, które jest oparte na wykorzystaniu dysku (dotyczy to zwłaszcza mniejszych firm).

Nowe polecenia archiwizujące
: W rozdziale poświęconym podstawowym narzędziom uwzględniono programy: *ntbackup*, *ditto* i *rsync*.

Amanda
: Całkowicie uaktualniono rozdział opisujący oprogramowanie Amanda, aby uwzględnić w nim postępy dokonane w ostatnich siedmiu latach.

Komercyjne narzędzia
: Zaktualizowano rozdział poświęcony komercyjnym narzędziom w celu uwzględnienia dokonań z dziedziny archiwizowania i odtwarzania danych z ostatnich siedmiu lat.

System HP-UX
: Ponieważ zmieniły się narzędzia *make_net_recovery* i *make_tape_recovery*, tak też się stało z omawiającym je rozdziałem.

Urządzenia archiwizujące
: Ależ przez siedem lat zmienił się sprzęt! Docelowe magazyny dyskowe, wirtualne biblioteki taśmowe i systemy z deduplikacją danych. Wszystko to zostało opisane.

W książce jest 11 *zupełnie nowych rozdziałów*, które znacznie rozszerzają jej zakres. Oto zagadnienia przedstawione w tych rozdziałach:

DB2
: Metoda archiwizowania bazy danych DB2 przy użyciu jej wbudowanych funkcji.

Exchange
: Metoda archiwizowania serwera Exchange za pomocą programu *ntbackup*.

SQL Server
: Metoda tworzenia kopii zapasowej danych serwera SQL Server przy użyciu jego wbudowanych funkcji.

MySQL
: Metoda archiwizowania i odtwarzania baz danych MySQL przy wykorzystaniu mechanizmów magazynowania: MyISAM, InnoDB i NDB.

PostgreSQL
: Metoda archiwizowania i odtwarzania tej popularnej darmowej bazy danych przy użyciu narzędzia *pg_dump* lub *pg_dumpall*.

BackupPC
: Objaśniono obsługę oprogramowania BackupPC, czyli kompletnego dyskowego systemu archiwizowania i odtwarzania wyposażonego w internetowy interfejs.

Bacula
: Omówiono użytkowanie darmowego produktu archiwizującego Bacula, który nocą przenosi dane z centrum danych i korzysta z najważniejszych zasobów sprzętowych komputerów.

Niemal ciągła ochrona danych
: Wyjaśniono, jak używać migawek i replikacji w celu archiwizacji danych.

System Solaris
: Omówiono przywracanie od podstaw komputera za pomocą narzędzia Flash Archive.

Przywracanie od podstaw komputera z systemami Linux i Windows
: Wyjaśniono, jak przy użyciu dysku Linux Live CD lub narzędzia Ghost for Linux przeprowadzić proces przywracania od podstaw komputera z systemami operacyjnymi Windows i Linux.

System Mac OS X
: Przedstawiono użycie wbudowanej funkcji przywracania od podstaw komputera z systemem operacyjnym Mac OS X (nie jest to zbyt trudne).

Co zostało pominięte?

Z różnych powodów w książce nie uwzględniono niektórych rozdziałów książki *Unix Backup & Recovery*. Wszystkie te rozdziały są obecnie dostępne pod adresem *http://www.backupcentral.com*. Oczywiście problemem jest to, że rozdziały te nie zostały uaktualnione. W związku z tym umieszczono je na stronie internetowej, aby każdy, kto chce, mógł pomóc w ich zaktualizowaniu.

- „Tru64 Bare-Metal Recovery”,
- „IRIX Bare-Metal Recovery”,
- „Informix Backup and Recovery”,
- „Clearcase Backup and Recovery”,
- „High Availability”.

Coś na temat witryny BackupCentral.com

Całkowicie przebudowaliśmy witrynę znajdującą się pod adresem *http://www.backupcentral.com*, używając systemu zarządzania treścią, forów i technologii MediaWiki. Moim celem numer jeden jest sprawić, aby łatwiej było zapewnić na niej dynamiczną treść. Zależy mi również na zbudowaniu silnej społeczności związanej z dziedziną archiwizowania i odtwarzania danych. Nowa witryna Backup Central ma do zaoferowania kilka świetnych rzeczy:

- Fora phpBB poświęcone różnym tematom związanym z archiwizowaniem. Jedno z forów dotyczy tej książki. Zachęcam do przyłączenia się do dyskusji.
- Każde forum dysponuje listą wysyłkową umożliwiającą śledzenie dyskusji za pośrednictwem forum lub poczty elektronicznej. Wszelkie posty zamieszczone na forum są przekazywane do listy wysyłkowej. Wiadomości pocztowe wysłane na adres listy wysyłkowej spowodują, że na forum pojawią się posty lub odpowiedzi.
- Wielokierunkowe połączenie między grupami dyskusyjnymi Usenet, listami wysyłkowymi i forami phpBB związanymi z archiwizowaniem. Jedną z rzeczy, które uświadomiłem sobie w czasie pisania książki, było to, że Usenet nadal funkcjonuje i ma się dobrze. Zależy mi na przedstawieniu tego znakomitego, choć niedocenianego zasobu społeczności związanej z witryną Backup Central, a także na utworzeniu innego portalu wykorzystującego Usenet. Każda interesująca mnie grupa dyskusyjna Usenet ma listę wysyłkową i forum. Wszystkie wiadomości kierowane do grupy dyskusyjnej Usenet, listy wysyłkowej lub forum trafiają na odpowiednie forum, listę wysyłkową lub grupę dyskusyjną.
- Strony Wiki bazują na tym samym oprogramowaniu MediaWiki, które obsługuje witrynę Wikipedia. Jedną z rzeczy, które Czytelnik znajdzie w witrynie Backup Central, jest strona Wiki dla każdego rozdziału książki. Strony te posłużą do aktualizowania i rozszerzania zagadnień poruszanych w książce. Na takie rozwiązanie zdecydowałem się z dwóch powodów:
  - Z pisaniem specjalistycznych książek wiąże się następujący problem: ledwie książka trafi do druku, pojawiają się w niej zmiany. W czasie gdy książka była przygotowywana do druku, serwer MySQL zaoferował kolejne trzy mechanizmy magazynowania danych, narzędzie QTParted miało zacząć obsługiwać system plików NTFS, a serwer dla systemu Windows oprogramowania Bacula miał stać się bardziej dostępny. Aby uwzględnić te zmiany, posłużymy się stronami Wiki.
  - Ja i moi współpracownicy nie znamy wszystkich odpowiedzi. Dołożyliśmy wszelkich starań, żeby zaoferować Czytelnikowi solidną książkę. Jednak nie mieliśmy do czynienia z wszystkim, z czym mógł mieć do czynienia Czytelnik. Bylibyśmy zadowoleni, gdyby Czytelnik pomógł nam rozwinąć tematy poruszone w książce, objaśnić sytuacje, w których określona procedura nie działa, lub przedstawić metody, przy użyciu których procedura powinna zostać zmodyfikowana (przykładowo mój przyjaciel próbował mi pomóc w zrozumieniu, jak sprawić, żeby program *rsync* lepiej radził sobie z milionami plików; ponieważ przeprowadzane przez niego testy nie zostały ukończone przed oddaniem książki do druku, poprosiłem Jasona o zawarcie wyników na stronie Wiki). Niech Czytelnik zajrzy na przebudowaną stronę internetową Backup Central i ocali świat lub przynajmniej jego dane.

Do zobaczenia pod adresem *http://www.backupcentral.com*!

Konwencje zastosowane w książce

W książce użyto następujących konwencji typograficznych:

Kursywa
: Przy użyciu tego stylu są wyróżnione nowe terminy, adresy URL, tytuły rozdziałów i książek, a także nazwy i rozszerzenia plików, katalogów, komputerów i programów.

`Tekst o stałej szerokości`
: Styl ten wyróżnia polecenia do wykonania, zmienne i opcje.

`Kursywa o stałej szerokości`
: Za pomocą tego stylu są wyróżniane łańcuchy zawarte w wierszach poleceń, w miejsce których należy wstawić rzeczywistą nazwę.

`Pogrubiona kursywa o stałej szerokości`
: Tym stylem są wyróżnione łańcuchy znajdujące się w wierszach poleceń, w miejsce których należy wstawić rzeczywistą nazwę.

`Listing`
: Stylem tym są identyfikowane wyniki wygenerowane przez polecenia.

`Pogrubiony listing`
: Przy użyciu tego stylu wyróżnia się wykonywane polecenia.

Książka jest efektem pracy zespołowej

To prawda. Choć na mojej tabliczce z nazwiskiem widnieje, że jestem specjalistą od archiwizowania i odtwarzania danych, nie oznacza to, że wiem wszystko na ten temat. W rzeczywistości nigdy nawet nie miałem do czynienia z częścią systemów operacyjnych lub baz danych omówionych w książce! Byłoby to oznaką lekceważenia Czytelnika, gdybym sam napisał rozdziały poświęcone produktom, których nie używałem. Jednak zależało mi na uwzględnieniu takich rozdziałów w książce. W związku z tym wynająłem zespół ekspertów, którzy napisali te rozdziały. W przybliżeniu 250 stron jest autorstwa innych osób. Współautorzy zostali przedstawieni na początku rozdziałów, w których tworzeniu brali udział.

Osoby, które brały udział w pisaniu książki

Nie jest prosto napisać rozdział w cudzej książce. Nie tylko trzeba umieć pisać, trzeba także robić to zgodnie z założeniami autora całości. Ponadto terminy są krótkie i cały proces tworzenia to mieszanka pośpiechu i czekania. Nie byłbym w stanie napisać książki bez wsparcia innych osób. W związku z tym pozwolę sobie oficjalnie podziękować wszystkim moim współpracownikom.

Amanda
: W tworzeniu tego rozdziału udział brali Dmitri Joukovski i Stefan G. Weichinger. Dziękuję za ukończenie tego rozdziału.

BackupPC
: Współtwórcą tego rozdziału jest Don „Kaczka” Harper. Kwa!

Bacula
: W pisaniu tego rozdziału brał udział Adam Thornton. Dzięki za uwzględnienie w książce narzędzia *Bacula*.

Narzędzia open source służące do niemal ciągłej ochrony danych
: Współautorami tego rozdziału są: Michael Rubel, Ben Escoto i David Cantrell. Rozdział kilka razy się zmieniał. Doceniam Waszą cierpliwość, gdy ostateczny kształt rozdziału układał mi się w głowie.

Przywracanie od podstaw komputera z systemem AIX
: W pisaniu rozdziału udział brał Mark Perino. Myślę, że jesteś najszybszym pisarzem w całym zespole.

Przywracanie od podstaw komputera z systemem HP-UX
: W pracach nad rozdziałem udział brali Eric Stahl i Ron Goodwyn. Znakomita współpraca, chłopaki.

Systemy Linux i Windows
: Współautorem tego rozdziału jest Reed Robins. Być może możemy to zrobić w ten sposób, w tamten sposób lub jeszcze inny! Czy wystarczająco zmieniłem zakres rozdziału? Dzięki.

Przywracanie od podstaw komputera z systemem Mac OS X
: W pracach nad rozdziałem uczestniczył Leon Towns-von Stauber. Dziękuję, Leon. Bez wątpienia jestem zadowolony z tego, że Mario namówił mnie, żebym do Ciebie zadzwonił. Ten rozdział jest perfekcyjny.

Przywracanie od podstaw komputera z systemem Solaris
: W pisaniu rozdziału udział brał Aaron Gersztoff. Aaron, uważaj, o co prosisz, dobrze?

Archiwizowanie i odtwarzanie bazy danych IBM DB2
: Współautorami tego rozdziału są: Jeff Richardson, Kulvir S. Bhogal i Kondal Yennaram. Wiele przeszliście i jestem za to niezmiernie wdzięczny.

Exchange
: W pracach nad tym rozdziałem uczestniczył Scott Harris. Więcej rysunków! Mniej rysunków! Zrób to w ten sposób! Nie, zrób to w taki sposób! Czyż nie jest fajne pisanie dla mnie?

SQL Server
: Współtwórcą tego rozdziału jest Scott Harris. Scooter, spójrz na to! Jesteś jedyną osobą, która była na tyle szalona, żeby napisać dla mnie dwa rozdziały. Dzięki.

Archiwizowanie i odtwarzanie bazy danych Sybase
: W pisaniu rozdziału udział brał Edward Barlow, który zaktualizował rozdział oryginalnie stworzony przez Bryn Smith. Następny współpracownik, bez którego nie mógłbym się obejść. Dzięki.

Oprogramowanie VMware i inne różności
: Współtwórcą tego rozdziału jest David Young. Kiedy wyprowadzisz się od swojej mamy?

Bez pomocy tych osób książka zawierałaby znacznie mniej informacji niż teraz.

Redaktorzy techniczni

Następna grupa osób, której muszę podziękować, to techniczni recenzenci. Było ich *wielu*! Problem z pisaniem książki poruszającej tak wiele zagadnień polega na tym, że trzeba korzystać z usług specjalistów i technicznych recenzentów. Z tego powodu większość redaktorów technicznych sprawdzała tylko jeden lub dwa rozdziały. Nie mógłbym się bez nich obejść. Choć jestem pewien, że kilka osób pominąłem, dołożyłem wszelkich starań, żeby wymienić je wszystkie (w kolejności alfabetycznej według imienia).

Adrin Kow
Andy Shellam
Anthony Johnson
Axel Schwenke
Ben Garrett
Brian Eliassen
Brian Peasland
Charles Whealton
Chris Thomas
Christoph Haas
Craig Barratt
D.A. Morgan
Dana Diederich
Daniel Callahan
Dave Mehler
David Boyd
Edward Conba
Eric Gilmore
Eric Stahl
Finn Henningsen
Frank Sweetser
Greg Lehey
Ian Gorrie
Ian Herd
James Bougor
Jayesh Thakrar
Jeff Badger
Jeff Frost
Jeff Harbert
Jeff Richardson
Jeffrey P. Humes
John Haight
John Hurley
John Madden
Kern Sibbald
Kumar Sundaram
Lenz Grimmer
Marcel Lans
Mark D. Powell
Mark Dawson
Mark Perino
Massimiliano Daneri
Matthew Huff
Megan Restuccia
Mike Harrold
Mohammed Mehdi
Neal A. Lucier
Norbert Munkel
Patrick Matthews
Paul Muggeridge
Ralph R. Hirtler
Rob Worman
Rodrigo Real
Satyaprakash Pandey
Scott Boss
Scott Harris
Shane Seymour
Simon Riggs
Steve Hanson
Stewart Smith
Tammy Bednar
Todd Toles
Víctor A. Rodríguez
Vitalis Jerome
Wil Coulbourn
William Cole

Przerażające historie

Osoby, które przyczyniły się do zamieszczenia w książce przerażających historii, sprawiły, że stała się bardziej interesująca. Jeśli nawet nie mogłem wykorzystać w książce historii jakiejś osoby, chcę jej podziękować za przesłanie.

Brian O'Neill
Brian Sakovitch
Chris Pritchard
David Bregman
David J. Young
Harry Tirrell
Hywel Matthews
Jack Coats
James Hunt
Jason Frankovitz
Jason Shupe
Jim Damoulakis
Jim Donnellan
John Merryman
Jorgen Lie
Karl Langdon
Kevin Suttle
Mark Perino
Michael Rice
Michael Tobin
Natalie Meek
Richard Ackerman
Scott Boss
Theo Van Dinter
William Birch
William S. Duncanson

Specjalne podziękowania

Było kilka osób, które okazały się niezmiernie pomocne w ten lub inny sposób podczas realizowania projektu. Chciałbym im szczególnie podziękować.

Anthony Johnson
Nie każdy prezes zarządu w ramach wolontariatu podjąłby się zrecenzowania rozdziału, który właściwie był poświęcony darmowej aplikacji konkurującej z produktem jego firmy. Mam nadzieję, że zarówno Ty, John, jak i Twoja firma macie się bardzo dobrze. Firma oferuje przyzwoite narzędzie służące do przywracania od podstaw komputerów z systemami Linux i AIX.

Brian Peasland
Skrytykowałeś kiepską jakość pierwszej roboczej wersji rozdziału poświęconego bazie danych Oracle — i słusznie. Nowa wersja rozdziału jest *znacznie* lepsza dzięki Twojej dokładnej i szczerej recenzji (sprawiłeś, kolego, że ponownie napisałem połowę rozdziału!).

Deb Cameron
Czyż nie jest fajne wydawanie książki mającej 18 autorów z trzech kontynentów, kilku stref czasowych i posługujących się różnymi językami, a także około 60 redaktorów technicznych? Zróbmy kiedyś coś takiego jeszcze raz!

Joshua D. Drake
Joshua, dziękuję, że poświęciłeś czas na te wszystkie rozmowy przez telefon, które miały pomóc mi zrozumieć serwer PostgreSQL. Następnym razem powinieneś być bardziej bezpośredni w wyrażaniu własnych opinii. Naprawdę nie wiadomo, jakie w danej chwili jest Twoje zdanie.

Lenz Grimmer
Byłeś moim przewodnikiem w społeczności związanej z serwerem MySQL. Zdecydowanie potrzebowałem pomocy. Dziękuję Tobie i całemu zespołowi zajmującemu się serwerem MySQL.

Lynn Stone
Dziękuję Ci za to, pomogłaś mi we wstępnej fazie realizacji projektu. Bez Ciebie nie byłoby to możliwe. Tylko ja znam Twoją sekretną tożsamość.

Tammy Bednar
Będąc tak zajętą kobietą, zapewniłaś mi dokładnie to, co było mi potrzebne do części projektu dotyczącej bazy danych Oracle. Oczywiście wykonałaś znacznie więcej pracy związanej z innymi produktami firmy Oracle, co zresztą widać. Mam nadzieję, że zauważysz w rozdziale poświęconym bazie danych Oracle moje wyrazy uznania dla Twoich osiągnięć.

Zmanda
Dziękuję za pomoc przy pisaniu rozdziału poświęconego oprogramowaniu Amanda i zapewnienie profesjonalnego wsparcia bardzo popularnemu narzędziu archiwizującemu open source.

Mario Obejas
Dziękuję Ci bardzo za skierowanie mnie do Leona.

Nie wiem wszystkiego

Jeśli czegoś się nauczyłem, pisząc tę książkę, to tego, że nie wiem wszystkiego na temat archiwizowania danych. Jeżeli Czytelnik zna lepszy sposób na wykonanie czegoś opisanego w książce, nauczył się jakichś specjalnych sztuczek lub napisał jakieś fajne narzędzia, które według niego mogą innym osobom pomóc w archiwizowaniu i odtwarzaniu danych, proszę o kontakt. Można mi wysłać wiadomość pocztową na adres *curtis@backupcentral.com*. Metody lub narzędzia zaproponowane przez Czytelnika mogą zostać uwzględnione w następnym wydaniu książki i od razu udostępnione pod adresem *http://www.backupcentral.com*.

Jak mogę podziękować?

Jak zacząć podziękowania kierowane do setek osób, które mi pomogły?

Do Boga: Oby każdy wyraz uznania dla tej książki trafił tylko do Ciebie.

Do mojej żony Celynn: Mówię „dziękuję" za wiele nocy, które spędziłaś sama, gdy ja korespondowałem, stukając w klawisze, z kimś mieszkającym w innym miejscu globu. Jesteś wyjątkową kobietą, która nigdy nie walczyła ze mną lub moim marzeniem. Kocham Cię.

Do mojej córki Niny: Miałaś zaledwie siedem lat, gdy pojawiła się moja pierwsza książka. Obecnie jesteś piękną młodą damą, która tak szybko dorasta. Zamierzam zdobyć broń i siedzieć na ganku.

Do mojej córki Marissy: Miałaś dopiero dwa lata, gdy wyszła moja pierwsza książka. Teraz jesteś śliczną dziewięciolatką. Jak ten czas płynie. Wybierzmy się razem do parku na rower.

Do moich rodziców: Cóż mogę powiedzieć? Zawsze we mnie wierzyliście i mówiliście mi: „Nieważne, czy będziesz kopał rowy. Po prostu bądź najlepszym kopaczem rowów na świecie". Cóż, będąc specem od archiwizacji, w branży komputerowej nie można bardziej zbliżyć się do kopacza rowów. Dlatego napisałem o tym książkę.

Podziękowania dla Boba Walkera za to, że pomógł mi znaleźć pierwszą pracę związaną z kopiami zapasowymi, a także dla Rona Rodrigueza za to, że aż zbyt chętnie dał mi tę pracę.

Wyrazy wdzięczności dla Susan Davidson, która nie zwolniła mnie, gdy w 1992 r. nie mogłem odtworzyć handlowej bazy danych. Ta druga szansa była wszystkim, co było mi potrzebne, aby stać się ekspertem od archiwizacji, którym jestem obecnie. Jeślibyś mnie wtedy zwolniła (jestem pewien, że kilka osób tego chciało), kto wie, gdzie obecnie bym był (jeżeli jesteś zainteresowana tą historią, zapoznaj się z ramką „Coś, z czym mi się udało" w rozdziale 1.).

Dziękuję firmie Collective Technologies za to, że pomogła mi poprawić umiejętności na tyle, abym stwierdził, że chcę się specjalizować w archiwizowaniu i odtwarzaniu danych. Dziękuję również za wspieranie mnie, gdy pisałem pierwszą książkę.

Podziękowania składam na ręce następujących osób: Jason Stege, Robin Young, Jeff Williams, Reed Robins i Elia Harris. Dziękuję za wiarę we mnie, gdy zakładałem własną firmę. Mam nadzieję, że według Was postąpiłem właściwie.

Wyrazy wdzięczności dla Marka Shirmana i wszystkich przyjaciół z GlassHouse za zapewnienie mi miejsca, w którym wreszcie poczułem, że korzystam ze swoich umiejętności.

Do rodziców mojej żony: Dziękuję za tak wspaniałą żonę, a także za traktowanie mnie jak własnego dziecka i wspieranie nas w poszukiwaniach. *Pahingi ng sinagang?*

Podziękowania dla wszystkich nauczycieli, którzy usilnie starali się wydobyć ze mnie pełnię możliwości — wreszcie się to udało.

Do wydawnictwa O'Reilly: Dziękuję za możliwość wydania tej bardzo potrzebnej książki.

Do moich redaktorów, Deb Cameron i Michaela Loukidesa: Musimy się wreszcie spotkać pewnego dnia! Nie wiem, jak Wy to robicie, że nieustannie czytacie tę samą książkę bez zmęczenia oczu. Jesteście znakomitymi redaktorami. Naprawdę mogę powiedzieć, że daliście z siebie wszystko w czasie realizowania projektu. Dziękuję, dziękuję i jeszcze raz dziękuję (nie edytujcie *tego* zdania, dobrze?).

Do Czytelnika: Dziękuję za nabycie książki. Mam nadzieję, że Czytelnik dowie się z niej tyle co ja, gdy ją pisałem.

CZĘŚĆ I

Wprowadzenie

Część I składa się z dwóch rozdziałów:

Rozdział 1. „Filozofia archiwizacji”
W rozdziale wyjaśniono, dlaczego powinno się archiwizować dane, a także w skrócie pokazano, jak to robić.

Rozdział 2. „Archiwizowanie wszystkich danych”
W rozdziale szczegółowo przedstawiono zasadnicze elementy dobrego systemu archiwizowania i odtwarzania danych.

ROZDZIAŁ 1.

Filozofia archiwizacji

Archiwizuję, a zatem będę.

Gdy patrzę na tytuł tego rozdziału, przypomina mi się stary występ Steve'a Martina, podczas którego powiedział z filozoficzną nutą, że właśnie ktoś dowiedział się wystarczająco, aby być nieszczęśliwym przez resztę swojego życia (Steve analizował ważne pytania, takie jak: „Czy w porządku jest wykrzyknięcie słowa »film« w remizie pełnej ludzi?". Obiecuję, że nie będę tego robić). Jednakże stwierdzenie „Filozofia archiwizowania" wydało mi się właściwe jako tytuł tego rozdziału, w którym wyjaśnię, *dlaczego* należy archiwizować dane (wspomnę oczywiście też o tym, *jak* to robić).

Rozbudowana archiwizacja przy niskim budżecie

Dobry system archiwizowania i odtwarzania danych jest kluczowy dla firmy dowolnej wielkości. Niestety, dział informatyczny nie zawsze dysponuje wymaganym budżetem i na system archiwizowania prawie nigdy nie są przeznaczane wystarczające środki. Jeśli uzna się, że potrzeba bardzo dobrego systemu archiwizowania, lecz nie posiada się odpowiednich środków na jego uruchomienie, trzeba wiedzieć, że książka ta przyda się właśnie w tego typu sytuacji. Konieczna jest rozbudowana archiwizacja przy niewielkim budżecie. Jeśli w przypadku Czytelnika tak właśnie jest, witam w klubie.

Skromny budżet nie oznacza jeszcze, że trzeba zrezygnować ze sporządzania kopii danych. Większość systemów archiwizacji zaprezentowanych w książce może być zastosowana w niewielkich środowiskach. Koszt instalacji uwzględniającej sprzęt wyniesie kilkaset złotych.

Klienci instytucjonalni nie muszą mieć powodów do obaw, ponieważ też znajdą w książce wiele interesujących rzeczy. Im więcej wykorzysta się rozwiązań przedstawionych w książce, tym więcej środków finansowych zaoszczędzi się na potrzeby innych projektów informatycznych. Mam nadzieję, że w czasie, gdy Czytelnik wdroży wszystkie zawarte w niej pomysły, ukończę pracę nad moją następną, równie interesującą książką. Zostaną w niej przedstawione komercyjne rozwiązania w dziedzinie ochrony danych, uwzględniające wieloplatformowe systemy archiwizowania i odtwarzania, ciągłą ochronę danych, niemal ciągłą ochronę danych, systemy archiwizacji danych bazujące na deduplikacji, replikację itp.

Gdy już Czytelnik dotarł do tego miejsca, może zadać następujące pytania:

- Dlaczego powinno się przeczytać tę książkę?
- Czy naprawdę można archiwizować dane za pomocą oprogramowania open source?
- Dlaczego powinno się użyć dysku?
- Dlaczego w ogóle powinno się archiwizować dane?
- Jak określić optymalną metodę archiwizowania?

Pora na odpowiedzi na wymienione pytania.

Dlaczego powinno się przeczytać tę książkę?

Jeżeli ktoś przez jakiś czas zajmował się administrowaniem systemem, może zadać sobie takie pytanie. Istnieje na nie wiele odpowiedzi. Być może instynkt samozachowawczy jest podstawowym czynnikiem motywującym. Chciałoby się mieć pewność, że nie straci się efektów pracy, gdy dojdzie do awarii dysku. Być może przyzwoity system archiwizowania już funkcjonuje i po prostu chciałoby się go ulepszyć. Być może Czytelnik szuka jakichś nowych pomysłów, jak poradzić sobie z kolejnymi potrzebami związanymi z archiwizowaniem i odtwarzaniem. Poniżej zamieściłem kilka powodów, dla których *według mnie* warto przeczytać tę książkę.

Schadenfreude

Schadenfreude jest niemieckim słowem na określenie radości z niepowodzenia innych. Właśnie dlatego ludzie oglądają w internecie te dziwne materiały wideo, na których jakiś przygłup próbuje zrobić coś niemądrego i w efekcie sam sobie robi krzywdę. Każda z ramek zawartych w książce przedstawia prawdziwą i groźną historię, która przydarzyła się komuś, kogo znam. Nie są to miejskie legendy lub makabryczne historie krążące między administratorami. Są to katastrofalne w skutkach zdarzenia, o których wiem z pierwszej ręki. Oczywiście czytaniu tych historii towarzyszy zachowanie opisywane słowem *schadenfreude*. Jednak każda historia niesie w sobie wniosek, który nie jest jedynym powodem, dla którego ją przytaczam. Rzeczy, o których ostrzegam w książce, naprawdę się wydarzyły. Ponieważ w przypadku braku przygotowania archiwizowanie danych może być bardzo utrudnione, należy z uwagą czytać książkę. Na początek można zapoznać się z zawartością ramki „Coś, z czym mi się udało" zamieszczonej w dalszej części rozdziału. Ramka przedstawia ważny epizod z mojej kariery.

Nigdy nie chciałbym znów wypowiedzieć tych słów

„Utraciliśmy dane z zaledwie kilku dni". W ramce „Coś, z czym mi się udało" napisałem, że straciliśmy dane z tylko kilku dni. Przeklinam dzień, w którym wypowiedziałem te słowa. Nigdy nie chciałbym ich użyć ponownie. Wtedy utwierdziłem się w przekonaniu o ważności kopii zapasowych. Nigdy ponownie czegoś nie będę przyjmował za pewnik. Zacząłem studiować wszystko na temat technologii archiwizowania danych. Książka odzwierciedla podjętą przeze mnie próbę zebrania w jednym miejscu tego, czego nauczyłem się o tanich metodach archiwizowania danych. Książka została napisana, aby nikt nie musiał już wypowiadać słów rozpoczynających ten akapit. Uważam, że *żadna strata danych nie jest do przyjęcia*. Założę się, że trudno byłoby znaleźć użytkownika, który miałby zdecydowanie inne zdanie na ten temat. Niezależnie od tego, czy jest to arkusz kalkulacyjny utworzony przez jedną osobę, czy baza danych klientów z fakturami sprzedaży z okresu wielu godzin lub dni, będąca efektem pracy

setek osób, należy dowiedzieć się od osoby używającej danych, jaka skala utraty danych będzie akceptowalna. Każde stwierdzenie, opinia, historia i rozdział w tej książce bazuje na przesłance, że każda utrata danych jest nie do przyjęcia. Poniżej dodatkowo zamieściłem bardzo ważne stwierdzenie.

Coś, z czym mi się udało

„Chcesz mi powiedzieć, że bezpowrotnie utraciliśmy kopie zapasowe *paryz* coś tam?". Nigdy nie zapomnę tych słów. Odpowiadałem za kopie zapasowe zaledwie od dwóch miesięcy i właśnie dowiedziałem się, że moja kariera dobiegła końca. Sześć tygodni wcześniej przenosiliśmy aplikację Oracle z jednego serwera na drugi. W procesie było coś, o czym zapomniałem — coś kluczowego. W tamtym czasie niewiele wiedziałem o archiwizowaniu baz danych i nie zdawałem sobie sprawy, że przed sporządzeniem bazy danych Oracle musiałem ją wyłączyć. Zadanie to było realizowane na starym serwerze przez narzędzie *cron*, o którego istnieniu nie miałem pojęcia. O wszystkim tym dowiedziałem się *po tym*, jak dysk nowego serwera zepsuł się.

„Po prostu przekaż nam ostatnią pełną kopię zapasową" — powiedzieli. Zacząłem przeglądać moje dzienniki. Właśnie wtedy zacząłem zauważać błędy. „Żaden problem — pomyślałem. — Po prostu użyję starszej kopii zapasowej". Starsze dzienniki nie wyglądały wcale lepiej. Zapamiętale przeglądałem kolejne dzienniki, aż znalazłem taki, który wyglądał na poprawny. Został sporządzony ponad 6 tygodni wcześniej. Gdy udałem się po wolumin, uświadomiłem sobie, że firma stosowała 6-tygodniowy cykl rotacji. W efekcie potrzebny wolumin został nadpisany dwa dni wcześniej.

To było to! W tamtej chwili wiedziałem, że muszę poszukać innej pracy. Była to baza firmy z danymi dotyczącymi zakupów. Utracone dane firmy o wielomiliardowych obrotach pochodziły z zamówień złożonych w okresie mniej więcej dwóch miesięcy.

A zatem przekazałem szefowi wieści. Właśnie wtedy usłyszałem to zdanie: „Chcesz mi powiedzieć, że bezpowrotnie utraciliśmy kopie zapasowe *paryz* coś tam?". Czy nie jest zadziwiające, że nie zapomniałem nazwy kopii? Nie pamiętam żadnej innej nazwy systemowej występującej w tej firmie, ale w pamięci utkwiła mi nazwa kopii zapasowej. Poczułem się tak mały, że mógłbym zmieścić się wewnątrz kasety 4-milimetrowej taśmy. Na szczęście w tamtym czasie w firmie pracował administrator systemu, którego nie mogę nazwać inaczej, jak magikiem. Zepsuty dysk został wskrzeszony, a następnie odzyskano z niego dane. Dzięki temu zostały utracone dane z zaledwie kilku dni. Nasz dział musiał rozesłać po całej firmie informację o konieczności ponownego wprowadzenia do bazy wszystkich zamówień, które wpisano w ciągu ostatnich dwóch dni. Powinienem był oprawić w ramkę kopię tego komunikatu, aby mi przypominał, co może się zdarzyć, gdy swojej pracy nie będę traktował wystarczająco poważnie. Jednak nie było to konieczne, ponieważ treść informacji trwale utkwiła w moim umyśle.

Część recenzentów książki powiedziała coś takiego: „To naprawdę śmiałe! Pisze Pan książkę na temat kopii zapasowych i na sam początek opisuje, jak sam jedną z takich kopii utracił. Naprawdę trzeba do tego mieć odwagę!". Dlaczego postanowiłem o tym wspomnieć? Przez te wszystkie lata z licznych awarii to jedno wydarzenie utkwiło mi w pamięci. Być może jest tak dlatego, że właśnie wtedy jedyny raz niemal straciłem pracę. Gdyby nie cudowne działania wspaniałego administratora, Joe Fitzpatricka, moja kariera mogła się zakończyć na samym początku. Opisałem to zdarzenie z następujących powodów:

- Zmieniło kierunek mojej kariery.
- Związanych z nim jest kilka wartościowych lekcji, z których sporo wyniosłem (więcej o nich w dalszej części książki).
- Mógłbym go uniknąć, gdybym posiadał książkę taką jak ta.

Muszę przyznać, że było naprawdę przerażające.

Biorąc pod uwagę obecnie dostępną technologię, nie ma powodu, aby miało dochodzić do utraty jakichkolwiek danych. Warunkiem jest, *aby* w przypadku kopii zapasowych był spełniony wymóg właściwej staranności i utrzymania priorytetu.

Ciekawość oprogramowania archiwizującego open source

Zaledwie kilka lat temu kopie zapasowe można było wykonać za pomocą niewielu skryptów i narzędzi: *dump*, *tar*, *cpio* lub *ntbackup*. Zapotrzebowanie na średniej klasy komputery zwiększało się w skali astronomicznej, natomiast proporcjonalnie wzrastała liczba większych baz danych, napędów lub systemów plików, a także długich nazw plików i ścieżek. Pojawienie się dużych baz danych i systemów plików przyczyniło się do powstania sporego rynku dla komercyjnych narzędzi archiwizujących. W tamtym czasie stworzono jeden lub dwa tego typu produkty. Później w sprzedaży były dostępne produkty innych firm.

Część tych pierwszych produktów była po prostu narzędziami z graficznym interfejsem użytkownika i funkcją zarządzania woluminami, opartymi na istniejących programach archiwizujących. Narzędzia te zapewniały większą funkcjonalność. Część firm uznała, że ich własne produkty miały wiele ograniczeń, których nie można było wyeliminować. Firmy te postanowiły opracować własne (nawet zastrzeżone) metody archiwizacji danych i podjęły próbę usunięcia ograniczeń, z którymi nie mogły sobie poradzić produkty oparte na narzędziu *dump* lub *tar*.

W ostatnich latach zapotrzebowanie na scentralizowane archiwizowanie i odtwarzanie danych spowodowało pojawienie się kilku narzędzi open source służących do wykonywania kopii zapasowych i ich odtwarzania. Sześć spośród nich omówiono w książce. Rynek narzędzi archiizujących open source rozwinął się podobnie do segmentu produktów komercyjnych. Oryginalny program archiwizujący open source o nazwie Amanda bazuje na narzędziu wybranym przez użytkownika. Produkt BackupPC pozostawia dane w ich oryginalnym formacie, a narzędzie Bacula używa niestandardowego formatu zaprojektowanego w celu wyeliminowania ograniczeń programu GNU o nazwie *tar*.

Obecnie na rynku produktów archiwizujących open source dostępnych jest kilka narzędzi. Całkiem możliwe jest to, że jeden lub więcej produktów open source przedstawionych w książce spełni oczekiwania Czytelnika dotyczące archiwizowania i odtwarzania danych. Obecnie ta książka jako jedyna omawia wszystkie te narzędzia.

Warto zapoznać się z archiwizowaniem danych wykorzystującym dysk

Jeśli ktoś nie słyszał o archiwizacji opartej na dysku lub archiwizacji D2D2T (*disk-to-disk-to-tape*), najwyższa pora, aby wyłączył cyfrowy rejestrator wideo DVR (*Digital Video Recorder*) i sięgnął po jedno lub dwa fachowe czasopisma (oczywiście używany rejestrator DVR jest jedynie urządzeniem archiwizującym obraz telewizyjny na dysku; jeśli ktoś okazjonalnie sporządza taśmy magnetowidowe z zapisem urządzenia DVR, ma nawet do czynienia z archiwizacją D2D2T). Zastosowanie dysku w systemach archiwizacji i odtwarzania znacznie zwiększyło się kilka lat temu i naprawdę eliminuje wiele problemów.

W rozdziale 9. omówiono urządzenia archiwizujące i bardziej szczegółowo wyjaśniono, dlaczego dyski stały się bardzo atrakcyjnym nośnikiem archiwizacyjnym. Oto krótkie podsumowanie, dlaczego tak jest.

Koszt
: Najważniejszym powodem, dla którego dysk stał się wyjątkowo atrakcyjnym nośnikiem kopii zapasowych, jest to, że w ciągu kilku ostatnich lat koszt dysku znacznie się zmniejszył. Koszt macierzy dyskowej w rozsądnej cenie obecnie w przybliżeniu jest taki sam jak koszt biblioteki taśmowej o podobnej pojemności, wypełnionej nośnikami. Gdy pod uwagę weźmie się kilka rzeczy, które można zrobić w przypadku dysków, takich jak eliminowanie pełnych kopii zapasowych lub nadmiarowych plików, okażą się one jeszcze tańsze.

Niezawodność
: W przeciwieństwie do taśm dyski są bliskie rozwiązaniom, które nie są podatne na zewnętrzne zanieczyszczenia. Ponadto rzeczywisty nośnik dysku twardego w porównaniu z nośnikiem taśmowym jest właśnie *twardy*. W efekcie pojedynczy napęd dyskowy z założenia jest bardziej niezawodny od napędu taśmowego. Napędy dyskowe stają się jeszcze bardziej niezawodne po umieszczeniu ich w macierzy RAID.

Elastyczność
: Ogólnie mówiąc, napędy taśmowe mogą działać tylko z dwoma szybkościami: *zatrzymania* i *bardzo dużą*. Choć niektóre napędy taśmowe obsługują zmienne szybkości, zwykle mogą zwolnić do prędkości stanowiącej 40% nominalnej szybkości urządzenia. Z kolei napędy dyskowe działają z dowolną wybraną szybkością. Aby uzyskać szybkość transferu wynoszącą kilkaset megabajtów na sekundę, wystarczy w ramach grupy RAID zestawić kilka napędów. Ta sama grupa RAID nie będzie miała problemu z zapisaniem danych napływających z szybkością 10 kB/s. W przeciwieństwie do napędów taśmowych napędy dyskowe nie mają problemu z naprzemiennym wolnym i szybkim zapisywaniem. Sprawia to, że dysk idealnie sprawdza się w przypadku nieprzewidywalnych strumieni archiwizowanych danych. Po zapisaniu takich losowych danych w szeregowy sposób na urządzeniu dyskowym z łatwością można przenieść zarchiwizowane dane na taśmę (jeśli jest to wymagane). Część osób całkowicie rezygnuje z tego kroku i zastępuje go replikacją. Warto spróbować zrobić coś takiego przy wykorzystaniu napędu taśmowego.

Archiwizacja oparta na dyskach jest wyjątkowo ekonomicznym rozwiązaniem, oferującym małym i średnim firmom całkowicie zautomatyzowane tworzenie kopii zapasowych. Choć duża biblioteka taśmowa może być bardzo tania (kilka złotych za gigabajt) i skalowalna, nie zawsze jest to prawdą w przypadku mniejszych bibliotek przeznaczonych dla rynku niewielkich i średnich firm. Dużym wyzwaniem jest skalowalność. Zwykle im biblioteka taśmowa jest tańsza, tym mniejszą ma skalowalność (oczywiście zawsze zdarzają się wyjątki). Dla porównania, część zupełnie zautomatyzowanych produktów archiwizujących open source wymienionych w książce może być zastosowana razem z jednym napędem dyskowym kosztującym mniej niż 300 złotych. Jeśli jeden napęd dyskowy nie wystarcza, wystarczy nabyć następny dysk i dodać go do menedżera woluminów. Można również kupić kontrolery RAID, które pozwalają zacząć od pojedynczego dysku i wraz ze wzrostem potrzeb dodawać kolejne. W ten sposób można zwiększyć pojemność z kilkuset gigabajtów do wielu terabajtów.

Dlaczego należy archiwizować dane?

Wszystko już słyszałem. Byłem oskarżany o to, że zajmuję się tylko kopiami zapasowymi. Powiedziano mi, że uważam, że cały świat obraca się wokół szpulki kasety. Stwierdziłem, że pewnego dnia świat spotka katastrofa i pozostanie mi kopia zapasowa. Pytanie brzmi: jak bardzo poważnie *Czytelnik* traktuje kwestię ochrony danych? Aby ułatwić rozstrzygnięcie tej kwestii, zastanówmy się, co się stanie, gdy nie będą dostępne poprawne kopie zapasowe.

Nie to miałem na myśli!

Administrowałem grupą testerów większej firmy tworzącej oprogramowanie. Podczas instalacji aplikacja tworzyła katalog *$HOME/foo*, żeby umieścić w nim dane użytkownika. Tester dbający o jakość oprogramowania przeprowadzał instalację z uprawnieniami superużytkownika. Aplikacja utworzyła katalog *$HOME/foo*. *$HOME* było dosłownie nazwą katalogu. Tester przesłał informację o błędzie i postanowił usunąć bezwartościowy katalog. Czytelnik prawdopodobnie domyślił się, że wykonano następujące polecenie:

```
# rm -rf $HOME
```

Polecenie zostało wykonane w standardowym systemie uniksowym, w przypadku którego dla superużytkownika w miejsce zmiennej `$HOME` nadal była wstawiana ścieżka /.

Gdy już przestałem się śmiać, udałem się po nośnik instalacyjny, aby ponownie zainstalować system (niestety, nie był dostępny żaden obraz systemu ani kopie zapasowe dla różnych serwerów testerów). Na szczęście większość krytycznych danych znajdowała się na serwerze NFS.

— William Birch

Ile będzie kosztować utrata danych?

Aby udzielić odpowiedzi na to pytanie, trzeba zastanowić się nad typem archiwizowanych danych. Jest to właściwa pora na skontaktowanie się z osobami, które wcale nie muszą uważać się za ekspertów od komputerów. Aby otrzymać odpowiedź, należy uzyskać informacje z innych działów. Biorąc pod uwagę, że dane mają postać zer i jedynek, o jakim właściwie typie informacji mowa? Czy stosuje się ręczne metody rozliczania lub rekordy finansowe firmy są przechowywane w jakimś programie księgowym? Czy gdy klient zadzwoni i złoży zamówienie, zapisuje się je na formularzu z kalką czy wprowadza dane do odpowiedniego programu przetwarzającego zamówienia? A jak wygląda to w przypadku takich rzeczy, jak budżety, zestawienia, inwentarze i innego rodzaju papierowe dokumenty, które przekłada się każdego dnia? Czy przechowuje się kopię każdej ważnej wysłanej notatki lub zadanie to spoczywa na komputerze?

Jeśli Czytelnik żyje jak większość ludzi, w znacznym stopniu jest zależny od tego, co określa się mianem komputerów. Nie pamięta już, ile efektów jego pracy zostało zapisanych w postaci niewielkich magnetycznych bitów na różnych obracających się talerzach. Być może pracuje w środowisku, w którym nigdy nie doszło do awarii dysku. W związku z tym nigdy nie musiał odtwarzać danych. Być może nigdy nie uszkodził klucza lub nie usunął ważnego pliku. Jeśli to prawda, powinien zapamiętać, co mówił mój ojciec: „Można wyróżnić dwie grupy motocyklistów: tych, co się przewrócili, i tych, których dopiero to czeka". To samo dotyczy napędów dyskowych. Jeżeli komuś nigdy nie zepsuł się dysk, zapewniam, że do tego dojdzie!

A zatem co się straci w przypadku utraty danych? Aby to stwierdzić, trzeba przyjrzeć się typom informacji, które mogą znajdować się w środowisku roboczym, a także zastanowić, co by się stało, gdyby wszystkie te dane przepadły. Większość tego, co można stracić, jest bardzo namacalna i da się określić w pieniądzu. To może zaskakiwać.

Utraceni klienci
: Całkiem możliwe, że spośród wszystkich strat ta jest najbardziej widoczna i dotkliwa. Jeśli cała baza danych klientów znajduje się na jakimś komputerze, jak będzie można poznać, kto jest kim, gdy dojdzie do jego awarii? Właściwie można utracić klientów i nigdy ich ponownie nie odszukać. Można również utracić klientów, którzy bazują na danych zlokalizowanych na jednym lub większej liczbie komputerów należących do firmy świadczącej usługi. Jeżeli klient dowie się, że przepadły jego dane, bez wątpienia nie będzie pod wrażeniem firmy, która do tego dopuściła. Może być tak, że utrata danych nie dotknie klienta w sposób bezpośredni. Jednak uzna, że skoro firma utraciła jego dane, może po prostu zakończyć z nią współpracę, ponieważ okazała się niekompetentna.

Zamówienia
: Niezależnie od usługi lub produktu oferowanego przez firmę istnieją metody rejestrowania zamówień dotyczących usługi lub produktu. Przeważnie jest to metoda wykorzystująca komputery. Utrata danych może oznaczać przepadnięcie zamówień złożonych w ciągu godzin, dni, a nawet tygodni. Mogą to być zamówienia, z którymi jest związana ciężka praca sprzedawców!

Morale
: Czytelnik może pomyśleć, jakby się czuł, gdyby był jednym ze sprzedawców, którego zamówienia przepadły. Spędził w dziale sprzedaży wiele dni lub tygodni, a teraz okazało się, że zamówienia zostały bezpowrotnie utracone. Być może powinien udać tam, gdzie ciężka praca nie pójdzie na marne. Im lepszy sprzedawca, tym większa szansa na to, że może opuścić firmę, gdy dojdzie do utraty jego danych. A jak będzie w przypadku przeciętnego pracownika? Jeśli reputacja komputerów jest taka, że mają przestoje i tracą dane, pracowników ogarnia poczucie bezradności. Być może powinni przenieść się tam, gdzie otrzymają sprzęt pozwalający normalnie pracować.

Reputacja
: A co z reputacją w branży? Wieści dotyczące poważnej utraty danych niewątpliwie szybko się rozniosą. Mogą dotrzeć do konkurencji, która na pewno przy pierwszej sposobności wykorzysta je dla własnych korzyści. Mogą też trafić do organizacji nadzorującej firmę. Na przykład dla organu nadzorującego bank niezbyt miła byłaby informacja, że administrator banku przyczynił się do poważnej utraty danych. Taka organizacja może naprawdę dokładnie przyjrzeć się poczynaniom administratora. Nikt nie marzy o czymś takim!

Budżet
: Wystarczy jeden przypadek utraty danych, aby dział komputerowy firmy zyskał złą reputację. Choć zła reputacja może przez jakiś czas pozostać, ze wszelkich sił trzeba próbować ją naprawić. Jest się tak dobrym jak ostatni proces odtwarzania danych (mój przyjaciel powiedział, że jest się tak dobrym jak *najgorzej* przeprowadzony proces odtwarzania). Jeżeli ludzie nie ufają wykonanym kopiom zapasowym, będą sami je sporządzać, powielając działania administratora. Pracownicy będą tracić czas i pieniądze na lokalne archiwizowanie swoich systemów. Każda osoba może zdecydować się na kupno własnego napędu i oprogramowania archiwizującego, a nawet przynieść z domu samodzielnie utworzony

skrypt. Wykonywanie przez pracowników kopii zapasowych w najlepszym razie będzie nieefektywne i kosztowne, a w najgorszym przyczyni się do dalszej utraty danych. Gdy każdy we własnym zakresie zajmuje się archiwizowaniem, można stracić sporo pieniędzy na opłacenie ludzi i zakup dodatkowego sprzętu.

Czas
: Ile osób obsługuje komputery? Ile pracy pójdzie na marne, gdy w zaprojektowanym środowisku dojdzie do utraty danych? Znam wiele firm mających kilku kontraktowych programistów nieustannie piszących kod. Jeśli przepadnie część kodu przechowywanego w systemie, o jakiej stracie finansowej będzie mowa? W rzeczywistości jest tak, że gdy w dowolnym dziale firmy wyniki pracy są przechowywane na komputerze i dochodzi do utraty danych, kosztuje to dużo czasu i pieniędzy.

Jaki będzie koszt przestoju?

Planując procedurę archiwizowania i odtwarzania danych, można wybierać spośród kilku opcji mających wpływ na szybkość odtwarzania. Im szybciej proces ten przebiega, tym bardziej kosztowny jest system archiwizowania. Przed podjęciem decyzji dotyczącej takich opcji trzeba zadać sobie następujące pytanie: „Jak kosztowny będzie przestój?". Gdy o tym myślę, przypomina mi się reklama kopiarki, która brzmiała tak: „Gdy kopiarka przestanie działać, wystarczy powiedzieć ludziom, że nie ma żadnego problemu, ponieważ po prostu użyje się papieru z kalką!". Jeżeli przestanie funkcjonować jeden z głównych systemów, czy użytkownicy nadal będą mogli kontynuować pracę, czy może cała firma przejdzie w stan spoczynku? Jeśli wystąpi drugi wariant, czy pracownicy będą w dalszym ciągu opłacani? Wtedy wysłanie ich do domu nie da żadnych oszczędności. Oto kilka godnych uwagi dodatkowych kosztów:

Postrzeganie przez klientów
: Klient nie znosi słuchać takich stwierdzeń, jak „Proszę zadzwonić ponownie, ponieważ nasze komputery obecnie nie są sprawne" lub „Połączenie nie jest aktywne". Zależnie od profilu działalności firmy klienci mogą po prostu udać się do konkurencji. Im dłużej systemy są wyłączone, tym więcej klientów usłyszy powyższe zdania.

Postrzeganie przez pracowników
: Nikt nie chce pracować w firmie, w której komputery nieustannie nie działają. Im więcej pracowników jest zależnych od komputerów, tym bardziej stwierdzenie to jest prawdziwe. Jeśli Czytelnik byłby sprzedawcą, który przez dzień lub dłużej nie może połączyć się z bazą danych, jak bardzo byłby zadowolony?

Czas
: Znów ma miejsce strata czasu. W efekcie nie ma postępów, a pracownicy zależni od dostępności wyłączonego systemu w rzeczywistości są opłacani za nic.

Znalezienie optymalnej metody archiwizowania

Używanie systemu bez kopii zapasowych można przyrównać do prowadzenia samochodu z prędkością 160 km/h po ruchliwej drodze dzień po wygaśnięciu polisy ubezpieczeniowej. Podobnie zastosowanie 3-węzłowego, intensywnie wykorzystywanego klastra na potrzeby mało istotnej aplikacji jest jak pełne ubezpieczenie wykupione dla 20-letniego samochodu. Tak jak polisy mają różne poziomy ubezpieczenia (na przykład kierowców rajdowych szczególnie

dotyczą różnego typu uszkodzenia), tak różne metody archiwizowania zapewniają różne poziomy odtwarzania danych.

Było blisko

Pamiętny moment miał miejsce, gdy dysponowaliśmy serwerem plików o pojemności 600 GB, który przez jakiś czas nie był poprawnie archiwizowany. Podczas wyjątkowo upalnego weekendu zepsuły się oba klimatyzatory w pomieszczeniu i w efekcie wzrosła temperatura. Oczekując na naprawienie klimatyzacji, wszystko wyłączyliśmy, a następnie, gdy sprzęt się ochłodził, rozpoczęliśmy archiwizowanie danych. Jak można było się spodziewać, zepsuły się dwa dyski znajdujące się obok siebie w tej samej macierzy RAID4. Z trudem udało nam się uniknąć całkowitej utraty danych dzięki temu, że znaleźliśmy wolny dysk i wymieniliśmy układy scalone między nim i jednym z uszkodzonych napędów. Dzięki temu uruchomiliśmy dysk i uzyskaliśmy do niego dostęp. Następnego dnia producent dostarczył nam nowe dyski, po czym spędziliśmy mnóstwo czasu na naprawianiu serwera kopii zapasowych.

— Theo Van Dinter

Nie należy przesadzać

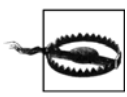
Nie wszystkie środowiska wymagają odtwarzania danych z dokładnością co do minuty. W przypadku wielu środowisk akceptowalna jest możliwość odtworzenia danych systemów zawartych w kopiach zapasowych wykonanych ostatniej nocy. W określonych środowiskach dopuszczalne jest przywrócenie systemu nawet z kopii sporządzonej w zeszłym tygodniu lub miesiącu. Wydanie tysięcy złotych i spędzenie setek godzin na wdrażaniu najwspanialszego na świecie rozwiązania archiwizującego będzie bezsensowne, jeśli nie jest wymagany taki poziom zabezpieczeń. Tego typu problem nie występuje w przypadku większości organizacji. Z kolei znaczna część organizacji nie przeznacza wystarczających środków lub nie dokłada odpowiednich starań, aby zapewnić systemy archiwizowania i odtwarzania. Jednak w innych sytuacjach pieniądze mogą zostać zmarnotrawione na niepotrzebnie rozbudowane systemy.

W obrębie jednej firmy zmieniają się też wymagania dotyczące możliwości odtwarzania danych na poszczególnych komputerach. Wymagania te mogą być determinowane przez ilość pracy, która może zostać zaprzepaszczona, lub przez możliwość zniechęcenia klienta. Przykładowo za dopuszczalną przez jednego lub dwóch pracowników może być uznana utrata wyników całodniowej pracy w edytorze tekstu. Jeśli dotyczyłoby to asystenta wiceprezydenta, pracującego nad budżetem departamentu, zakres tolerancji mógłby wyglądać inaczej. Prawdopodobnie zupełnie nie do zaakceptowania będzie utrata wpisów wprowadzanych do firmowej bazy danych sprzedaży, używanej przez setki osób, nawet jeśli chodzi o wpisy z ostatniej godziny.

Chodzi o to, że *wymagania dotyczące archiwizowania są określane przez wymogi stawiane procesowi odtwarzania danych*. Trudność tkwi w znalezieniu i zastosowaniu narzędzia będącego w stanie zapewnić oczekiwany poziom odtwarzania. Weźmy pod uwagę katalogi domowe użytkowników. Jeżeli są one przechowywane lokalnie na stacji roboczej każdego użytkownika, utrata zapisanych na dysku efektów pracy z minionego popołudnia oznaczałaby zmarnowanych kilka godzin. Jeśli jednak katalogi użytkowników znajdują się na serwerze plików NFS obsługującym tysiące użytkowników, mogą pójść na marne tysiące godzin, gdy zastosuje się tylko tradycyjne narzędzia archiwizujące.

Jeśli niedopuszczalna jest utrata danych sieciowego serwera plików, pod uwagę można wziąć użycie technologii *obrazu* (ang. *snapshot*). Odpowiednie oprogramowanie umożliwia wykonanie obrazu napędu lub systemu plików w określonej chwili, a następnie zastosowanie tego obrazu do sporządzenia kopii zapasowej napędu lub systemu plików. Jeżeli kopia zapasowa napędu lub systemu plików zostanie wykonana za pomocą obrazu, będzie uwzględniać spójną zawartość napędu lub systemu plików z chwili utworzenia obrazu. Jeżeli takie rozwiązanie zainteresuje Czytelnika, może przeczytać rozdział 7., w którym opisano emulowanie technologii obrazu za pomocą narzędzia *rsync* i dowiązań twardych.

Choć czasami wymagane narzędzie dołączono do używanego systemu operacyjnego lub platformy bazodanowej, po prostu nie jest właściwie użytkowane. Czasami narzędzia archiwizujące nie są w ogóle stosowane. Jeśli na przykład dysponuje się produkcyjną bazą danych Oracle, połączenie gorących nocnych kopii zapasowych z archiwizowaniem dzienników powtórzeń zapewnia możliwość odtworzenia danych z dokładnością co do minuty. Jeśli jednak utraci się dysk wchodzący w skład bazy danych, która nie archiwizuje swoich dzienników transakcji, przepadną wszystkie dane od czasu wykonania ostatniej „zimnej" kopii zapasowej. Więcej informacji można znaleźć w części piątej książki.



Jeżeli dysponuje się produkcyjną instancją i nie używa funkcji rejestrowania transakcji mechanizmu bazodanowego, należy ją włączyć jak najszybciej!

A zatem nie wystarczy znaleźć właściwe narzędzie zapewniające wymagany poziom odtwarzania danych, trzeba także z niego odpowiednio korzystać.

Uzyskanie wymaganego poziomu zabezpieczeń

Niektóre środowiska nie mogą sobie pozwolić nawet na minutę przestoju i powinny być wyposażone w najlepszy system archiwizowania, niezależnie od jego ceny. Wynika to z rozmiaru straty, która nastąpi, gdy systemy będą niedostępne choćby przez krótki okres (wiem o jednej firmie, która twierdzi, że traci ponad milion dolarów na minutę, gdy systemy są wyłączone). A nawet jeśli pracuje się w środowisku, które pozwala na przestój, wydanie ogromnej ilości pieniędzy na natychmiast dostępny *„gorący" serwer*[1] jest zupełnym marnotrawstwem.

Przyjrzyjmy się tabeli 1.1. Nikt nie powinien być zależny od samochodu lub komputera, nie mając przynajmniej podstawowego poziomu zabezpieczenia. Jeśli jedyny samochód nie jest ubezpieczony i pijany kierowca spowoduje poważny wypadek, jak pokryć taką stratę? A co zrobić, gdy zepsuje się dysk twardy w systemie przechowującym krytyczne informacje i przepadną wszystkie dane? Część osób zapomina o tym, że ta zależność działa w dwie strony. Jeżeli ktoś nabył już trzeci samochód i ma on 20 lat (i wcale nie jest to model retro), prawdopodobnie wykupi dla niego wyłącznie ubezpieczenie OC. Można będzie żyć bez tego pojazdu w razie jego zniszczenia. Wydanie setek dodatkowych złotych rocznie na ubezpieczenie samochodu wartego 150 zł nie ma po prostu sensu. Podobnie, jeżeli zarządzane komputery znajdują się w środowisku, w którym można się bez nich obyć przez kilka dni, czy naprawdę będą potrzebne

[1] „Gorący" serwer to komputer znajdujący się w stanie oczekiwania, który może natychmiast rozpocząć proces przywracania środowiska.

lustrzane napędy z możliwością wymiany podczas pracy? Dla środowiska należy określić właściwy poziom ochrony.

Tabela 1.1. Porównanie ubezpieczenia samochodu i ochrony danych

Typ zabezpieczenia	Ubezpieczenie samochodu	Komputerowe kopie zapasowe
Minimalne zabezpieczenie	Kolizje i odpowiedzialność cywilna (chroni jedynie przed utratą koszulki, gdy spowoduje się wypadek).	• Regularne nocne wykonywanie kopii zapasowych (chroni przed utratą wyników pracy, gdy napęd dyskowy zepsuje się).
Nieoczekiwane nieszczęśliwe zdarzenia	Pełne zabezpieczenie (wandalizm, boskie wyroki itp.).	• Rejestrujące systemy plików. • Zasilacze awaryjne UPS (*Uninterruptible Power Supply*).
Możliwość natychmiastowego prowadzenia samochodu	Zabezpieczenie dotyczące wynajęcia samochodu (otrzymuje się inny pojazd, gdy własny na skutek wypadku trafił do naprawy).	• Macierz RAID. • Kopie lustrzane. • Zastosowanie napędów wymienianych podczas pracy. • System HA (*High-Availability*) o wysokim stopniu dostępności.
Poważne nieszczęśliwe zdarzenia	Inna firma przejmie polisę i wymieni samochód, jeśli zarówno on, jak i dotychczasowa firma ubezpieczeniowa zostały zniszczone w wyniku trzęsienia ziemi.	• Wysłanie kopii zarchiwizowanych woluminów do zewnętrznej lokalizacji, aby zabezpieczyć się przed sytuacją, gdy zostanie zniszczone zarówno pomieszczenie komputerowe, jak i biblioteka nośników. • Wysłanie kopii zapasowych za pośrednictwem dedykowanej sieci do dużego systemu magazynowania znajdującego się po stronie dostawcy usług składowania.
Maksymalna ochrona	Firma ubezpieczeniowa nie tylko zgodzi się na wcześniej wymienione warunki, ale również na umieszczenie tego samego modelu samochodu (będącego do pełnej dyspozycji) w innej lokalizacji, gdy wszystkie takie pojazdy w miejscu zamieszkania lub siedzibie firmy zostały zniszczone.	• Wykonywanie kopii lustrzanych w czasie rzeczywistym na systemie umożliwiającym wymianę nośników podczas pracy i znajdującym się w innym oddziale firmy. • Wysłanie kopii zapasowych za pośrednictwem sieci lub kuriera do dostawcy „gorących" serwerów.

Trzeba zrównoważyć koszt określonego rozwiązania archiwizującego z przewidywaną stratą finansową spowodowaną przestojem, przed którym rozwiązanie ma chronić. Dla przykładu załóżmy, że analizuje się dwa warianty archiwizowania. Pierwszy wariant uwzględnia wysyłanie kopii zarchiwizowanych woluminów do zewnętrznego dostawcy usług magazynowania (miesięczny koszt wynosi 500 zł). Drugi wariant polega na zastosowaniu natychmiast dostępnego komputera w stanie oczekiwania, umieszczonego w innym mieście. Komputer ten odbiera co minutę z serwera produkcyjnego replikowane dane. Przyjmijmy, że miesięczny koszt tego rozwiązania to 5000 zł.

Firma znajduje się w Utopii, gdzie dotąd nie miały miejsca żadne klęski żywiołowe, a dyski są w całości kopiowane. Stwierdzono, że koszt dziennego przestoju wyniósłby tylko 500 zł. Czy naprawdę warto wydać 60 000 zł rocznie na zabezpieczenie przed czymś, co prawdopodobnie nigdy nie nastąpi? Jeśli centrum danych przydarzy się coś katastrofalnego w skutkach, czy równie dobrze nie sprawdzą się jednodniowe kopie przechowywane na zewnątrz? Choć firma miałaby dodatkowy dzień przestoju lub trochę więcej, wcześniej stwierdzono, że na coś takiego może sobie pozwolić. W przypadku takiego środowiska prawdopodobnie znacznie bardziej odpowiednie byłoby rozwiązanie kosztujące rocznie 6000 zł.

Jednak czy zabezpieczono się przed wszystkim, przed czym należałoby się zabezpieczyć? Czy firma jest zlokalizowana w obszarze podatnym na klęski żywiołowe, a jeśli tak, to czy stosuje ochronę przed tego typu zdarzeniami? Być może pod uwagę trzeba wziąć innego typu magazynowanie na zewnątrz firmy. Jeśli dysponuje się bazą klientów wymagającą regularnego dostępu do danych znajdujących się na komputerach, czy zapewniono możliwość szybkiego odtworzenia na wypadek awarii? Być może z myślą o serwerach bazodanowych powinno się uwzględnić użycie „gorącej" lokalizacji lub przechowywanie kopii lustrzanych w wielu miejscach. Tabela 1.1 to przegląd różnych poziomów zabezpieczenia.

Dlaczego używa się terminu „wolumin" zamiast „taśma"?

Większość narzędzi archiwizujących pierwotnie stworzono z myślą o zapisywaniu danych na taśmie. W związku z tym książki i sieciowe podręczniki przeważnie omawiają archiwizowanie na taśmie. Jednak wiele osób umieszcza kopie zapasowe na dyskach CD, dyskach magneto-optycznych, a nawet napędach dyskowych. Tego typu nośniki mają wiele zalet, ponieważ pod względem działania przypominają bardziej napędy dyskowe niż taśmowe. Dostęp losowy do zarchiwizowanych danych jest prostszy, a ponadto można je odczytywać w blokach o dowolnym wybranym rozmiarze. Wynika to stąd, że napędy dyskowe, w przeciwieństwie do taśmowych, nie zapisują przerw między rekordami.

Ponieważ wiele osób nie korzysta już z taśm, w książce, gdy jest to właściwe, stosuje się bardziej ogólny termin *wolumin*. Zamiast terminu *napęd taśmowy* Czytelnik napotka też pojęcie *napęd archiwizujący*. To dlatego, że napędem archiwizującym może być zarówno nagrywarka dysków CD, jak i napęd dyskowy. W książce terminy *taśma* i *napęd taśmowy* są używane tylko wtedy, gdy jest to konieczne i właściwe.

Witryna BackupCentral.com oferuje dla każdego rozdziału książki stronę umożliwiającą internautom zamieszczanie własnych uwag i opinii. Pod adresem *http://www.backup➥central.com* można znaleźć zaktualizowane informacje lub dodać do nich własne.

ROZDZIAŁ 2.

Archiwizowanie wszystkich danych

Gdy już Czytelnik zapoznał się z filozoficznymi wywodami z rozdziału 1., pora przyjrzeć się niektórym istotnym pojęciom dotyczącym archiwizowania i odtwarzania danych. Dzięki nim będzie wiadomo, co uwzględnić w kopii zapasowej, kiedy przeprowadzać archiwizację danych itp.

Nie wolno pomijać tego rozdziału!

Zwykły Czytelnik może przyjąć, że niniejszy rozdział stanowi wprowadzenie do podstawowych zagadnień związanych z archiwizowaniem. Choć faktycznie jest to cel tego rozdziału, prawdziwe jest również to, że wielu doświadczonym administratorom przedstawione tu pojęcie są nieznane. Jest tak między innymi dlatego, że administratorzy nieustannie są odciągani od typowych zajęć, takich jak archiwizowanie, aby robić rzeczy uważane za „ważniejsze" (na przykład instalowanie nowych serwerów i stwierdzanie, które systemy wolno działają). Ponadto administratorzy nawet przez kilka lat mogą nie musieć odtwarzać danych. Konieczność regularnego stosowania kopii zapasowych bez wątpienia zmieniłaby sposób postrzegania ich ważności.

Napisałem tę książkę, ponieważ przez kilka lat archiwizacja i odtwarzanie danych stanowiły dla mnie podstawowy obszar działalności. Chciałbym przekazać wiedzę, którą zdobyłem w tym okresie. W niniejszym rozdziale pokrótce omówiłem, w jaki sposób powinny funkcjonować kopie zapasowe. Wyjaśniłem też wiele podstawowych, a zarazem wyjątkowo ważnych pojęć, na których powinien bazować każdy dobry plan archiwizowania. Na takim planie są oparte wszystkie wdrożenia przedstawione w książce.

Niewykonalne zadanie, którego nikt nie chce się podjąć

Czy ktokolwiek czytający tę książkę powie, że utrata danych jest czymś dobrym? Nie wierzę w to. A zatem dlaczego traktuje się kopie zapasowe tak mało poważnie? Czasami czuję się jak Rodney Dangerfield, gdy walczę o to, aby wykonywać lepsze kopie zapasowe — „Mówię Ci, że nie ma dla mnie za grosz szacunku". Kopie zapasowe często nie są uwzględniane podczas projektowania systemu. Czy gdy ktoś kupuje nowy serwer, pyta o jego wpływ na obecnie stosowaną metodę archiwizowania danych? Niektóre działy informatyczne nie kontrolują nawet procesu nabywania nowych systemów, ponieważ czasami są one kupowane przez inne centra, odpowiedzialne za finanse. Czy kiedykolwiek próbowano wyjaśnić kierownikowi innego działu, dlaczego jego terabajtowy serwer bazodanowy nie będzie archiwizowany na dołączonym niezależnym napędzie taśmowym używającym taśm o gigabajtowej pojemności?

Następną często pomijaną kwestią jest personel odpowiedzialny za archiwizację. Czy kiedykolwiek próbowano szukać osób, które będą zajmować się kopiami zapasowymi? Często jest to dodatkowy obowiązek, którym obarcza się różne osoby, podobnie gdy ja, moja siostra i brat ustalaliśmy, kto tym razem będzie zmywać naczynia. Jeśli szczęśliwie wyznaczono osobę odpowiedzialną za archiwizowanie danych, zwykle będzie to najmłodszy stażem pracownik firmy. Wiem o tym, ponieważ właśnie w ten sposób zacząłem moją pierwszą pracę. W rzeczywistości wiele osób tak właśnie zaczyna pracować. Jak można czemuś tak ważnemu nadać tak niski priorytet? Być może powinno się to zmienić. Czy jedna książka spowoduje zmianę tak długotrwałej tradycji zatrudniania pracowników? Prawdopodobnie nie, ale być może okaże się pomocna. W najgorszym razie, gdy ktoś odpowiedzialny za sporządzanie kopii zapasowych kupi tę książkę, będzie miał kompletny przewodnik pozwalający wykonać to naprawdę pokaźne zadanie.

Można zapytać, co w tym takiego pokaźnego? Dlaczego kopie zapasowe są tak istotne przy nowoczesnych systemach komputerowych i niezawodnych napędach dyskowych? Dlatego, że komputery w dalszym ciągu psują się. Ponadto firmy bardziej niż kiedykolwiek wcześniej ufają w ich niezawodność. Niezależnie od tego, jak dobry jest producent używanego systemu Unix lub jak niezawodne są napędy dyskowe, systemy nie będą zawsze sprawne. Nie pomoże tu nawet zatrudnienie Dogberta jako administratora sieci. W przypadku systemów komputerowych obowiązuje prawo Murphy'ego. Systemy nie tylko będą mieć sporadyczne awarie, ale też będzie to następować w porze najmniej odpowiedniej dla firmy i jej klientów. W takiej chwili — a ta chwila kiedyś nadejdzie — zadaniem osoby wykonującej kopie zapasowe będzie odtworzenie danych znajdujących się na dysku lub dyskach, które odmówiły posłuszeństwa. Typowe pytanie brzmi: „Ile to potrwa?". Jedyna dopuszczalna odpowiedź to: „Już zrobione".

Kto chciałby być tym, który nie poradził sobie z odtwarzaniem danych i w efekcie spowodował, że baza danych klientów była niedostępna przez dodatkowe trzy godziny? Kto chciałby być tym, który musi rozesłać po całej firmie informację o konieczności ponownego wprowadzenia wszystkich zamówień z ostatnich dwóch dni? Kto chciałby być tym, który będzie o tym myśleć codziennie podczas sprawdzania wyników operacji archiwizacji przeprowadzonej zeszłej nocy? Jeśli ktoś nie doprowadził do utraty danych, oznacza to, że po prostu robi to, co powinien. Jeżeli nie wykonuje rzetelnie swoich obowiązków, napotka na spore problemy. Kto chce dostać taką pracę? Z pewnością nikt.

Czytelnik kupił tę książkę, ponieważ ma niewykonalne zadanie, którego nikt nie chce się podjąć. Niezależnie od tego, czy archiwizowaniem danych Czytelnik zajmuje się od jakiegoś czasu, czy dopiero zaczął się tym zajmować, może stwierdzić, że jest to spore zadanie. Ilość danych jest ogromna, ich natura ciągle się zmienia, tymczasem dostępne narzędzia nigdy nie wydają się spełniać oczekiwań. Wiem to, ponieważ sam się z tym borykałem. Spędziłem wiele miesięcy, próbując wdrożyć „rozwiązania" oferowane przez systemy operacyjne i bazy danych, które nie były odpowiednio przygotowane. Spotkałem się z firmami wydającymi pieniądze na kosztowne komercyjne narzędzia tylko po to, aby zostać ze złym oprogramowaniem, niespełniającym ich wymagań. Widziałem instalacje nowszych i większych serwerów niemających nawet jednego napędu archiwizującego. Spędziłem również długie noce i weekendy w pomieszczeniach komputerowych, próbując odtworzyć dane w „rozsądnym" czasie. Niestety, termin „rozsądny" jest definiowany przez końcowego użytkownika, który nie ma pojęcia, jak trudnym zadaniem jest odtwarzanie danych.

Obecnie istnieją rozwiązania dla niemal każdego problemu dotyczącego kopii zapasowych. Dla niewielkiego sklepu z zaledwie kilkoma komputerami, z których każdy będzie działał pod kontrolą identycznego systemu operacyjnego, istnieje dobre rozwiązanie. Dla dużego sklepu z setkami komputerów z różnymi odmianami systemów Unix, Linux, Windows i Mac OS lub tylko kilkoma bazami danych o pojemności wielu terabajtów również dostępne jest odpowiednie rozwiązanie. Największym problemem jest dezinformacja. Większość osób po prostu nie wie, co jest dostępne. W związku z tym obchodzą się bez żadnego rozwiązania lub korzystają z gorszego (zwykle poleconego przez obrotnego sprzedawcę). Oto sześć ważnych pytań, które trzeba cały czas zadawać sobie i innymi:

Dlaczego?
: Dlaczego stosujemy ochronę przed awarią? Czy naprawdę stanie się coś złego, gdy dojdzie do utraty danych? Jakie będą tego konsekwencje? Jakiego typu dane mamy i jaka jest ich wartość?

Co?
: Co będzie archiwizowane? Zawartość całego komputera czy tylko wybrane napędy lub systemy plików? Jakie systemy operacyjne będą archiwizowane? Co jeszcze, poza standardowymi napędami lub systemami plików, powinno zostać uwzględnione w kopii zapasowej?

Kiedy?
: Kiedy jest najlepszy moment na archiwizowanie systemu? Jak często powinno się wykonywać pełną kopię zapasową? Kiedy powinno się sporządzać przyrostową kopię zapasową?

Gdzie?
: Gdzie zostanie umieszczona kopia zapasowa? Jakie jest najlepsze miejsce do przechowywania woluminów z kopiami zapasowymi?

Kto?
: Kto zapewni sprzęt, oprogramowanie i usługi instalacyjne, które połączą wszystko w jeden system?

Jak?
: W jaki sposób przeprowadzimy archiwizację? Istnieje kilka różnych metod ochrony przed utratą danych. Należy zapoznać się z odmiennymi rozwiązaniami, takimi jak przechowywanie poza obrębem określonej lokalizacji, replikacja, wykonywanie kopii lustrzanych, macierze RAID, a także różnymi poziomami ochrony oferowanymi przez poszczególne metody (każde z wymienionych zagadnień omówiono szczegółowo w dalszej części książki).

Dlaczego archiwizuje się dane?

Jeśli Czytelnik nie umie odpowiedzieć na to pytanie, naprawdę nie ma sensu, aby czytał dalej. Dobrą wiadomością jest to, że na to pytanie naprawdę łatwo odpowiedzieć. Wystarczy pomyśleć o wszystkim, co może przydarzyć się przechowywanym danym, i przyjrzeć się każdemu ich rodzajowi. Należy się zaznajomić z każdą jednostką firmy tworzącą dane, a także dowiedzieć się, jak na ich funkcjonowanie wpłynie utrata lub uszkodzenie danych. Wszystko to uzasadnia dalsze działanie.

Co należy archiwizować?

Doświadczenie pokazuje, że jednym z najczęstszych powodów utraty danych jest to, że nigdy ich nie uwzględniono w kopii zapasowej. Decyzja dotycząca tego, *co* należy archiwizować, jest istotna.

Przygotowanie się na najgorsze

Przed podjęciem decyzji, jakie pliki należy dodać do kopii zapasowych, powinno się zaprosić na obiad najbardziej pesymistycznie nastawionego pracownika technicznego firmy. W rzeczywistości warto spotkać się z kilkoma takimi osobami. Dodatkowo należy poprosić je o przedstawienie sytuacji, przed którymi chciałyby być chronione. Sytuacje te należy uwzględnić przy decydowaniu, co powinno zostać dodane do kopii zapasowej. Sytuacje te będą też pomocne podczas określania sposobu archiwizowania. Gościom zaproszonym na obiad należy zadać następujące pytanie: „Jakie są absolutnie najgorsze scenariusze, które mogą spowodować utratę danych?". Oto kilka możliwych odpowiedzi:

- Cały system pada ofiarą pożaru i doszczętnie spala się. Pozostaje sterta nierozpoznawalnych, stopionych metalowych elementów i sczerniałego, kopcącego się plastyku.
- Ponieważ komputer był tak bardzo ważny, trzeba go było replikować za pomocą następnego węzła, znajdującego się tuż obok. Oczywiście ten komputer zapali się razem z pierwszym.
- Jest centralny serwer, który kontroluje wszystkie kopie zapasowe, monitoruje lokalizacje woluminów z kopiami zapasowymi i typ przechowywanych na nich plików itp. Serwer, który został zniszczony, umieszczono obok tego „serwera archiwizującego". W efekcie intensywny żar spowodował również jego dewastację.
- Katastrofalna reakcja łańcuchowa uszkodziła serwery DHCP i Active Directory, główny serwer NIS, serwery NFS i CIFS z katalogami użytkowników, a także serwer bazodanowy inwentaryzujący wszystkie woluminy z kopiami zapasowymi i ich lokalizacjami. Ostatni z wymienionych komputerów przechowuje też telefoniczną bazę danych z wszystkimi umowami dotyczącymi usług, numerami dostawców i procedurami eskalacji.
- Ktoś nie mógł zapamiętać numeru do nowego dostawcy zewnętrznego magazynu, więc przykleił go taśmą do ściany tuż obok serwera archiwizującego. Oczywiście płomienie właśnie spaliły kartkę do tego stopnia, że nie można jej odczytać.
- Płomienie uaktywniły system gaszenia i woda leje się na woluminy z kopiami zapasowymi. Ale zły dzień…

Jak postąpić, gdy jeden z powyższych scenariuszy faktycznie się spełni? Czy w ogóle wiadomo, od czego zacząć? Czy wiadomo:

- Jaki wolumin zawiera kopię wykonaną ostatniej nocy?
- Gdzie kopię zapisano?
- Jak skontaktować się z dostawcą zewnętrznego magazynu, aby pobrać kopie woluminów z zarchiwizowanymi danymi? A czy po ich znalezieniu serwer i sprzęt sieciowy będą gotowe do przeprowadzenia procesu odtwarzania?
- Do kogo zadzwonić w celu wymiany sprzętu w niedzielę o drugiej w nocy?
- Jaka była struktura sieci, zanim spaliły się wszystkie kable?

Najpierw trzeba odtworzyć serwer archiwizujący, ponieważ zawiera wszystkie niezbędne informacje. Przyjmijmy, że udało się znaleźć w portfelu wizytówkę firmy archiwizującej i uzyskać każdy zarchiwizowany wolumin. Ponieważ przepadła baza danych nośników, skąd będzie wiadomo, który z nich przechowuje kopię wykonaną ostatniej nocy? A czas ucieka…

Załóżmy, że udało się przeszukać wszystkie woluminy i znaleźć ten, którego trzeba użyć do odtworzenia serwera archiwizującego (łatwiej powiedzieć, niż wykonać!). Dzięki własnym umiejętnościom i sprytowi, a także wydatnej pomocy obsługi technicznej, odtworzono dane. Wszystko jest dostępne i funkcjonuje. Ile dysków znajdowało się w komputerze, który został zniszczony? Jakie to były modele dysków? Jak zostały podzielone na partycje? Czy część dysków nie została za pomocą paskowania połączona w celu uzyskania większych woluminów? Czy niektóre dyski nie stanowiły kopii lustrzanej innych dysków? Gdzie te informacje są przechowywane? Czy w ogóle wiadomo, jaką pojemność miały napędy lub systemy plików? To naprawdę staje się skomplikowane…

Moje oko sprawdzono

Firma biotechnologiczna z kilkoma serwerami uważanymi za systemy sprawdzone pod kątem zastosowań FDA CFR21 straciła krytyczną bazę danych uruchomioną na jednym z serwerów. Gdy pracownicy firmy udali się do serwera archiwizującego, aby odtworzyć dane, ku swemu przerażeniu odkryli, że serwer z bazą danych nie był objęty archiwizacją od mniej więcej trzech miesięcy. W jakiś sposób serwer usunięto z harmonogramu archiwizacji. W związku z tym nie były generowane żadne „błędy". W efekcie pracownicy firmy pozostali bez czegokolwiek, co pozwoliłoby uzyskać aktualną kopię zapasową. Problem był na tyle poważny, że dotarł do samego prezesa zarządu.

— Jim Damoulakis

Czy nie przeprowadzono w zeszłym tygodniu sporej aktualizacji jądra na trzech komputerach (poprawkę eliminującą wszystkie ataki sieciowe polegające na zmasowanym wysyłaniu pakietów, które przeciążały sieć w środku dnia)? Czyż nie wykonano kopii zapasowej jądra po jego aktualizacji? Oczywiście poprawka uaktualniła pliki w obrębie całego napędu z systemem operacyjnym. Czyż nie sporządzono pełnej kopii zapasowej? Jak odtworzy się zawartość napędu z systemem operacyjnym? Czy naprawdę zamierza się przejść przez proces ponownej instalacji systemu operacyjnego tylko po to, aby można było uruchomić narzędzie *restore* i jeszcze raz nadpisać system?

Systemy plików nie narzekają na pojemność, dopóki mają wystarczający rozmiar, aby móc przechowywać odtwarzane dane. Dlatego niezbyt trudne jest skonfigurowanie i uaktywnienie takich systemów plików. A jak wygląda to w przypadku bazy danych używającej niesformatowanych partycji? Wiadomo, że będzie to znacznie bardziej złożone. Systemy plików będą wymagały umieszczenia urządzeń: */dev/rdsk/c7t3d0s7*, */dev/dsk/c8t3d0s7* i */dev/dsk/c8t4d0s7* dokładnie tam, gdzie znajdowały się wcześniej, i poddania ich partycjonowaniu tak samo jak przed awarią. Dodatkowo właścicielem urządzeń musi być użytkownik bazy danych. Czy wiadomo, których napędów właścicielem był ten użytkownik przed wystąpieniem awarii? A których po?

Może się to zdarzyć.

W części czwartej omówiono powyższe sytuacje.

Inwentaryzowanie

Trzeba się upewnić, że w razie nieszczęśliwego zdarzenia są dostępne następujące kluczowe informacje:

Kopie kopii zapasowych
: Wiele firm zaczęło centralizować zarządzanie swoimi kopiami zapasowymi, co według mnie jest dobrym posunięciem. Jednak gdy centralizuje się magazynowanie informacji dotyczących wszystkich kopii, z całym planem archiwizowania jest związany pojedynczy punkt awarii. Nie będzie można odtworzyć serwera archiwizującego, ponieważ nie jest dostępna baza danych kopii zapasowych. Nie ma takiej bazy, gdyż trzeba najpierw odtworzyć serwer archiwizujący. Operacja taka byłaby pierwszym krokiem w przypadku dowolnej awarii wielu systemów. W kwestii takiej jak inwentaryzacja nośników nie należy lekceważyć wartości wydrukowanego zestawienia przechowywanego poza obszarem firmy. Wydruk taki może po prostu uchronić przed wieloma problemami. Jeśli punkt awarii jest jeden, odtwarzanie serwera archiwizującego powinno być jak najprostszą i jak najlepiej udokumentowaną operacją. Można nawet rozważyć utworzenie narzędziem *tar*, *ntbackup* lub *rsync* specjalnej kopii zapasowej danych, która jeszcze bardziej uprości odtwarzanie po wystąpieniu awarii.

Jakie urządzenia peryferyjne były dostępne?
: Zakładając, że regularnie archiwizuje się konfigurację napędów dyskowych, można dysponować listą wszystkich dysków. Czy jednak wiadomo, jakie są modele używanych napędów? Jeśli wszystkie napędy są marki X i mają pojemność 500 GB, nie będzie problemu. Jednak wiele serwerów jest wyposażonych w kilka różnych napędów instalowanych w dłuższym okresie. W obrębie jednego komputera może się znajdować kombinacja napędów o pojemnościach 40, 100 i 500 GB. Trzeba zadbać o to, aby w jakiś sposób zapisać te informacje. Systemy Unix i Mac OS rejestrują je w pliku *messages*, a system Windows w rejestrze. Dlatego mam nadzieję, że Czytelnik archiwizuje plik *messages* i rejestr systemu Windows.

W jaki sposób dokonano podziału na partycje?
: To pytanie może *naprawdę* być istotne, gdy trzeba będzie odtworzyć zawartość napędu z systemem operacyjnym lub bazą danych. Oba typy napędów zwykle są partycjonowane przy użyciu niestandardowych partycji, które muszą być określone dokładnie tak samo jak wcześniej, aby można było poprawnie przeprowadzić proces odtwarzania. Zazwyczaj informacje na temat partycji nie są nigdzie przechowywane w systemie. W związku z tym trzeba zrobić coś dodatkowego, aby je zapisać. Przykładowo w systemie Solaris dla każdego napędu można uruchomić program *prtvtoc* i informacje zapisać w pliku. W internecie można poszukać skryptów, które gromadzą tego typu informacje. Dostępnych jest też kilka przeznaczonych do tego darmowych narzędzi.

Jak skonfigurowano menedżery woluminów?
: Dostępnych jest kilka menedżerów woluminów dołączonych do konkretnych systemów operacyjnych. Spośród nich wymieńmy: Veritas Volume Manager, Windows Dynamic Drives, Solstice (Online) Disk Suite i Logical Volume Manager firmy HP. Jak skonfiguro-

wano używany menedżer woluminów? Jakie urządzenia są objęte tworzeniem kopii lustrzanych? Jak skonfigurowano urządzenia wielodyskowe? Może trudno w to uwierzyć, ale takie informacje nie zawsze są rejestrowane przez zwykłe narzędzia archiwizujące. Przez wiele miesięcy korzystałem z menedżera Logical Volume Manager, zanim się dowiedziałem o istnieniu narzędzia *lvmcfgbackup* (archiwizuje dane konfiguracyjne menedżera LVM). Jeśli poprawnie udokumentowano konfigurację menedżera woluminów, czasami odtwarzanie w ogóle może nie być konieczne. Jeśli na przykład zepsuje się dysk systemu operacyjnego, wystarczy skonfigurować dyski tak jak wcześniej, a następnie w identyczny sposób ponownie określić paskowanie. W efekcie dane powinny być nienaruszone. Robiłem coś takiego kilkakrotnie.

Jak skonfigurowano bazy danych?
Miałem do czynienia z wieloma przestojami baz danych. Gdy pytam administratora baz danych, jak je skonfigurował, prawie zawsze odpowiedź brzmi: „Nie jestem pewien…". Należy odszukać informacje na temat konfiguracji i zapisać je w widocznym miejscu.

Czy udokumentowano konfigurację serwerów: DHCP, Active Directory, NFS i CIFS?
Dokumentować i jeszcze raz dokumentować! Istnieją setki powodów, aby poprawnie dokumentować takie rzeczy, jak konfiguracja tych serwerów. Jednym z nich jest odtwarzanie po wystąpieniu awarii. Dobra dokumentacja jest konkretnym elementem planu archiwizowania. Dokumentacja powinna być regularnie aktualizowana i dostępna. Nikt nie powinien mówić: „Od wielu lat nie konfigurowałem od podstaw serwerów NIS, Active Directory i NFS. Jak w takiej sytuacji ponownie wykonać tego typu operację? Czy ktoś ma egzemplarz mojej książki?". W rzeczywistości w tej sytuacji najlepszym rozwiązaniem jest zautomatyzowanie tworzenia nowych serwerów. Jeśli system operacyjny pozwala na to, należy poświęcić czas na napisanie skryptów automatyzujących instalowanie różnych usług i konfigurujących je pod kątem stosowanego środowiska. Skrypty należy zawrzeć w pakiecie uruchamianym każdorazowo podczas konfigurowania nowego serwera. Jeszcze lepsze będzie sprawdzenie, czy producent systemu operacyjnego oferuje produkty automatyzujące instalację nowych serwerów (na przykład Jumpstart firmy Sun, Ignite-UX firmy HP, Linux Kickstart i funkcje powielania systemu Mac OS).

Czy dysponuje się planem działania?
Mało przyjemne scenariusze zostały przedstawione na początku, aby zmotywować Czytelnika, by zaczął przygotowywać plan. Nie należy czekać z zakupem szufli, aż przed domem nagromadzi się 6-metrowa warstwa śniegu! Śnieg spadnie. Pytanie tylko kiedy. Warto zaprosić pesymistów z firmy na obiad i pozwolić im, aby wyobrazili sobie najgorsze scenariusze, które mogą mieć miejsce, a następnie przygotować plan działania. Trzeba dysponować w pełni udokumentowanym krok po kroku planem przewidującym koniec świata komputerowego. Jeśli nawet plan będzie wymagał niewielkiej modyfikacji, gdy faktycznie trzeba będzie z niego skorzystać, dobrze będzie móc od czegoś zacząć. Jest to o wiele lepsze niż stwierdzenie: „Co teraz wiemy? Czy ktoś widział moje zestawienie?" (sporządzono jego papierowy wydruk, zgadza się?).

Trzeba wiedzieć, co tkwi w komputerach!
Dla osoby zajmującej się archiwizowaniem i odtwarzaniem danych najlepszym zabezpieczeniem przed niemal każdego rodzaju stratą jest znajomość chronionych komputerów. Jeśli określony serwer przestanie działać, powinno się od razu wiedzieć, że znajduje się na nim baza danych Oracle lub SQL Server, która obsługuje konkretne woluminy. Dzięki takiej wiedzy można natychmiast rozpocząć odtwarzanie serwera. Należy mocno zaangażować się w instalację każdego nowego systemu lub bazy danych. Powinno się wiedzieć, jakie są

używane platformy bazodanowe i jak je skonfigurowano. Powinno się dysponować informacjami na temat wszystkich nowych napędów, systemów plików, baz danych lub systemów. Trzeba bardzo dobrze zaznajomić się z każdym komputerem, z jego przeznaczeniem i zawartością. Wiedza taka jest o tyle istotna, że należy przewidzieć wykonywanie specjalnych kopii zapasowych.

Warto przeglądać dzienniki

Było to moje pierwsze zlecenie po ukończeniu uczelni. Miałem przede wszystkim obsługiwać zwykłe komputery, ucząc się u boku dobrze opłacanego konsultanta od Uniksa, któremu nadam imię *Fred*.

Obsługiwaliśmy aplikację walutową ForEx o nazwie Opus, którą uruchomiono pod systemem SunOS. Gdy program zapisywał transakcje walutowe, połowę informacji umieszczał w ścieżce. Jeśli na przykład ktoś zrealizował 15 czerwca walutową transakcję dolarowo-funtową z kimś z banku Bank of New York, ścieżka i nazwa pliku wyglądały tak:

```
/opt/app/opus/transactions/portfolio/third-
party/...itd...itd.../USD/CAI/GBP/BONY/ask/19970615120453.2372149821335
```

Choć taki mało sympatyczny zapis nie był z pewnością winą Freda, postanowił uwzględnić go w procesie archiwizowania. Zauważyłem, że zdefiniował zadanie *tar* z opcją `-v`, które generowało dzienniki na tyle duże, że nie chciało mu się ich przeglądać. Gdy usunąłem opcję `-v` i zacząłem sprawdzać dzienniki, stwierdziłem, że kopie zapasowe były niepoprawnie wykonywane. Wersja narzędzia *tar* dodawana w tamtym czasie do systemu SunOS obcinała ścieżki plików o długości przekraczającej 100 znaków. Ścieżki transakcji walutowych były za długie o mniej więcej 9 znaków. Fred po prostu archiwizował narzędziem *tar* ogromne drzewo katalogów pozbawione plików. Jeśli kiedykolwiek przejrzałby dzienniki, wiedziałby o tym.

Następnego dnia zostałem głównym administratorem systemu Unix. Firma nie przedłużyła kontraktu z Fredem.

— Jim „Sparky" Donnellan

Czy archiwizuje się to, co faktycznie zaplanowano?

Pamiętam administratora pracującego u jednego z moich poprzednich pracodawców, który mówił: „Czy znalazło się to na taśmie?". Zawsze zadawał to pytanie z charakterystycznym uśmiechem. Był to taki jego sposób witania się z osobą zajmującą się archiwizowaniem. Jego pytanie ma sens. Istnieje kilka ogólnych metod tworzenia kopii zapasowych, dzięki którym można w znacznym stopniu zwiększyć ich efektywność. Zanim zastanowimy się, czy archiwizować część systemu czy cały system, przyjrzyjmy się powszechnej praktyce używania list dołączeń i stwarzanym przez nie zagrożeniom. Ponadto przeanalizujmy niektóre metody, których można uniknąć w przypadku stosowania list dołączeń.

Czym są listy dołączeń i wykluczeń? Ogólnie mówiąc, można wyróżnić dwie metody archiwizowania systemu:

- Można poinstruować system archiwizujący, aby zarchiwizował wszystko z wyjątkiem tego, co znajduje się na *liście wykluczeń*. Oto przykład:
  - Serwery z systemami Unix, Linux i Mac OS:

```
Dołącz: *
Wyklucz: /tmp, /junk1, /junk2
```

- Serwery z systemem Windows:

```
Dołącz: *
Wyklucz: *.tmp, *Temporary Internet Files*, ~*.*, *.mp3
```

- Można poinstruować system archiwizujący, aby zarchiwizował to, co znajduje się na *liście dołączeń*. Oto przykład:
  - Serwery z systemami Unix, Linux i Mac OS:

```
Dołącz: /data1, /data2, /data3
```

  - Serwery z systemem Windows:

```
Dołącz: D:\, E:\
```

Po przyjrzeniu się powyższym przykładom można zadać sobie pytanie: „Co się stanie, gdy utworzy się katalog */data4* lub doda napęd *F:*\?". Ktoś musi pamiętać o dodaniu katalogu lub napędu do listy dołączeń, w przeciwnym razie nie zostaną uwzględnione w kopii zapasowej. Jest to przepis na spore problemy. Jeśli nie jest się jedyną osobą, która dodaje napędy lub systemy plików, i nie ma się doskonałej pamięci, zawsze przeoczy się jakiś napęd lub system plików. Dopóki są inni administratorzy i w głowie ma się substancję szarą, coś zostanie pominięte.

Nie cierpię, gdy to się dzieje

Pracowałem w większej firmie wydawniczej, gdy przestał działać serwer obrazów. Gdy zainteresowane osoby udały się do administratora kopii zapasowych i poprosiły o odtworzenie wszystkich obrazów znajdujących się na serwerze, okazało się, że nie dysponował żadnymi danymi z tego komputera. Powodem było to, że po dodaniu rok wcześniej serwera do środowiska produkcyjnego nikt oficjalnie nie zażądał, aby uwzględnić go w systemie archiwizacji. W efekcie firma straciła tysiące obrazów.

— *Chris Pritchard*

Jeśli jednak narzędzie archiwizujące nie obsługuje automatycznego wykrywania napędu lub systemu plików, niewiele wysiłku trzeba, aby stwierdzić: „Trzeba wszystko zarchiwizować". W jaki sposób sporządza się listę systemów, napędów, systemów plików i baz danych, które mają zostać uwzględnione przez proces archiwizacji? W tym celu trzeba poszukać plików, takich jak */etc/vfstab* lub rejestry systemu Windows, a następnie wyodrębnić listę napędów lub systemów plików, dla których wykona się kopię zapasową. Można następnie za pomocą list wykluczeń wyłączyć wszelkie napędy lub systemy plików, które nie mają być archiwizowane.

Serwer bazodanowy Oracle dla systemu Unix ma plik o nazwie *oratab*, który może posłużyć do przechowywania listy wszystkich instancji bazodanowych serwera[1]. Oczywiście system Windows przechowuje takie informacje w rejestrze. Za pomocą pliku *oratab* można wyszczególniać wszystkie instancje wymagające archiwizacji. Niestety, bazy danych Informix i Sybase nie oferują takiego pliku. Trzeba go utworzyć ręcznie. Zachęcam do zrobienia tego z wielu powodów. Dysponując takim plikiem, znacznie łatwiej normalizować proces uruchamiania systemu i wykonywania kopii zapasowych. Jeśli tak skonfiguruje się skrypty startowe, że baza danych nie będzie uaktywniana, gdy nie będzie jej w tym pliku, można być niemal pewnym,

[1] Instancję serwera Oracle można zainstalować bez umieszczania jej w tym pliku. Jednakże instancja nie zostanie uruchomiona po ponownym załadowaniu systemu. Zwykle oznacza to, że administrator baz danych woli sam uwzględnić instancję w pliku. Więcej na ten temat można znaleźć w rozdziale 15.

że w pliku znajdą się wszystkie istotne bazy danych. Oznacza to oczywiście, że wszelkie ważne bazy danych są archiwizowane bez konieczności ręcznej interwencji administratora. Oznacza to również, że w przypadku każdego komputera można użyć tych samych skryptów uruchamiających bazy danych Informix i Sybase. Nie trzeba na stałe umieszczać w skryptach nazwy każdej bazy danych.

Jak stwierdzić, jakie systemy zarchiwizować? Choć nigdy nie uzyskałem pełnej odpowiedzi na to pytanie, zawsze chciałem napisać skrypty, z których jeden monitorowałby różne bazy danych, szukając nowych systemów. Zależało mi na uzyskaniu z serwera DNS (*Domain Name System*) kompletnej listy wszystkich hostów i porównaniu jej z główną listą. Po znalezieniu nowego adresu IP chciałem spróbować stwierdzić, czy jest on aktualny. Jeżeli tak by było, oznaczałoby to, że pojawił się nowy komputer, który prawdopodobnie wymaga archiwizacji. Taki skrypt byłby bezcenny. Zapewniłby, że w sieci nie byłoby żadnych nowych systemów, w przypadku których nie byłoby wiadomo, czy wykonuje się kopie zapasowe, czy nie. Po zlokalizowaniu nowego adresu IP za pomocą narzędzia *nmap* można zidentyfikować typ systemu, któremu adres przypisano. Narzędzie to wysyła niepoprawny pakiet TCP pod adres IP. Odpowiedź otrzymana spod tego adresu pozwala określić, pod kontrolą jakiego systemu operacyjnego komputer działa.

Obecnie kilka komercyjnych pakietów oprogramowania zarządzającego ochroną danych oferuje taką możliwość.

Czy archiwizować cały system czy jego część?

Zakładając, że wzięto pod uwagę rzeczy, które nie są uwzględniane przez standardowe systemowe kopie zapasowe, trzeba zdecydować, czy zarchiwizuje się cały system czy tylko jego wybrane napędy lub systemy plików. W tym przypadku należy wspomnieć o dwóch różnych szkołach. Z tego, co się orientuję, z wariantem archiwizowania wybranych systemów plików związanych jest zbyt wiele pułapek. Archiwizowanie wszystkiego jest prostsze i bezpieczniejsze od sporządzania kopii zapasowej za pomocą listy. Można zauważyć, że w większości książek pada stwierdzenie: „Choć najlepiej archiwizować wszystko, większość osób postępuje inaczej". Takiego zdania Czytelnik nie znajdzie w tej książce. Uważam, że nieuwzględnienie wszystkiego w kopii zapasowej jest bardzo niebezpieczne. Pod uwagę warto wziąć poniższe porównanie dwóch metod.

Archiwizowanie tylko wybranych napędów lub systemów plików

Oto argumenty za i przeciw sporządzaniu kopii zapasowych dla niektórych danych:

Oszczędność przestrzeni nośnika i mniejsze obciążenie sieci. To pierwszy argument, zwykle uznawany za zaletę metody archiwizowania wybranych systemów plików. Polega to na tym, że w kopii zapasowej umieszcza się mniejszą ilość danych. Zwolennicy tej szkoły zalecają tworzenie dwóch grup kopii zapasowych: z danymi systemu operacyjnego i zwykłymi danymi. Chodzi o to, że kopie zapasowe systemu operacyjnego byłyby wykonywane rzadziej. Niektórzy nawet sugerują, aby takie kopie sporządzać tylko po dokonaniu znaczącej modyfikacji, takiej jak zastosowanie poprawki zabezpieczeń systemu Windows,

zaktualizowanie systemu, zainstalowanie poprawki lub przebudowanie jądra. Z kolei zwykłe dane byłyby archiwizowane codziennie.

Pierwszy problem jest taki, że ten argument stracił na aktualności. Wystarczy przyjrzeć się rozmiarowi danych typowego nowoczesnego komputera. Obecnie ilość zwykłych danych znacznie przekracza ilość danych systemu operacyjnego. Nie zaoszczędzi się zbyt wiele przestrzeni lub przepustowości sieci przez zrezygnowanie z archiwizowania systemu operacyjnego, nawet podczas wykonywania pełnej kopii zapasowej. Gdy pod uwagę weźmie się przyrostowe kopie zapasowe, proporcja ta będzie jeszcze większa. Jeżeli nie ma czegoś ważnego, co powinno zostać zarchiwizowane, partycje systemu operacyjnego nie spowodują dużego wzrostu rozmiaru przyrostowej kopii zapasowej! Do ważnych danych można zaliczyć pliki systemów: Unix, Linux i Mac OS, takie jak */etc/passwd*, */etc/hosts*, *syslog*, */var/adm/messages*, i wszelkie inne pliki, które okażą się przydatne w przypadku awarii systemu operacyjnego. Do istotnych danych należy również dołączyć rejestr systemu Windows. Plik wymiany jest raczej jedyną zupełnie bezwartościową porcją informacji, która może zostać umieszczona na dysku systemu operacyjnego. Plik wymiany można wyłączyć za pomocą listy wykluczeń.

Trudniejsze administrowanie. Zwolennicy okazjonalnego wykonywania kopii zapasowych powiedzieliby, że ważne pliki, takie jak wyżej wymienione, można dodać do specjalnej kopii zapasowej. W tym przypadku problem polega na tym, że jest to znacznie trudniejsze od archiwizacji wszystkiego. Zakładając, że z większości kopii zapasowych wyłączy się pliki konfiguracyjne, trzeba będzie pamiętać o przeprowadzeniu ręcznej archiwizacji każdorazowo po zmodyfikowaniu pliku konfiguracyjnego lub bazy danych. Oznacza to, że po wprowadzeniu zmiany konieczne będzie wykonanie czegoś *specjalnego*, *co jest czymś złym*. Jeśli po prostu wszystko się zarchiwizuje, można administrować systemami stosownie do wymagań bez konieczności pamiętania o wykonaniu kopii zapasowej przed wprowadzeniem jakiejś zmiany.

Łatwiej jest dzielić kopię między woluminami. Jedna z bardzo niewielu rzeczy, które mogą być uznane za plus, ma miejsce wtedy, gdy napędy lub systemy plików rozdzieli się między wieloma kopiami zapasowymi. Łatwiejsze jest rozmieszczenie kopii zapasowych na wielu woluminach. Jeśli kopia zapasowa systemu nie mieści się na jednym woluminie, prościej zautomatyzować archiwizację przez podzielenie kopii na dwie obsługiwane przez różne listy dołączeń. Jednak aby z tego skorzystać, zamiast list wykluczeń trzeba zastosować listy dołączeń, z którymi są związane wcześniej zaprezentowane ograniczenia. Powinno się sprawdzić, czy narzędzie archiwizujące oferuje lepszą metodę rozwiązania tego problemu.

Łatwiej napisać odpowiedni skrypt, niż samemu przetwarzać pliki fstab i oratab lub rejestr systemu Windows. Trudno się spierać z takim stwierdzeniem. Jeśli poświęci się czas na to, aby zadanie wykonać dobrze za pierwszym razem, nigdy nie trzeba będzie znów używać list dołączeń. Przychodzi mi w tym miejscu na myśl moje inne ulubione stwierdzenie: „Nigdy nie ma czasu na zrobienie czegoś poprawnie, a zawsze jest czas na wykonanie czegoś jeszcze raz". Warto znaleźć czas na wykonanie tego zadania dobrze za pierwszym razem.

Najgorsze, co może się wydarzyć? Coś przeoczono! W tym przypadku początkowo największe korzyści są takie, że zaoszczędzi się trochę czasu, nie pisząc skryptów, a także zyska kilka bajtów przepustowości sieci. Najgorszy możliwy efekt uboczny to przeoczenie napędu lub systemu plików z budżetem szefa, który właśnie został usunięty.

Archiwizowanie całego systemu

Lista zalet archiwizowania całego systemu jest krótsza i znacznie bardziej przekonująca. Oto ona:

Pełna automatyzacja. Gdy uda się utworzyć działający skrypt lub program, wystarczy monitorować generowane przez niego dzienniki. Można spokojnie spać, wiedząc, że wszystkie dane znajdą się w kopii zapasowej.

Najgorsze, co może się wydarzyć? Straci się przyjaciela pracującego w dziale sieciowym. Można zwiększyć obciążenie sieci o kilka punktów procentowych, co może się nie spodobać osobom zajmującym się okablowaniem sieciowym (oczywiście taki stan rzeczy będzie trwał do momentu, gdy będzie trzeba odtworzyć serwer, na którym osoby te przechowują źródłową bazę danych DNS).

Archiwizowanie wybranych napędów lub systemów plików jest jednym z najczęściej popełnianych błędów, na które napotykam podczas analizowania konfiguracji archiwizacji. Bardzo łatwo wpaść w taką pułapkę przez wybiórcze archiwizowanie dla zaoszczędzenia czasu. Dopóki jednak nie przekona się o tym samemu, można nie wiedzieć, w jak dużym niebezpieczeństwo się znalazło. Jeśli konfiguracja archiwizacji bazuje na listach dołączeń, mam nadzieję, że to omówienie przekona Czytelnika do ponownego przemyślenia podjętej decyzji.

Decydowanie o momencie przeprowadzania archiwizacji

Mogłoby się wydawać, że jest to najbardziej oczywiste zagadnienie. Czyż nie każdy archiwizuje swój system każdej nocy? W takim razie w czym tkwi problem? Właściwsze pytanie brzmi: „Jakie poziomy są uaktywniane i kiedy?". Zawsze jest to poważne pytanie. Jak często wykonuje się pełną kopię zapasową? Jak często sporządza się przyrostowe kopie zapasowe? Czy stosuje się różne poziomy archiwizacji przyrostowej uwzględniającej wyłącznie dzisiejsze zmiany, czy ciągłą archiwizację przyrostową obejmującą wszystko, co zmieniło się od czasu wykonania ostatniej pełnej kopii zapasowej? Każdy ma własne odpowiedzi na te pytania. Pewne jest tylko to, że każdej nocy powinien być stosowany *co najmniej* jeden z poziomów archiwizacji. Zanim przejdziemy dalej, warto zdefiniować kilka pojęć.

Poziomy archiwizowania

Poniżej zaprezentowano różne poziomy archiwizowania. Każdy używa tych terminów w inny sposób.

Pełna kopia/Poziom 0
: Pełna kopia zapasowa.

Poziom 1
: Przyrostowa kopia zapasowa archiwizująca wszystko, co zmieniło się od czasu wykonania ostatniej kopii zapasowej poziomu 0. Kolejne kopie poziomu 1 nadal uwzględniają wszystko, co zostało zmodyfikowane od chwili sporządzenia ostatniej pełnej kopii zapasowej poziomu 0.

Poziomy 2 – 9

Każdy poziom archiwizuje wszystko, co zmieniło się od momentu wykonania ostatniej kopii następnego najniższego poziomu. Oznacza to, że poziom 2 archiwizuje wszystko, co zmieniło się od chwili utworzenia kopii zapasowej poziomu 1 lub poziomu 0 (jeśli brak poziomu 1). W przypadku niektórych produktów kolejne kopie zapasowe poziomu 9 uwzględniają to, co się zmieniło od czasu sporządzenia ostatniej kopii tego poziomu. Jednak jest to dalekie od powszechności.

Kopia przyrostowa

Zwykle tego typu kopia zawiera wszystko, co zmieniło się od momentu wykonania ostatniej kopii zapasowej dowolnego rodzaju.

Kopia różnicowa

Większość osób traktuje kopię różnicową jako taką, która archiwizuje wszystko, co zostało zmodyfikowane od czasu sporządzenia ostatniej pełnej kopii zapasowej. Jednak nie jest to powszechne rozumowanie. W systemie Windows różnicową jest kopia, która nie usuwa bitu archiwizacji. A zatem, gdy wykona się pełną kopię zapasową, a następnie kilka kopii różnicowych, w tradycyjnym znaczeniu będą one funkcjonowały jak kopie różnicowe. Jeśli jednak w systemie utworzy się choćby jedną kopię przyrostową, usunie ona bit archiwizacji i w efekcie następna kopia różnicowa uwzględni tylko te pliki, które zmodyfikowano od czasu sporządzenia ostatniej kopii przyrostowej. Właśnie z tego powodu kopia różnicowa nie jest tożsama z kopią zapasową poziomu 1.

Skumulowana kopia przyrostowa

Preferuję używanie tego określenia zamiast pojęcia kopii różnicowej. Termin identyfikuje kopię zapasową uwzględniającą wszystkie pliki, które zostały zmodyfikowane od momentu wykonania ostatniej pełnej kopii zapasowej.

Narzędzia archiwizujące i administratorzy korzystający z nich nie stosują tych terminów w jednoznaczny sposób. Trzeba się upewnić, jak używany produkt interpretuje określony termin!

Często zadawane jest mi następujące pytanie: „Czy mam archiwizować dane każdej nocy?". W rzeczywistości pytanie to powinno brzmieć tak: „Czy nawet w weekend?". Nikt nie pracuje w weekend, zgadza się? *Tak*, z wyjątkiem klienta, który zrobił wielką awanturę w zeszły weekend. Czytelnik może się domyślić, o jakim kliencie mowa. Taki klient zamiast do działu pomocy technicznej dzwoni do szefa, gdy pojawi się problem. I jeśli nie zastanie szefa lub ten nie poradzi sobie z problemem wystarczająco szybko, klient skontaktuje się z jego przełożonym. W zeszły weekend taki klient był naprawdę zapracowany i spędził go w całości nad przyszłorocznym budżetem. Około pierwszej nad ranem w poniedziałek klient wreszcie ukończył przygotowywanie budżetu. Około czwartej rano przestał działać dysk z domowym katalogiem klienta (czyż wszystko nie odmawia posłuszeństwa w poniedziałkowy poranek?). Od piątkowego wieczoru nie sporządzono kopii zapasowej danych. W efekcie dzwoni szef. Do głowy przychodzą różne powody, dla których szef postanowił zadzwonić. Czy *Czytelnik* chciałby być tą osobą, która powie temu klientowi, że można było zapisać jego plik, lecz w weekend nie została przeprowadzona archiwizacja danych?

Bit archiwizacji systemu Windows jest złem!

Bit archiwizacji systemu Windows jest złem, które musi zostać powstrzymane. W najgorszym razie producenci sprzętu archiwizującego powinni umożliwić niekorzystanie z tego bitu, bez żadnych konsekwencji. Jeśli dla pliku w systemie Windows ustawi się bit gotowości do archiwizacji, będzie to oznaczać, że plik jest nowy lub zmodyfikowany i powinien zostać dołączony do przyrostowej kopii zapasowej. Po zarchiwizowaniu pliku bit archiwizacji jest usuwany. A zatem pierwszy problem związany z tym bitem polega na tym, że powinien być nazywany *bitem kopii zapasowej*. Kopie zapasowe nie są archiwami.

Jednak największym problemem z bitem archiwizacji jest to, że związany z nim proces przyjmuje, że bit ten usunie tylko jedna aplikacja, podczas gdy w rzeczywistości może być ich kilka. Pierwszy program archiwizujący użyty do utworzenia kopii zapasowej katalogu usuwa bit archiwizacji, więc następny program nie dokona archiwizacji tych samych plików katalogu. Załóżmy, że użytkownik postanowi zastosować narzędzie *ntbackup* w celu zarchiwizowania na dysku CD swoich plików znajdujących się na firmowym serwerze. Jeżeli tak postąpi, program *ntbackup* usunie bit archiwizacji i firmowy system archiwizacji odpowiedzialny za wykonywanie kopii zapasowej plików katalogu nie doda ich do tworzonej przyrostowej kopii zapasowej. Ponieważ dla plików nie jest ustawiony bit archiwizacji, wynika z tego, że nie ma potrzeby uwzględniania ich w kopii zapasowej. Oznacza to, że każdy użytkownik może spowodować niezgodne z zamierzeniami działanie całego systemu archiwizacji.

Zwolennicy bitu archiwizacji podkreślają, że jest on ustawiany dla nowo zainstalowanego oprogramowania, nawet wtedy, gdy pliki są stare. Program archiwizujący, który używa jedynie czasu modyfikacji, nie zauważa takich plików, jeśli są starsze niż najnowsza przyrostowa kopia zapasowa. Dlatego taki program powinien być może stosować kombinację bitu archiwizacji i czasu modyfikacji. Jeżeli to lub to zmieni się, plik powinien zostać uwzględniony w przyrostowej kopii zapasowej.

W przypadku archiwizowania danych systemów uniksowych nie występuje bit archiwizacji. W związku z tym aplikacje archiwizujące używają wartości *mtime* (gdy zawartość pliku w ostatnim czasie zmieniła się) lub *ctime* (gdy atrybuty pliku zostały w ostatnim czasie zmodyfikowane). Podczas tworzenia kopii zapasowej w systemie Windows różne aplikacje archiwizujące w odmienny sposób stosują bit archiwizacji. Część aplikacji używa go w połączeniu z wartością *mtime* lub *ctime*. Część narzędzi stosuje wyłącznie bit archiwizacji, natomiast część w ogóle z niego nie korzysta (biorąc pod uwagę to, co napisałem na temat bitu archiwizacji, może nie być to takie złe rozwiązanie).

Microsoft zaoferował w systemie Windows 2000 i jego następcach alternatywę dla bitu archiwizacji w postaci *dziennika zmian* (ang. *change journal*). Produkty archiwizujące obsługujące dziennik zmian mogą z niego korzystać przy określaniu, które pliki zostały zmodyfikowane. Tym samym nie muszą w tym celu szukać bitu archiwizacji. Choć domyślnie dziennik zmian nie jest aktywny, można go włączyć poleceniem `fsutil usn createjournal`. Trzeba określić wartość parametru `MaximumSize`, która będzie na tyle duża, aby przechować wszystkie zmiany dokonane między sporządzeniem kolejnych kopii zapasowych. Ponieważ w jednym rekordzie o rozmiarze 4 kB mieści się 30 lub 40 zmian, w dzienniku o pojemności 75 MB można pomieścić 500 000 zmian (jeśli dziennik zmian nie jest wystarczająco duży, w celu zrobienia miejsca najstarsze modyfikacje są usuwane z początku dziennika, dlatego ważne jest, żeby dziennik miał odpowiednią pojemność). Sugeruję określić największą liczbę plików, jaką do tej pory umieszczono w przyrostowej kopii zapasowej, a następnie dla dziennika ustawić dwukrotnie większą pojemność. Dzięki temu zwiększy się integralność systemu archiwizacji kosztem niewielkiego wzrostu przestrzeni zajmowanej przez dziennik zmian.

Jakie poziomy są stosowane i kiedy?

Jeśli chodzi o odpowiedź na to pytanie, szkół jest kilka. Poniżej przedstawiono kilka propozycji harmonogramów archiwizacji.

Harmonogram tygodniowy — same pełne kopie zapasowe poziomu 0

W tabeli 2.1 zaprezentowano harmonogram archiwizacji dla paranoika (nie oznacza to wcale nic złego). Kopia zapasowa poziomu 0 jest sporządzana codziennie na oddzielnym woluminie (proszę nie nadpisywać wczorajszej poprawnej kopii zapasowej poziomu 0 dzisiejszą kopią, która może być uszkodzona!). Jeżeli system jest naprawdę niewielki, taki harmonogram może być w sam raz. Jeśli jednak systemy mają duży rozmiar, ten harmonogram archiwizacji nie będzie zbyt dobrze skalowalny. Poza tym w przypadku obecnie dostępnego komercyjnego oprogramowania archiwizującego stosowanie takiego harmonogramu naprawdę nie jest konieczne.

Tabela 2.1. Same pełne kopie zapasowe

Niedziela	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota
Pełna kopia/ poziom 0	Pełna kopia/ poziom 0	Pełna kopia/ poziom 0	Pełna kopia/ poziom 0	Pełna kopia/ poziom 0	Pełna kopia/ poziom 0	Pełna kopia/ poziom 0

Harmonogram tygodniowy — tygodniowa pełna kopia zapasowa i codzienne różnicowe kopie poziomu 1

Zaletą harmonogramu przedstawionego w tabeli 2.2 jest to, że dane z okresu tygodnia zwykle trzeba będzie odtwarzać tylko z dwóch woluminów: z pełną kopią poziomu 0 i najnowszą różnicową kopią poziomu 1. Wynika to stąd, że każda różnicowa kopia zapasowa poziomu 1 uwzględnia wszystkie zmiany dokonane od momentu utworzenia w niedzielę pełnej kopii. Następną zaletą tego typu harmonogramu jest to, że uzyskuje się wiele kopii plików modyfikowanych na początku tygodnia. Prawdopodobnie jest to najlepszy harmonogram do zastosowania, gdy korzysta się z prostych narzędzi, takich jak *dump*, *tar* lub *cpio*, ponieważ wymagają one od administratora pełnego zarządzania woluminami. Odtwarzanie z dwóch woluminów jest *znacznie łatwiejsze* niż z sześciu (można mi zaufać)!

Tabela 2.2. Tygodniowa pełna kopia zapasowa i codzienne różnicowe kopie poziomu 1

Niedziela	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota
Pełna kopia/ poziom 0	Różnicowa kopia/poziom 1	Różnicowa kopia/poziom 1	Różnicowa kopia/poziom 1	Różnicowa kopia/poziom 1	Różnicowa kopia/poziom 1	Różnicowa kopia/poziom 1

Harmonogram tygodniowy — tygodniowa pełna kopia zapasowa i codzienne kopie poziomów

Jeśli produkt archiwizujący obsługuje wiele poziomów, można zastosować harmonogram pokazany w tabeli 2.3. Zaletą tego harmonogramu jest to, że wymaga mniej czasu i nośników niż wcześniej omówiony. Ma on dwie wady. Po pierwsze, każdy zmodyfikowany plik jest archiwizowany tylko raz, co powoduje bardzo dużą obawę o utratę danych, gdy dojdzie do

Tabela 2.3. Tygodniowa pełna kopia zapasowa i codzienne kopie poziomów

Niedziela	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota
Pełna kopia/ poziom 0	1	2	3	4	5	6

uszkodzenia dowolnego z nośników. Po drugie, w celu pełnego odtworzenia danych w piątek potrzebnych będzie 6 woluminów. Jeżeli używa się dobrego narzędzia archiwizującego open source lub komercyjnego, druga z wymienionych wad tak naprawdę nie stanowi problemu, ponieważ aplikacje te całkowicie wyręczają użytkownika w zarządzaniu woluminami (włącznie z wymienianiem taśm — za pomocą zautomatyzowanego zmieniacza).

Harmonogram tygodniowy — miesięczna pełna kopia zapasowa i codzienne przyrostowe kopie (wieża Hanoi)

Jeden z najbardziej interesujących pomysłów, z jakimi się spotkałem, nosi nazwę planu archiwizacyjnego wieży Hanoi. Bazuje on na starożytnej łamigłówce o tej samej nazwie, wykorzystującej ciąg matematyczny. Łamigłówka składa się z trzech kołków i kilku zakładanych na nich pierścieni o różnych średnicach. Pierścień nie może być umieszczony na pierścieniu o mniejszej średnicy. Celem gry jest przemieszczenie wszystkich pierścieni z pierwszego kołka na trzeci przy wykorzystaniu drugiego kołka, służącego w razie potrzeby do tymczasowego magazynowania[2].

Zadaniem większości harmonogramów archiwizacji jest umieszczenie zmodyfikowanych plików na więcej niż jednym woluminie przy jednoczesnym zredukowaniu całkowitej wykorzystanej pojemności woluminów. Plan archiwizacyjny wieży Hanoi radzi sobie z tym lepiej od każdego innego harmonogramu. Jeśli na potrzeby poziomów archiwizacji zastosuje się tego typu plan, większość zmodyfikowanych plików zostanie zarchiwizowana najwyżej dwa razy. Poniżej przedstawiono dwie różne wersje ciągu (nawiasem mówiąc, są one powiązane z liczbą pierścieni znajdujących się na trzech kołkach).

0 3 2 5 4 7 6 9 8 9
0 3 2 4 3 5 4 6 5 7 6 8 7 9 8

W rzeczywistości te ciągi są naprawdę proste. Każdy składa się z dwóch przeplatających się zestawów liczb (zestaw 2 3 4 5 6 7 8 9 przeplata się z zestawem 3 4 5 6 7 8 9). W tabeli 2.4 pokazano harmonogram ilustrujący działanie planu archiwizacji wieży Hanoi.

Tabela 2.4. Podstawowy harmonogram planu archiwizacji wieży Hanoi

Niedziela	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota
0	3	2	5	4	7	6

Harmonogram rozpoczyna się od niedzielnego poziomu (pełna kopia zapasowa). Załóżmy, że plik zmodyfikowano w poniedziałek. Poniedziałkowa kopia zapasowa poziomu 3 uwzględni wszystko, co się zmieniło od chwili wykonania kopii zapasowej poziomu 0. Oznacza to, że zmodyfikowany plik znajdzie się w poniedziałkowej kopii. Przyjmijmy, że we wtorek zmodyfikowano następny plik. A zatem we wtorkową noc kopia zapasowa poziomu 2 musi poszukać niższego poziomu, prawda? Ponieważ poniedziałkowy poziom 3 nie jest niższy, kopia

[2] Aby zapoznać się z całą historią gry i uzyskać adres URL, pod którym można w nią zagrać, należy zajrzeć na stronę znajdującą się pod adresem *http://www.math.toronto.edu/mathnet/games/towers.html*.

odwołuje się również do poziomu 0. W efekcie zarchiwizowany zostanie plik zmodyfikowany w poniedziałek, a także plik zmieniony we wtorek. W środę kopia zapasowa poziomu 5 uwzględnia tylko to, co zostało zmodyfikowane tego dnia, ponieważ odwołuje się ona do wtorkowej kopii poziomu 2. Jednak w czwartek kopia poziomu 4 nie odwołuje się do środowego poziomu 5, lecz do wtorkowego poziomu 2.

Warto zauważyć, że plik zmodyfikowany we wtorek został zarchiwizowany tylko raz. Aby poradzić sobie z tym problemem, stosuje się zmodyfikowaną wersję ciągu matematycznego planu archiwizacji wieży Hanoi, w przypadku którego każdego tygodnia poziom archiwizacji obniża się do poziomu 1 (tabela 2.5).

Tabela 2.5. Miesięczny harmonogram planu archiwizacji wieży Hanoi

Dzień tygodnia	Pierwszy tydzień	Drugi tydzień	Trzeci tydzień	Czwarty tydzień
Niedziela	0	1	1	1
Poniedziałek	3	3	3	3
Wtorek	2	2	2	2
Środa	5	5	5	5
Czwartek	4	4	4	4
Piątek	7	7	7	7
Sobota	6	6	6	6

Jeśli nie będzie to stanowić problemu dla Czytelnika i stosowanej metody archiwizacji[3], a także dla używanego systemu archiwizacji, namawiam do wybrania harmonogramu z tabeli 2.5. Każdej niedzieli uzyska się kompletną przyrostową kopię zapasową wszystkiego, co zmieniło się od czasu wykonania pełnej miesięcznej kopii. W pozostałe dni tygodnia każdy zmodyfikowany plik jest archiwizowany dwukrotnie, z wyjątkiem plików zmienionych w środę. W ten sposób można się uchronić przed awarią nośnika lepiej niż w przypadku każdego z wcześniej opisanych harmonogramów. Choć oczywiście do przeprowadzenia pełnego procesu odtwarzania będzie wymagany więcej niż jeden wolumin, nie stanowi to problemu, gdy dysponuje się zaawansowanym narzędziem archiwizującym z funkcją zarządzania woluminami.

„W środku nocy..."

Ten cytat z piosenki Billy'ego Joela identyfikuje porę, która zwykle najlepiej nadaje się do archiwizowania danych. Wykonywanie kopii zapasowych powinno być tak zaplanowane, aby nie miało miejsca w normalnych godzinach pracy firmy. Choć czasami nie można tego uniknąć, nie powinno się to zdarzać zbyt często. Oto dwa podstawowe powody:

Spójność
: Pomijając przypadek sklepu całodobowego, w nocy pliki są najbardziej stabilne (oczywiście w tym czasie mogą być aktywne zadania wsadowe przetwarzające dane i klienci przeglądający witrynę WWW; w związku z tym nie wszystkie pliki będą dostępne). Jeżeli

[3] Zawsze jest to brane pod uwagę w przypadku każdej rekomendacji podanej w książce. Jeżeli zalecenie jest kłopotliwe dla Czytelnika lub trudne z punktu widzenia używanej przez niego metody archiwizacji, nie jest dobrze! Jeśli kopie zapasowe powodują dyskomfort, nawet nie będzie próbował odtworzyć z nich danych! Administrator systemu zawsze powinien dążyć do upraszczania wszystkiego.

operację archiwizowania wykonuje się w dzień, pliki są modyfikowane i prawdopodobnie również otwarte. Takie pliki trudniej zarchiwizować. Część pakietów archiwizacyjnych radzi sobie z otwartymi plikami lepiej od innych, natomiast część w ogóle nie może ich dodać do kopii zapasowej. Ponadto, jeśli plik jest modyfikowany w ciągu dnia, nie będzie miało się pewności, która jego wersja faktycznie znalazła się w kopii zapasowej.

Szybkość

Następnym powodem, dla którego nie należy tworzyć kopii zapasowych w ciągu dnia, jest większe obciążenie sieci, a tym samym mniejsza szybkość. Przepustowość sieci będąca do dyspozycji procesu archiwizacji znacznie spada, gdy sieć jest normalnie obciążona. Jeżeli archiwizowanie nocą też jest problematyczne, można zastosować specjalną sieć tylko na potrzeby sporządzania kopii zapasowych. Wykonywanie kopii w ciągu dnia może znacząco wpłynąć na szybkość innych aplikacji, a ponadto nie jest dobrą praktyką regularne spowalnianie systemów, gdy korzystają z nich użytkownicy.

Oczywiście w przypadku obecnej globalnej i internetowej ekonomii pojęcie nocy jest względne. W sklepie, w którym komputery są cały czas dostępne, zadania trzeba wykonywać w dość odmienny sposób. Czytelnik może zajrzeć do rozdziału 8., aby dowiedzieć się, co robią producenci, żeby łatwiej poradzić sobie z takim wyzwaniem.

To ma dobre zakończenie (prawie)

Choć nigdy nie zdarzyło mi się, żeby serwer został fizycznie unicestwiony, straciłem całe serwery, konfigurację, wszystko. Na szczęście miałem ich konfigurację zapisaną. Pamiętam też sytuację, w której straciliśmy serwer z firmową bazą danych Informix przechowującą wszystkie dane dotyczące woluminów i ich lokalizacji. Pamiętam, jak powiedziałem: „Jak ja sobie teraz z tym poradzę?". Na szczęście mieliśmy w zwyczaju przesyłać każdego dnia wydruk woluminów do pracowników innego oddziału firmy. Poprosiłem ich o wydruk z zestawieniem wszystkich woluminów z jednego dnia. *Udało się!*

Co to ma być? Czy Czytelnik pomyślał, że jestem niegodziwą i złośliwą osobą, która ma na celu wywoływanie koszmarów następnego tygodnia? Czy Czytelnik nie ma pojęcia, jak zdobyłby takie informacje, gdyby ich potrzebował? Czy powie, że przez jakiś czas nie będzie z tego powodu spał? Dobrze! Lepiej zarwać część nocy, niż stracić dane. Jednym z podstawowych celów tej książki jest przestraszyć Czytelnika. Osoba odpowiedzialna za kopie zapasowe, która jest zadowolona z siebie, stanowi zagrożenie. Opisana sytuacja była frustrująca i spowodowała utratę danych, które normalnie nie są uwzględniane przez standardowe kopie zapasowe.

Decydowanie o metodzie archiwizowania danych

Po podjęciu decyzji o *porze* archiwizowania danych trzeba zastanowić się, *w jaki* sposób zostanie to zrealizowane. Jednak najpierw należy przyjrzeć się problemom, przed którymi można się uchronić.

Trzeba być gotowym na wszystko — 10 rodzajów nieszczęśliwych wydarzeń

Jak już wspomniano, sposób przeprowadzenia procesu odtworzenia danych determinuje metodę tworzenia kopii zapasowych. Jedno z pytań, które trzeba sobie postawić, brzmi: „Przed czym zamierzam się uchronić?". Czy w środowisku pracy znajdują się osoby z większymi uprawnieniami, które korzystają ze swoich komputerów w inteligentny sposób i nigdy nie popełniają głupich błędów? Czy firma straci mnóstwo istotnych danych, gdy pliki znajdujące się na komputerach osobistych użytkowników zostaną przypadkowo usunięte? Czy jeśli huragan zniszczy zupełnie firmę, będzie możliwe jej dalsze funkcjonowanie? Trzeba się upewnić, że ma się świadomość wszystkich potencjalnych powodów utraty danych, a następnie, że na wszystkie ma się przygotowane metody archiwizacji. Najbardziej kompletną listę potencjalnych powodów utraty danych, z którą się spotkałem, można znaleźć w książce *Practical Unix and Internet Security* wydawnictwa O'Reilly, napisanej przez Simsona Garfinkela i Gene'a Spafforda. Poniżej zamieściłem tę listę wraz z moimi komentarzami.

Błąd użytkownika
: Jak dotąd to było powodem największej liczby odtworzeń w każdym środowisku, z którym miałem do czynienia. „Cześć. Coś się stało z moim plikiem i przypadkowo wcisnąłem jakiś klawisz. Czy mogę cię *prosić* o odtworzenie tego pliku?". Czyż to zadanie nie jest naprawdę proste? A co Czytelnik powie na takie oto typowe pytanie: „Czy można odtworzyć coś, co usunięto około godzinę temu?". Będzie to możliwe w przypadku stosowania systemów ciągłej ochrony danych i obrazów. Jednak nie wtedy, gdy kopie zapasowe wykonuje się co noc.

Błąd obsługi technicznej komputerów
: Choć ma to miejsce rzadziej niż błędy użytkowników (chyba że mają uprawnienia administratora), gdy wystąpi, to naprawdę będzie co robić! Co się dzieje, gdy administrator za pomocą narzędzia *newfs* usunie system plików urządzenia bazy danych lub katalog użytkownika z dokumentami? W takich przypadkach operacje odtwarzania muszą być wykonane przez administratora *naprawdę szybko*, ponieważ sam zawinił. Sposoby, dzięki którym administrator może się chronić przed tego typu błędami, są te same co w przypadku błędów popełnianych przez użytkowników. Może wystarczyć typowe wykonywanie nocnych kopii zapasowych lub obrazów danych.

Awaria sprzętu
: Choć większość książek omawia metody ochrony przed awariami sprzętu, zwykle nie wspomina, że awaria może mieć dwie postaci: awarii napędu dysku i awarii całego komputera. Ta uwaga jest ważna, gdyż aby zabezpieczyć się przed tymi typami awarii, trzeba zastosować dwie zupełnie różne metody. Wiele osób nie bierze tego pod uwagę, gdy opracowuje plan ochrony danych. Przykładowo, gdy doszło do uszkodzenia napędu lub systemu plików, często słyszałem coś takiego: „Myślałem, że dla dysku była tworzona kopia lustrzana!". Kopia lustrzana nie chroni przed awarią całego komputera. Mój przyjaciel mawiał, że jeśli wolne elektrony krążące w komputerze postanowią uszkodzić napęd lub system plików, to gdy komputer się zepsuje, kopia lustrzana jedynie sprawi, że zniszczenia będą skuteczniejsze. Również obrazy danych nie uchronią przed awarią sprzętową (jeśli nie znajdują się na woluminie archiwizacyjnym).

Awaria dysku

Obecnie ochrona komputerów przed awarią napędu dysku jest dość prosta. Trzeba jedynie zdecydować, jakiego poziomu bezpieczeństwa się wymaga. Choć mirroring, często określany jako RAID 1, oferuje najlepszą ochronę, podwaja początkowy koszt inwestycji związanej z zakupem napędów i kontrolerów. Z tego właśnie powodu większość osób decyduje się na inne poziomy macierzy RAID (*Redundant Arrays of Independent Disks*), wśród których największą popularnością cieszą się macierze RAID 5 i RAID 6. Woluminy RAID 5 zabezpieczają przed awarią jednego napędu przez obliczanie bloków parzystości i przechowywanie ich na każdym napędzie dysku. Macierz RAID 6 zwiększa poziom ochrony przez przechowywanie dwóch bloków parzystości. Dzięki temu dopuszczalna jest awaria więcej niż jednego napędu.

Awaria komputera

Większa część ochrony przed awarią całego komputera polega na dobrych procedurach administrowania komputerami. Należy właściwie dokumentować komputery. Korzystając z dzienników systemowych i wszelkich innych dostępnych metod monitorowania, należy dokładnie obserwować działanie komputerów. Należy reagować na komunikaty dotyczące wadliwych dysków, kontrolerów, procesorów i modułów pamięci. Ostrzeżenia związane z awariami sprzętu dają szansę usunięcia problemów, zanim spowodują poważniejsze konsekwencje. Inna metoda ochrony polega na zastosowaniu rejestrującego systemu plików. Rejestrowanie systemu plików wygląda podobnie jak w przypadku bazy danych. Monitorowane są zatwierdzone i częściowo zatwierdzone operacje zapisu w systemie plików. Gdy komputer jest uruchamiany, rejestrujący system plików może wycofać częściowo zatwierdzone operacje zapisu, tym samym zapobiegając uszkodzeniu systemu plików.

Dziennik zmian systemu Windows nie powoduje, że system plików NTFS staje się rejestrującym systemem plików. Dziennik zmian zawiera jedynie listę plików, które zostały zmodyfikowane, bez samych zmian. W związku z tym nie jest w stanie wycofać żadnych zmian.

Awaria oprogramowania

Ochrona przed awarią oprogramowania może być trudna. Błędy systemów operacyjnych, baz danych i aplikacji zarządzających systemami mogą spowodować utratę danych. Również w tym przypadku stopień możliwej ochrony przed tego typu błędami zależy od rodzaju wykonywanych kopii zapasowych. Często tworzone obrazy danych lub systemy ciągłej ochrony danych to jedyne rozwiązania pozwalające rzeczywiście zabezpieczyć się przed utratą danych (może być ich sporo) w wyniku awarii oprogramowania.

Elektroniczne włamania, wandalizm i kradzież

W ostatnich latach miało miejsce kilka takich incydentów. O wielu z nich można było usłyszeć w ogólnokrajowych wiadomościach. Jeśli traci się dane na skutek dowolnego z takich zdarzeń, ma się do czynienia z zupełnie odmienną sytuacją. Choć można odtworzyć dane, możliwe jest, że nigdy nie uzyska się pewności, co się z nimi stało, gdy miał do nich dostęp ktoś inny. A zatem trzeba zrobić wszystko, aby zagwarantować, że nigdy do tego nie dojdzie. Jeżeli ktoś chce się uchronić przed utratą danych w taki sposób, mocno go zachęcam do przeczytania książki *Practical Unix and Internet Security* autorstwa Simsona Garfinkela i Gene'a Spaffordа (wydawnictwo O'Reilly), z której zapożyczyłem tę listę powodów utraty danych.

Klęski żywiołowe

Czy jest się przygotowanym na huragan, tornado, trzęsienie ziemi lub powódź? Jeśli nie, nie jest się w tym osamotnionym. Można sobie wyobrazić, że całe województwo zostało zniszczone. Jeżeli korzysta się z magazynowania kopii zapasowych poza obrębem firmy, czy miejsce składowania jest blisko firmy? Czy magazyn jest przygotowany na każdego typu klęskę żywiołową, która wystąpi na obszarze województwa? Jeśli na przykład biuro i firma magazynująca jego dane znajdują się na obszarze powodzi, czy przechowuje ona kopie zapasowe na pierwszym piętrze? Jeśli nie, dane mogą zostać utracone po jednym intensywnym deszczu. Aby naprawdę uchronić się przed poważną klęską żywiołową, należy znaleźć odległe miejsce składowania kopii zapasowych (więcej na ten temat można znaleźć w dalszej części rozdziału, w punkcie „Przechowywanie poza obrębem firmy").

„Jak udały się kopie zapasowe z zeszłej nocy?"

Chyba słyszałem już tysiące dramatycznych historii o błędach popełnianych przez administratorów (takich jak wykonanie polecenia `rm -r /*`). Pamiętam kogoś, kto chciał usunąć niepotrzebny plik zawarty w katalogu */bin* i mający nazwę *?*&(&^JI($SF))FS%$#T* lub podobną. W tym celu wykonał polecenie `rm /bin/?*` (usuwa wszystkie pliki, których nazwa zaczyna się dowolnym znakiem). Jednak zdarzyło się także coś, czego byłem naocznym świadkiem. W dalszym ciągu mnie to rozbawia.

Konsultantowi zlecono wyczyszczenie katalogów domowych użytkowników. Zdaje się, że firma, w której byłem zatrudniony, bardzo dobrze radziła sobie z usuwaniem kont osób odchodzących z pracy, lecz już nie tak dobrze wyglądało to w przypadku usuwania ich katalogów domowych. Konsultant napisał program, który przede wszystkim wykonywał następujące operacje:

1. Po wpisaniu polecenia `cd` przechodził do katalogu */home1*.
2. Przy użyciu narzędzia *find* szukał katalogów, dla których w pliku haseł nie istniały wpisy, a ponadto ich właścicielem nie był użytkownik *root* lub inny administrator.
3. Dla katalogu wykonywał polecenie `rm -r`.

Katalog domowy każdego użytkownika znajdował się poniżej katalogu, którego nazwą była pierwsza litera identyfikatora użytkownika. Przykładowo ścieżka katalogu domowego użytkownika *cnowak* wyglądała tak: */home1/c/cnowak*. Pora opisać to, co się wydarzyło. Zamysł był taki, że właścicielem katalogu */home1/c* jest użytkownik *root*, dlatego katalog nie zostanie usunięty. Niestety, raz jeden lub dwóch administratorów poleceniem `cd` przeszło do katalogu */home1/c/cnowak*, aby spróbować wyeliminować problem z prawem własności. W tym celu administrator wykonał polecenie `chown cnowak .*`. Jeśli kiedykolwiek Czytelnik zastosował to polecenie z uprawnieniami użytkownika *root*, to wie, że `.*` uwzględnia `..`, co w tym przypadku oznacza */home1/c*. A zatem właścicielem katalogu */home1/c* staje się użytkownik *cnowak*!

Konsultant nie przewidział tego i zinterpretował katalog */home1/c* jako katalog domowy, a następnie poszukał w pliku haseł użytkownika *c*. Oczywiście, w pliku nie było takiego użytkownika, dlatego program wykonał polecenie `rm -r /home1/c`. Nie jestem pewien, kiedy mój znajomy zdał sobie sprawę z tego, co się stało, ale pamiętam, że otrzymałem dziwny telefon, gdy wychodziłem z domu. Kolega zapytał mnie tajemniczo i z zakłopotaniem: „Jak udały się kopie zapasowe katalogu */home1* z zeszłej nocy?". „Świetnie, jak zawsze — odpowiedziałem. — A dlaczego pytasz?". Jest coś pięknego w mocy, jaką emanuje osoba archiwizująca dane w tej magicznej chwili, gdy ktoś *naprawdę* musi mieć możliwość odtworzenia kilku plików. Tak więc jest się kimś, kto przychodzi wcześnie i zostaje do późna, obserwując obracające się napędy archiwizacyjne. W jednej chwili można stać się dla kogoś najważniejszą osobą! *To jest fajne*.

Inne nieszczęśliwe zdarzenia

Pamiętam, jak testowaliśmy nasz plan odtwarzania danych po awarii dla firmy, w której pracowałem. Przyjęliśmy, że cysterna wybuchła na ulicy, która biegła przy centrum danych firmy. Plan polegał na odtworzeniu danych w alternatywnym budynku. Oznaczało to, że trzeba by było przechowywać nośniki z danymi poza obrębem firmy i dysponować dodatkowym budynkiem przystosowanym tak, aby pomieścił wszystkie używane systemy. Dobrym sposobem jest tu rozmieszczenie systemów produkcyjnych i projektowych w różnych budynkach. Jeśli systemy produkcyjne zepsują się lub nastąpi przerwa w zasilaniu budynku, w którym się znajdują, mogą zostać zastąpione przez systemy projektowe.

Zarchiwizowane informacje

Okropnie jest uświadomić sobie, że zaginął rzadko używany, lecz bardzo ważny plik. Jednak jeszcze bardziej przerażająca jest sytuacja, kiedy plik jest niedostępny dłużej, niż trwa cykl retencji. Przykładowo kopie zapasowe przechowuje się tylko przez trzy miesiące, po czym ponownie wykorzystuje najstarszy wolumin, nadpisując wszelkie zapisane na nim dane. Jeśli coś takiego ma miejsce, *nie będzie możliwe* odtworzenie żadnych plików, które nie istnieją od ponad trzech miesięcy. Nie jest istotne, jak bardzo użytkownik będzie podkreślał ważność plików, ani to, ile telefonów wykona do przełożonych administratora. Nie będzie on w stanie *nigdy* odtworzyć tych plików. Właśnie z tego powodu część kopii zapasowych powinno się przechowywać trochę dłużej. Normalną praktyką jest przeznaczenie każdego miesiąca jednej pełnej kopii zapasowej do magazynowania przez kilka lat. Jeżeli zamierza się przechowywać takie kopie przez długi czas, trzeba zapoznać się z zawartością ramki „Czy magazynuje się archiwa zbyt długo?", a także z podrozdziałem „Kopie zapasowe a archiwa danych" znajdującym się w rozdziale 24.

Automatyzowanie archiwizowania danych

Jeśli Czytelnik pracuje w sklepie o umiarkowanym budżecie, prawdopodobnie po spojrzeniu na tytuł tego podrozdziału stwierdził: „Byłoby fajnie, gdybyśmy mogli sobie na to pozwolić". Choć automatyzacja wykorzystująca kosztowne szafy i zmieniarki jest fajna, nie o taką automatyzację mi chodzi. Można wyróżnić dwa typy automatyzacji. Pierwszy umożliwia przeprowadzenie całego cyklu tworzenia kopii zapasowych bez konieczności ręcznej interwencji operatora, takiej jak wyciąganie i wsuwanie nowych woluminów. Choć ten typ automatyzacji może znacznie wszystko ułatwić, wiąże się z większymi kosztami. Jeżeli nie można sobie na to pozwolić, tańszą opcją jest nakazać systemowi archiwizowania powiadamianie operatora, że sam musi wykonać jakąś czynność. W najgorszym razie system powinien powiadomić, że trzeba (lub zapomniano) wymienić wolumin. Jeśli coś pójdzie nie tak, trzeba o tym wiedzieć. Zbyt wiele razy ludzie przeglądali dzienniki archiwizacji tylko wtedy, gdy musieli przeprowadzić proces odtwarzania danych. Właśnie w takim momencie dowiadują się, że od kilku dni lub tygodni nie udało się wykonać kopii zapasowych. Bardziej „inteligentny" system archiwizowania może wysłać do operatora wiadomość pocztową lub komunikat na pager, jeśli coś nie pójdzie zgodnie z oczekiwaniami.

Drugi typ automatyzacji jest znacznie bardziej istotny. Odnosi się on do sposobu „myślenia" mechanizmu tworzenia kopii zapasowych. Mechanizm archiwizacji powinien wiedzieć, co ma uwzględnić w kopii zapasowej, bez interwencji operatora. Jeżeli administrator bazy danych instaluje nową bazę, mechanizm archiwizacji powinien o tym wiedzieć. Jeśli administrator systemu instaluje nowy napęd lub system plików, kopie zapasowe powinny go automatycznie

uwzględnić. Jest to rodzaj automatyzacji, który ma zasadnicze znaczenie dla zapewnienia poprawnych kopii zapasowych. Dobry system archiwizowania nie powinien być zależny od tego, że ludzki mózg ma coś zapamiętać.

Czy magazynuje się archiwa zbyt długo?

Niektóre organizacje rządowe ustalają przepisy, które mówią, jak długo określonego typu dane mogą być przechowywane w archiwach firmy. Nie chodzi o przepisy, które nakazują składować dane przez określoną liczbę lat, ale o te, które narzucają obowiązek *usuwania* danych po upływie iluś lat. Przykładowo można zostać poinformowanym, że dział kadr może przechowywać dane dotyczące nagan tylko przez dwa lata. Jeśli pracownik przypuszcza, że jego szanse na awans zmniejszają się, ponieważ akta nagan mają więcej niż dwa lata, może zaskarżyć firmę i żądać odszkodowania za poniesione szkody. W aktach sądowych znalazło się wiele spraw bazujących na takich przepisach.

Co się stanie, gdy akta dotyczące nagan mają w rzeczywistości postać pliku znajdującego się na czyimś komputerze? Prawo dotyczy również komputerów. Pliki muszą zostać usunięte. Co jednak będzie, gdy plik zapisano na woluminie archiwizującym przechowywanym przez bliżej nieokreślony okres? Wiele firm ustala zasady archiwizowania danych, które nakazują, aby co roku dla każdego systemu przez nieokreślony czas magazynować jeden wolumin. W ostatnich latach część firm przegrała sprawy sądowe z powodu stosowania takich zasad.

Jedynym sposobem obejścia takich komplikacji jest wyłączenie ze zwykłych kopii zapasowych wszelkich katalogów zawierających tego typu informacje i archiwizowanie ich w ramach innego harmonogramu, zgodnego z lokalnymi przepisami dotyczącymi przechowywania dokumentów. Przyznam, że jest to uciążliwe. W mojej książce Czytelnik nie przeczyta, że bezcelowe robienie czegoś specjalnego jest czymś dobrym. Jednak w tych czasach pełnych konfliktów nie należy pomijać tej kwestii.

Uwzględnianie w planie rozbudowy

Inny powszechny problem ma miejsce, gdy z czasem system archiwizowania jest rozbudowywany. To, co działa w przypadku jednego lub dwóch komputerów, niekoniecznie poradzi sobie z 200 komputerami. Wraz ze wzrostem objętości danych zwiększa się zapotrzebowanie na ustandaryzowany system archiwizowania. Jest to problem, ponieważ większość administratorów, tworząc skrypt powłoki, który wykonuje kopie zapasową danych z pięciu lub sześciu komputerów, nie uwzględnia tego, że w przyszłości komputerów może być więcej. Pamiętam swoje początki, gdy byłem operatorem kopii zapasowych. Miałem pod opieką 10 lub 11 komputerów. Jeden z nich, z systemem Ultrix, robił wrażenie. W tych czasach mówiliśmy, że był ogromny (przechowywał prawie 8 gigabajtów danych!). Największy napęd taśmowy Exabyte, jakim dysponowaliśmy, oferował pojemność 10 GB (z kompresją). Dla komputera z danymi o pojemności 8 GB stosowaliśmy duży napęd taśmowy 10 GB. Mieliśmy skrypt archiwizujący, moim zdaniem naprawdę dobry, który bez żadnych modyfikacji działał przez dwa lata.

Wtedy pojawiły się komputery HP. *Najmniejszy* z nich miał pojemność 20 GB, największy oferował znacznie więcej. Jednak te duże serwery były wyposażone w niewielki napęd DDS o pojemności 2 GB (4 GB z kompresją). Twórca skryptu archiwizującego nigdy nie pomyślał o komputerze, który będzie przechowywał dane o objętości większej niż pojemność taśmy. Pewnego dnia obudziłem się i dowiedziałem, że nasz serwer miał awarię. Później wiele miesięcy

zajęło mi tworzenie skryptu powłoki, który obsługiwałby dzielenie zawartości napędu lub systemu plików na dwie taśmy. W końcu zrezygnowałem i kupiłem komercyjny produkt. Zmierzam do tego, że jeśli pomyślałbym o tym wcześniej, być może byłbym w stanie poradzić sobie z wyżej przedstawionym ograniczeniem bez siedzenia tyle po nocach.

Gdy projektuje się system archiwizowania lub centrum danych, trzeba wziąć pod uwagę to, że komputery będą pojemniejsze i zwiększy się ich liczba. Ponadto należy zaplanować, co się zrobi, gdy stanie się coś złego. Można mi wierzyć, że tak *będzie*. Znacznie lepiej dla zdrowia psychicznego (nie wspominając o gwarancji utrzymania pracy) będzie przewidzieć nieuniknione i uwzględnić to od razu podczas projektowania systemu archiwizowania. System ten jest czymś, co powinno być wykonane dobrze już za pierwszym razem. Jeśli *przed* zaprojektowaniem systemu archiwizowania poświęci się trochę czasu na wyobrażenie sobie, w jaki sposób można spowodować jego awarię, zaoszczędzi się wiele pieniędzy na tabletkach nasennych i przeciwbólowych.

Nie należy zapominać o uniksowych wartościach mtime, atime i ctime

Systemy operacyjne Unix, Linux i Mac OS dla każdego pliku rejestrują trzy różne czasy. Pierwszy to czas modyfikacji *mtime*. Wartość *mtime* jest zmieniana każdorazowo, gdy plik zostaje zmodyfikowany (na przykład po dodaniu wierszy do pliku dziennika). Następny czas to czas dostępu *atime*. Wartość *atime* jest modyfikowana zawsze, gdy plik jest używany (na przykład po uruchomieniu skryptu lub odczytaniu dokumentu). Ostatnim czasem jest czas zmiany *ctime*. Wartość *ctime* jest aktualizowana każdorazowo, gdy są modyfikowane atrybuty pliku, takie jak uprawnienia lub prawo własności.

Administratorzy korzystają z wartości *ctime*, szukając śladów włamań hakerów, którzy mogą zmieniać uprawnienia pliku lub próbować lokalizować słabe punkty systemu. Administratorzy monitorują też wartość *atime*, gdy szukają dużych plików nieużywanych od dłuższego czasu (takie pliki mogą zostać zarchiwizowane, a następnie usunięte).

Kopie zapasowe modyfikują wartość atime

Można się zastanawiać, co ma to wspólnego z kopiami zapasowymi. Trzeba wiedzieć, że każde narzędzie wykonujące kopię zapasową przy wykorzystaniu systemu plików podczas odczytu archiwizowanego pliku modyfikuje wartość *atime*. Niemal wszystkie komercyjne narzędzia, a także programy: *tar*, *cpio* i *dd*[4], pozwalają na coś takiego. Ponieważ program *dump* odczytuje system plików za pośrednictwem pliku urządzenia, nie zmienia wartości *atime*.

Wartość atime może zostać zresetowana, lecz nie bez konsekwencji

Zanim program archiwizujący umieści plik w kopii zapasowej, może sprawdzić jego wartość *atime*. Po zarchiwizowaniu pliku wartość *atime* oczywiście zmieni się. Program może następnie za pomocą wywołania systemowego *utime* zresetować wartość *atime* do jej oryginalnego ustawienia. Jednak zmiana wartości *atime* jest uważana za modyfikację atrybutu, z czym wiąże się zmiana wartości *ctime*. Oznacza to, że gdy użyje się takiego narzędzia, jak *cpio* lub *gtar*,

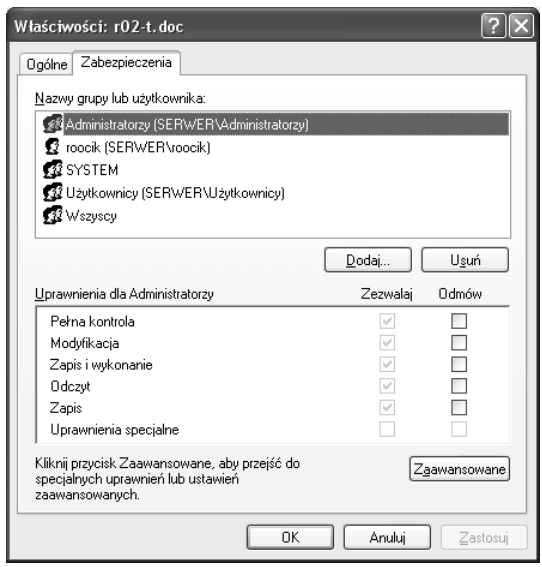
[4] Oczywiście program *dd* modyfikuje wartość *atime*, gdy za jego pomocą kopiuje się pojedynczy plik systemu plików. Jeśli przy użyciu tego programu kopiuje się plik urządzenia, nie zostaną zmienione czasy dostępu plików systemu plików.

które może resetować wartość *atime*, dla każdego zarchiwizowanego pliku zostanie zmieniona wartość *ctime*. Jeśli ktoś dysponuje systemem monitorującym zmiany wartości *ctime*, bez wątpienia uzna, że ma do czynienia z hakerem!

Zrozumienie, w jaki sposób używane narzędzie traktuje wartości *atime* i *ctime*, jest niezbędne.

Nie wolno zapomnieć o listach kontroli dostępu ACL

Pliki systemu Windows zapisane w systemie plików NTFS i część plików nowszych linuksowych systemów plików korzysta z list kontroli dostępu ACL (*Access Control List*), umożliwiających nadawanie lub odbieranie użytkownikom uprawnień. Listy ACL decydują o tym, kto może odczytywać, zapisywać, wykonywać i modyfikować plik lub mieć nad nim pełną kontrolę. Rysunek 2.1 przedstawia przykład list ACL.



Rysunek 2.1. Przykład listy kontroli dostępu

Trzeba sprawdzić, jak wykorzystywany produkt archiwizujący obsługuje listy ACL. Właściwa metoda polega na archiwizowaniu i odtwarzaniu list ACL. Choć w przypadku komercyjnych produktów jest to standardowa funkcja, niestety, nie wszystkie programy open source ją oferują. Podczas analizowania narzędzi open source trzeba pamiętać o sprawdzeniu kwestii przetwarzania list ACL.

Trzeba pamiętać o rozwidleniu zasobów systemu Mac OS

Pliki systemu Mac OS zapisane w systemach plików: MFS, HFS lub HFS Plus mają dwa rozwidlenia: danych i zasobów. *Rozwidlenie danych* zawiera rzeczywiste dane pliku, takie jak tekst. *Rozwidlenie zasobów* przechowuje powiązane dane strukturalne pliku, takie jak przesunięcia, menu, okna dialogowe i ikony. Dwa rozwidlenia są ze sobą ściśle powiązane, tworząc jeden plik. Choć zwykle rozwidlenie zasobów jest stosowane przez pliki wykonywalne, może z niego

korzystać każdy plik, a także inne aplikacje. Przykładowo program przetwarzający tekst może przechowywać tekst pliku w rozwidleniu danych, a obrazy pliku — w rozwidleniu zasobów.

Podobnie do list ACL systemu Windows rozwidlenia zasobów wymagają archiwizowania. Jednak nie wszystkie produkty archiwizujące wykonują to w poprawny sposób. Trzeba się upewnić, co system archiwizowania robi z rozwidleniem danych i zasobów.

Należy dbać o maksymalne uproszczenie

W przypadku kopii zapasowych stwierdzenie to jest dwukrotnie lub trzykrotnie ważniejsze. Im bardziej skomplikowany schemat archiwizowania, tym bardziej prawdopodobne, że zawiedzie. Jeśli schemat nie jest zrozumiały, nie będzie można go wdrożyć. Trzeba pamiętać o tym każdorazowo, gdy do systemu archiwizowania dodaje się nowy mniej lub bardziej ważny element. Każda zmiana powoduje zagrożenie danych. Ponadto każda modyfikacja może jeszcze bardziej zwiększyć złożoność systemu archiwizowania i sprawić, że będzie trudniejszy do objaśnienia nowej osobie odpowiedzialnej za archiwizowanie. Jeden z szefów działu obsługi technicznej komercyjnego produktu archiwizującego stwierdził, że cały czas widzi to samo. Ktoś naprawdę dobrze opanuje oprogramowanie i napisze różne skrypty automatyzujące określone zadania. Kopie zapasowe są dobrze naoliwioną maszyną do momentu, gdy zajmie się nimi praktykant. Ponieważ nie rozumie on wszystkich szczegółów systemu, wszystko zaczyna zawodzić. Nagle dane znajdują się w niebezpieczeństwie. Trzeba o tym pamiętać następnym razem, gdy postanowi się dodać do skryptu archiwizującego jakąś nową fajną funkcję.

Poniższy komentarz dotyczy również wcześniejszego podrozdziału „Uwzględnianie w planie rozbudowy". Jednym z często popełnianych błędów decyzyjnych jest rezygnowanie ze stosowania automatyzacji od samego początku. Znacznie łatwiej po prostu umieścić gdzieś na stałe w pliku listę dołączeń lub wstawić ją do konfiguracji narzędzia *cron* bądź w samym wpisie zaplanowanego zadania. Jednak w takim przypadku metod archiwizowania jest wiele. Jeśli każdy komputer ma własny, specjalnie dostosowany system archiwizowania, bardzo trudno monitorować mechanizm kopii zapasowych i objaśnić go nowemu pracownikowi.

Należy pamiętać, że słowo *specjalne* nie ma pozytywnego znaczenia. Wystarczy to na okrągło powtarzać, aby w końcu w to uwierzyć.

Nic szczególnego się nie stanie, gdy zarządza się dwoma lub trzema komputerami. Jednak sytuacja zmieni się w przypadku 200 komputerów. Jeżeli trzeba zapamiętywać szczegóły związane z każdym systemem każdorazowo podczas przeglądania dzienników, bez wątpienia wszystko wymknie się spod kontroli. Wyjątki charakterystyczne dla poszczególnych systemów oznaczają również, że pewne rzeczy się przeoczy. Czy administrator pamięta, że dziewięć miesięcy temu dla komputera *apollo* zdefiniowano wykluczenie katalogu */home**? Lepiej, żeby pamiętał, gdyż właśnie *apollo* stał się głównym serwerem NFS i ma już siedem katalogów *home*.

Jeśli obcej osobie nie można objaśnić systemu archiwizowania w najwyżej kilka godzin, prawdopodobnie wszystko jest zbyt skomplikowane. Powinno się przyjrzeć kwestiom związanym z wdrożeniem, takim jak scentralizowane rejestrowanie, ustandaryzowane skrypty archiwizujące i określony poziom automatyzacji.

Należy czytać instrukcje

Adres IP serwera archiwizującego dużej firmy programistycznej był cały czas zmieniany na inny, prawdopodobnie losowo wybrany. Jedyny identyfikowalny schemat był taki, że każdy nowy adres IP, który zostałby przypisany serwerowi, był adresem jednego z archiwizowanych klientów. Skontaktowano się z działem obsługi dostawców, a ponadto nad rozwiązaniem problemu nieustannie pracowali wszyscy inżynierowie. Jednak nikt nie potrafił znaleźć przyczyny problemu.

Okazało się, że operator archiwizacji wyznaczony do rozwiązywania problemów z kopiami zapasowymi stosował standardowe procedury diagnostyczne ustalone dla grupy. Jednak nowy operator pomieszał kilka poleceń. W efekcie, gdy próbował przeprowadzić podstawową operację rozpoznawania nazw archiwizowanych hostów, zamiast polecenia `nslookup` *`nazwa_hosta`* było wykonywane polecenie `ifconfig -a` *`nazwa_hosta`*. Powodowało ono zmianę adresu IP serwera archiwizującego na adres dowolnego hosta, który miał problemy z archiwizowaniem. Działo się to w losowych porach dnia i tylko w te dni, w które pracował nowy operator.

— Jorgen Lie

Przechowywanie kopii zapasowych

Żadnego pożytku nie będzie z naprawdę dobrych kopii zapasowych, gdy dojdzie do zniszczenia, zgubienia lub pomieszania woluminów archiwizacyjnych. Trzeba dysponować dobrze zdefiniowanym procesem magazynowania nośników.

Ogólne informacje na temat przechowywania

Jeśli Czytelnik przeczytał wcześniejszą treść książki, wie, że kopie zapasowe traktuję bardzo poważnie. Jeżeli wykonane kopie zapasowe są istotne, czy nośnik, na którym je umieszczono, nie jest równie ważny? Czyż nie jest to oczywiste? Cóż, nie wywnioskuje się tego na podstawie obserwacji większości „bibliotek" woluminów. "Stosy" woluminów to raczej bardziej odpowiednie określenie. Z iloma pomieszczeniami komputerowymi, w których woluminy były rozmieszczone w różnych miejscach, Czytelnik miał do czynienia? Woluminy są układane w stosy, stawiane za komputerami, a napęd taśmowy naprawdę dobrze sprawdza się w roli podstawki na kawę (prawda, że nie chcemy, aby na nowym napędzie zostały jakiekolwiek okręgi po filiżance kawy?).

Czy kiedykolwiek Czytelnik potrzebował woluminu i nie mógł go znaleźć? Zdarzyło mi się to. Jest to okropne uczucie, gdy wie się, że plik zapisano na woluminie, lecz nie można znaleźć tego nieszczęsnego woluminu! Dlaczego zatem traktuje się woluminy archiwizujące jak brudne rzeczy do prania? Trzeba uporządkować woluminy! Woluminom należy nadać etykiety, skatalogować je, przypisać im unikatowe nazwy lub liczby i umieścić w określonego typu pojemniku z zachowaniem jakiegoś logicznego uporządkowania. Jeśli tego nie zrobimy, nawiedzi nas demon archiwizowania!

Możliwość szybkiego przeprowadzenia dużej operacji odtwarzania zależy bezpośrednio od tego, jak dobrze uporządkowano nośniki.

Lokalne przechowywanie

Jak wygląda sprawa szafy używanej do lokalnego przechowywania nośników z woluminami? Czy Czytelnik jej nie posiada? A może korzysta z szafki na akta? Można zastosować dowolny pojemnik. Jeśli jednak można sobie pozwolić na pojemniki przeznaczone do magazynowania nośników, warto je kupić. . Nikt nie pożałuje takiego zakupu. Przeprowadzanie operacji odtwarzania jest znacznie mniej stresujące, gdy bez problemów można znaleźć wolumin. Kilka firm produkuje takie pojemniki. Można także kupić szafy ognioodporne. Jednak należy pamiętać, że ognioodporność nie oznacza żaroodporności. Tego typu pojemniki na nośniki mają za zadanie ochronić przed krótkotrwałym ogniem, który szybko zostanie ugaszony przez system zraszania. Jeśli tuż przy pojemniku przez długi czas będzie płonął ogień lub w pomieszczeniu znacząco wzrośnie temperatura, z woluminów może nie być żadnego pożytku (jest to kolejny dobry powód, dla którego woluminy trzeba również składować poza obrębem firmy).

Niech najbardziej zorganizowana osoba w firmie zaprojektuje system magazynowania nośników. Oto moja propozycja. Warto poprosić najlepszego pracownika działu administracyjnego, aby przyjrzał się systemowi magazynowania i porównał go ze swoją szafą na dokumenty. Należy wyjaśnić, że oczekuje się najszczerszej opinii.

12 000 złotych sztab

Instytucja finansowa, w której raz pracowałem, miała magazyn liczący ponad 12 000 nośników. Nigdy nie zdarzyło się, żeby przepadł choćby jeden z nich. Jak się to nam udało? Traktowaliśmy każdy wolumin, jakby był sztabą złota. Nasz system magazynowania był oparty na kilku elementach:

- *Każdy wolumin* miał unikatowy liczbowy identyfikator.
- Identyfikator miał postać kodu paskowego umieszczonego na woluminie (mogę zapewnić, że etykietowanie więcej niż 500 oryginalnych 5,25-calowych dyskietek instalacyjnych dodanych do serwerów AT&T 3b2/1000 nie było niczym przyjemnym; jednak udało się nam to zrobić z pomocą grupy stażystów!).
- Numer, nazwa, przeznaczenie, typ nośnika, data użycia i lokalizacja każdego woluminu były przechowywane w bazie danych Informix.
- Ta sama baza danych rejestrowała każdą operację przeniesienia woluminu. Gdy wolumin zabrano do innego budynku w celu przeprowadzenia archiwizacji lub odtworzenia, było to zapisywane w bazie. Jeśli wolumin wysłano do dostawcy zewnętrznego magazynu, informacja o tym trafiała do bazy danych. Jeżeli administrator wypożyczył wolumin archiwizujący lub instalacyjny dysk CD, było to rejestrowane w polu „Wypożyczone".
- Gdy w celu odtworzenia danych na krótki okres wyciągano z biblioteki jeden z nośników, odpowiednią uwagę zamieszczano ręcznie w dzienniku. W przypadku wykonywanych w ciągu dnia dużych operacji przenoszenia dużej liczby woluminów posługiwaliśmy się czytnikiem kodów kreskowych współpracującym ze skryptem powłoki automatycznie aktualizującym bazę danych.
- Co kwartał przeprowadzaliśmy pełną inwentaryzację, a raz w miesiącu wybiórczą. Jeśli po wykonaniu wybiórczej inwentaryzacji okazało się, że jest zbyt wiele błędów, oznaczało to konieczność przeprowadzenia następnej pełnej inwentaryzacji.

- Podczas inwentaryzacji porównywaliśmy każdy wolumin z wydrukiem z bazy danych, a także każdy wpis wydruku z rzeczywistym woluminem (druga część procesu uwzględniała wychwytywanie administratorów, którzy w swoich biurkach pochowali kopie zapasowe lub nośniki instalacyjne).
- Woluminy były przechowywane w zamykanych na klucz szafach firmy Wrightline. Dostęp do woluminów mieli wyłącznie operatorzy archiwizacji (były to te same osoby, które odpowiadały za zaginięcie jakiegoś nośnika).
- Inwentaryzacje były nazywane samoinspekcjami. Raz w roku dodatkowo dział inspekcji dokonywał wewnętrznego audytu, a zewnętrzny audyt przeprowadzała organizacja OCC (*Office of the Comptroller of the Currency*). Jej pracownicy przeglądali dzienniki, szukając niespójności. Posiadali umiejętność identyfikowania trochę dziwnie wyglądających wpisów. Gdy na takie natrafili, mówili: „Czy możemy to zobaczyć…".
- Cały proces był dokładnie dokumentowany. Choć prawdopodobnie instytucja ulepszyła nieznacznie te procedury, w dalszym ciągu z nich korzysta.

Organizacja OCC traktuje kwestię ochrony danych bardzo poważnie (nawiasem mówiąc, w Stanach Zjednoczonych organizacja ta ma możliwość zakazania komuś prowadzenia działalności bankowej). Trzeba się upewnić, że tego typu organizacja jest usatysfakcjonowana procedurami stosowanymi przez kontrolowaną firmę.

Przechowywanie poza obrębem firmy

Po uporządkowaniu nośników przechowywanych lokalnie pora wziąć pod uwagę magazynowanie ich na zewnątrz. Można wyróżnić dwie metody składowania danych poza obrębem firmy:

- zdalne składowanie danych na nośnikach (z użyciem taśm).
- zdalne elektroniczne składowanie danych (bez użycia taśm).

Choć druga metoda może być kosztowna, nie jest tak droga, jak niektórzy myślą. Ponadto znacznie łatwiej z niej skorzystać w przypadku nieszczęśliwych wydarzeń i nie straci się taśm, gdyż takowych nie ma. Oczywiście właśnie na coś takiego (zniszczenie nośników lub budynku, w którym się znajdują) ma przygotować magazynowanie kopii zapasowych poza obrębem firmy. Jeśli dysponuje się w innym miejscu kompletnym zestawem kopii zapasowych, będzie można odtworzyć dane nawet w przypadku najgorszej lokalnej klęski żywiołowej.

Wybieranie dostawcy zdalnego składowania danych na nośnikach

Zadanie to jest tak istotne jak wybieranie oprogramowania archiwizującego. Wybranie złego dostawcy może być zgubne w skutkach. Będzie się od niego zależnym, a przecież pełni on rolę ostatniej linii obrony. Właśnie za to mu się płaci. W związku z tym stosowane przez dostawcę procedury przechowywania i segregowania muszą być bez zarzutu. Muszą być lepsze niż w scenariuszu przedstawionym w podrozdziale „12 000 złotych sztab". Procedura rejestrowania operacji przenoszenia używana przez dostawcę musi być pozbawiona luk. Oto lista rzeczy, które należy uwzględnić podczas wybierania dostawcy usługi zewnętrznego magazynowania:

Inwentaryzowanie poszczególnych nośników
: Pierwszy dostawca usługi składowania danych na nośnikach, z którego oferty skorzystałem, wszystkie moje woluminy przechowywał w pudełkach. Nigdy nie inwentaryzował poszczególnych nośników. Sam musiałem wiedzieć, w którym pudełku znajdował się

każdy wolumin. Gdy potrzebowałem woluminu z jednego z pudełek, pracownik dostawcy musiał pojechać po niego i mi go przywieźć. Wtedy jednak w żaden sposób nie było rejestrowane, gdzie wolumin faktycznie się znajdował. Coś takiego określa się mianem *składowania kontenerowego*. Większość firm zdalnie składujących dane na nośnikach oferuje również magazynowanie poszczególnych nośników. Rozwiązanie to zapewnia monitorowanie każdego woluminu.

Inwentaryzowanie z użyciem kodu kreskowego, bazujące na lokalizacji
: Każdy wolumin powinien mieć kod kreskowy umożliwiający dostawcy usługi składowania skanowanie wszystkich pobranych i oddanych woluminów. Dostawca powinien wczytywać do bazy magazynu kody kreskowe otrzymanych woluminów i usuwać z bazy kody woluminów oddanych klientowi.

Podwójna elektroniczna kontrola
: Jeśli samemu rejestruje się położenie każdego woluminu i to samo robi dostawca, obie strony powinny się wzajemnie podwójnie sprawdzać. Jedna lub obie strony mogą wydrukować zestawienie lokalizacji woluminów wyeksportowane z bazy danych. Można napisać program, który porównuje położenie każdego woluminu w dwóch różnych magazynach. Nie jestem w stanie powiedzieć, ile razy taki program oszczędził mi problemów. Dobrze jest znaleźć błąd w chwili, gdy wystąpi, niż tygodnie później, gdy trzeba użyć woluminu, który zaginął.

Testowanie usługi wybranego dostawcy

Należy sprawdzić, czy dostawca jest czujny. Jeden z umożliwiających to sposobów polega na przekonaniu się, czy dostawca pozwoli na zostawienie nas samych w magazynie. Ponieważ jest się klientem takiej firmy, można jej pracowników poprosić o zgodę na przeprowadzenie inwentaryzacji nośników. Trzeba upewnić się, czy firma udzieli nieograniczonego dostępu do wnętrza magazynu. Jeśli pracownicy pozostawią klienta w środku bez żadnego nadzoru, uzyska on dostęp do nośników innych firm. Oznacza to, że osoby z innych firm mogą mieć dostęp do naszych nośników. Od takiego dostawcy należy nie odejść, lecz uciec.

Należy organizować niespodziewane inspekcje, a także wyrywkowe kontrole. Należy poprosić o oddanie losowych woluminów, aby się przekonać, jak szybko dostawca jest w stanie je odnaleźć. Należy poprosić o właśnie wysłane woluminy. Woluminy będące w trakcie inwentaryzowania są najtrudniejsze do odnalezienia. Jednak dostawca powinien być w stanie to zrobić. Jeżeli regularnie każdego dnia wysyła się dostawcy pięć woluminów do zinwentaryzowania, pewnego dnia należy przekazać cztery woluminy, lecz w dokumentach zlecić inwentaryzację pięciu woluminów. Ma to na celu sprawdzenie, czy dostawca zwróci na to uwagę. Jeśli nie, trzeba interweniować! Procedury dostawcy powinny chronić klienta przed tego typu ludzkimi błędami. Jeżeli tak nie jest, procedury wymagają ulepszenia. Trzeba być nieprzewidywalnym, w przeciwnym razie można zostać zignorowanym. Sprawienie, że dostawca będzie cały czas czujny, spowoduje, że dobrze zapamięta klienta i to, jak ważne są dla niego jego woluminy (przy okazji należy wspomnieć, że możliwość przeprowadzania niespodziewanych inspekcji i wyrywkowych kontroli powinna być zawarta w umowie; trzeba się upewnić, że nie stanowi to problemu dla dostawcy, w przeciwnym razie wiadomo, co należy zrobić).

Dostawcy przechowują dwa rodzaje woluminów: takie, które ciągle są wymieniane, i takie, które są magazynowane przez nieokreślony czas. Gdy wymienia się cyklicznie woluminy, są one inwentaryzowane. W przypadku archiwizowanych woluminów sprawa wygląda zupeł-

nie inaczej. Jeśli wolumin przeleżał u dostawcy dwa lata i nigdy nie był używany, jak stwierdzić, że nic się z nim nie stało? Powinno się przynajmniej raz w roku (zaleca się dwa) przeprowadzić pełną inwentaryzację takich woluminów.

Należy wysyłać oryginały i zatrzymywać kopie. Jedną z rzeczy, które powinno się regularnie testować, jest stosowana procedura kopiowania. Jeśli wysyła się woluminy poza obręb firmy, niektóre produkty archiwizujące oferują możliwość wysyłania oryginałów lub kopii. Jeżeli jest to możliwe, należy wysyłać oryginały. Gdy konieczne okaże się odtworzenie danych, należy użyć kopii. Gdy coś pójdzie nie tak, zawsze można udać się po oryginał. Ten proces weryfikuje procedurę kopiowania każdorazowo podczas odtwarzania danych. Można wyeliminować błędy obecne w procesie, zanim dojdzie do jakiegoś nieszczęśliwego zdarzenia. Pamiętam kilka sytuacji, w których napęd uszkodził wolumin lub rozlała się na nim woda. Naprawdę bardzo potrzebna nam była wtedy kopia przechowywana na zewnątrz firmy. Jest to nieodpowiednia pora, aby dowiadywać się, że procedura kopiowania nie jest dobra!

Zdalne elektroniczne składowanie danych

Elektroniczne składowanie danych staje się dość popularne. Choć może być kosztowne, jest czymś wspaniałym. Jeśli można sobie pozwolić na takie rozwiązanie, szczególnie do niego zachęcam. Chodzi o to, że kopie zapasowe są wysyłane bezpośrednio do systemu magazynowania dostawcy usługi elektronicznego składowania danych. Trzeba sobie zadać jedno pytanie: „Co się stanie, gdy *magazyn dostawcy* całkowicie spłonie?". Wszystkie dane mogą zostać utracone. Nie można pozwolić na coś takiego. Trzeba się upewnić, że firma magazynująca dane to niejedyne miejsce, w którym umieszczono zarchiwizowane dane. Dodatkowo należy wiedzieć, jak przeprowadzi się dużą operację odtwarzania. Choć niewielkie łącze sieciowe może być wystarczające do ciągłego wykonywania przyrostowych kopii zapasowych, prawdopodobnie nie sprawdzi się w przypadku odtwarzania 100 GB danych. Jeśli stanowi to problem, należy skontaktować się z dostawcą usługi elektronicznego składowania danych w sprawie opcji lokalnego odtwarzania danych.

Testowanie kopii zapasowych

Chciałbym, żeby na ten temat było tyle do powiedzenia, aby pojawiła się potrzeba stworzenia oddzielnego rozdziału. Wynika to stąd, że testowanie kopii zapasowych jest bardzo ważną kwestią. Nie mogę powiedzieć, ile słyszałem historii o osobach, które zwlekały z przetestowaniem swoich kopii zapasowych, aż stwierdziły, że muszą przeprowadzić proces pełnego odtwarzania danych. Właśnie w takich momentach okazuje się, że użyto złego urządzenia lub współczynnika bloków bądź urządzenie wygenerowało błędy wejścia-wyjścia. Kwestia testowania kopii zapasowych nie może być lekceważona, w przeciwnym razie niemiła niespodzianka jest nieunikniona.

Należy wszystko testować!

Ważne jest, aby testować każdego typu operację odtwarzania. Jeśli sprawdza się kopie zapasowe systemu plików, trzeba przeprowadzić następujące operacje:

- Odtworzenie wielu pojedynczych plików. Czy znajdzie się igłę w stogu siana?
- Odtworzenie starszej wersji pliku.

- Odtworzenie całej zawartości napędu lub systemu plików i porównanie uzyskanych wyników z oryginalnymi. Czy są jednakowego rozmiaru itd.?
- Próba odtworzenia systemu, przy założeniu, że w całości uległ awarii.
- Skorzystanie z alternatywnej kopii zapasowej, przy założeniu, że określony wolumin jest zły.
- Pobranie kilku woluminów od dostawcy usługi zewnętrznego magazynowania.
- Próba poradzenia sobie z sytuacją, w której został zniszczony serwer archiwizujący (jest to trudne zadanie!). Test ten jest wyjątkowo istotny, gdy korzysta się z darmowego lub komercyjnego narzędzia archiwizującego. Ponieważ część produktów nie radzi sobie dobrze z takimi sytuacjami, można mieć spore problemy.

Jeśli testuje się odtwarzanie bazy danych, trzeba wykonać następujące operacje:

- Odtworzenie części bazy danych, przy założeniu, że utracono tylko jeden plik danych lub napęd dyskowy (jeśli jest to możliwe).
- Odtworzenie całej bazy danych na innym serwerze (w tym przypadku można się dowiedzieć o plikach, których nie dołączono).
- Odtworzenie stanu bazy danych z określonej chwili z przeszłości (ćwiczenie to okaże się pomocne, gdy trzeba będzie odtwarzać dane po błędzie administratora bazy danych lub użytkownika).
- Skorzystanie ze starszej kopii zapasowej, przy założeniu, że nie udało się wykonanie kopii zeszłej nocy. Teoretycznie, jeśli na woluminie archiwizacyjnym zapisano wszystkie dzienniki transakcji, powinno być możliwe użycie kopii sporządzonej kilka tygodni wcześniej, a następnie odtworzenie danych do chwili obecnej za pomocą tych dzienników. Jest to kolejny mocny argument za zastosowaniem dzienników transakcji.

Często należy wykonywać operację testowania

Jak już wspomniałem, jeden dzień należy posiedzieć z kilkoma naprawdę pesymistycznymi osobami i poprosić je o wyobrażenie sobie sytuacji kwalifikujących się do przetestowania. Trzeba sprawdzić możliwość *regularnego* wykonania operacji odtwarzania w przypadku każdej sytuacji. To, co działa w tym miesiącu, może zawieść w następnym. Jedyną pewną rzeczą jest zmienność. Jedną z propozycji jest utworzenie listy procedur odtwarzania i losowe testowanie co miesiąc ich podzbioru. Zmienia się kierownictwo, sprzęt, sieci, a także wersje systemów operacyjnych i baz danych. Każda zmiana, o której wiadomo, powinna pociągać za sobą odpowiedni test.

Monitorowanie kopii zapasowych

Jeśli nie monitoruje się kopii zapasowych, pewne jest to, że nie będą się zachowywać zgodnie z oczekiwaniami. Można tu przytoczyć analogię do garnka z zupą, która się nie ugotuje, gdy nie będzie się jej doglądać. Każda kopia zapasowa powinna mieć codziennie sprawdzany dziennik. Zadanie to może zostać też zautomatyzowane. Oto kilka przykładów:

Generowanie podsumowania. Narzędzie *dump* wyświetla całą masę mniej istotnych dla mnie komunikatów, takich jak `Pass I`, `Pass II`, `% done` itp. Gdy monitoruję wykonane za pomocą tego programu kopie zapasowe setek napędów lub systemów plików, większość komunikatów

nie przedstawia zbyt dużej wartości. W rzeczywistości zależy mi, aby dowiedzieć się, co zostało zarchiwizowane, w jakim miejscu, kiedy, na jakim poziomie, a także czy został wygenerowany nadal popularny komunikat `DUMP IS DONE`. W celu uzyskania zestawienia tylko tych wierszy w pierwszej kolejności wykonuję polecenie `grep -v`, aby wykluczyć zbędne informacje i pozostawić jedynie kilka wierszy. Znacznie prościej jest przeglądać mniejszą ilość informacji. Taka metoda może też być zastosowana w przypadku innych poleceń archiwizujących w systemach: Unix, Linux i Mac OS.

W systemie Windows coś podobnego realizuje się za pomocą zaplanowanego zadania. Można spowodować, że zadanie utworzy dziennik w jakimś miejscu na dysku *C:*, a następnie dziennik zostanie przesłany do administratora za pośrednictwem programu pocztowego SMTP, takiego jak *blat* (darmowe narzędzie trybu wiersza poleceń przesyłające wiadomość pocztową za pomocą protokołu SMTP lub post do grupy dyskusyjnej Usenet przy użyciu protokołu NNTP). Program *blat* można pobrać pod adresem *http://www.blat.net*.

Wyświetlanie wszystkiego, co nietypowe. Zadanie to można zrealizować na dwa sposoby. Jeśli znane są komunikaty pojawiające się, gdy coś pójdzie nie tak, można je wychwycić narzędziem *grep*. Inna metoda polega na zastosowaniu polecenia `grep -v` usuwającego wszystkie spodziewane wiersze i sprawdzeniu tego, co pozostało. Jeżeli nie ma nic — znakomicie! Jeśli pozostały wiersze, prawdopodobnie zawierają błędy. Można zobaczyć wiersze `I/O error` i `Write error` lub coś jeszcze innego. Nie są one wskazane w przypadku kopii zapasowych.

Jeśli to samo chce się osiągnąć w przypadku zaplanowanego zadania systemu Windows, trzeba skorzystać z jakiegoś emulatora uniksowego, takiego jak Cygwin, UWIN lub GnuWin32, umożliwiającego uruchomienie w systemie Windows narzędzia *grep* i innych poleceń powłoki.

Zawsze można to jeszcze bardziej ulepszyć

Nie ma znaczenia to, jak dobre są kopie zapasowe, ponieważ zawsze mogą być lepsze. Można spędzać każdą bezsenną godzinę na dostrajaniu i ulepszaniu każdego elementu programu archiwizującego, a także wiedzieć wszystko na temat kopii zapasowych, a i tak możliwe jest dalsze doskonalenie. Moje kopie zapasowe nigdy nie będą wystarczająco dobre. Zawsze znajdzie się jakiś nowy szczegół jakiegoś innego pakietu archiwizacyjnego, większy lub bardziej „inteligentny" automatyczny zmieniacz, szybszy napęd archiwizujący lub jakiś scenariusz, który nie zostanie przeze mnie uwzględniony. Jednak trzeba sobie zdać sprawę z tego, że każda wprowadzona zmiana stwarza zagrożenie utraty danych. W książce często podkreślam, że każdorazowo, gdy pojawia się czynnik ludzki, sprawy się komplikują. Choć można być najlepszym na świecie programistą skryptów powłoki lub języka Perl, nadal będzie się popełniać błędy.

Jeśli coś nie jest barokowe, nie należy tego przyozdabiać

Barok wymagał przyozdabiania. Bez przesady! Jedno niezmienne tempo? I klawesyn? To musiało przeminąć. Na szczęście Bartolomeo Cristofori w 1709 r. wynalazł fortepian i ofiarował nam instrument, który mógł grać głośno (*fortissimo*) i cicho (*piano*). Niedługo później rozpoczął się klasycyzm i muzyka zaczęła cechować się zmiennością tempa i uczuciem.

Jeśli jednak coś *nie* jest zepsute, nie należy tego naprawiać! Choć pewnie każdy już o tym słyszał, gdy pod uwagę weźmie się ryzyko związane z każdą zmianą, w świecie archiwizowania zdanie to nabiera jeszcze większego znaczenia. Czytając tę książkę lub jakieś czasopismo bądź

rozmawiając z innymi administratorami, bez wątpienia Czytelnik stworzy listę rzeczy, które chciałby zrobić. Należy skoncentrować się na *lukach* lub scenariuszach, których stosowany plan archiwizowania i odtwarzania po prostu nie uwzględnia. Zanim pomyśli się o utworzeniu fajnego programu z menu, który chciało się napisać z myślą o operacjach odtwarzania, trzeba się martwić tym, że żaden z woluminów nie jest magazynowany poza obrębem firmy. Zanim zacznie się dekorować stoisko, trzeba zająć się wszystkim tym, co zasadnicze. Zanim pod uwagę weźmie się wprowadzenie nowej zmiany, należy zadać sobie pytanie, czy coś jest jeszcze ważniejsze i czy modyfikacja jest naprawdę konieczne i warta ryzyka.

Postępowanie zgodnie z odpowiednimi procedurami wdrożeniowymi

Nie należy wprowadzać zmiany w systemie archiwizacji, a następnie wdrażać jej jednocześnie na wszystkich komputerach. Zmianę należy testować na systemie projektowym, a jeszcze lepiej na systemie, który normalnie nie jest archiwizowany. W ten sposób nie wystawi się żadnej kopii zapasowej na niebezpieczeństwo. Inną dobrą praktyką jest testowanie modyfikacji równolegle z tym, co już zrobiono. Im większa zmiana, tym bardziej istotne jest przeprowadzenie równoległej konwersji. Jest tak szczególnie wtedy, gdy używa się nowej metody, a nie jedynie rozszerza dotychczas stosowaną. Starej metody należy używać do momentu, aż uzyska się pewność, że nowa metoda działa! Należy postępować zgodnie z planem podobnym do następującego:

- Zmiany dotyczące archiwizowania należy sprawdzać w miejscu, w którym naprawdę nie wyrządzą nikomu żadnych szkód, gdy spowodują na przykład awarię systemu!
- Operację należy testować na niewielką skalę przy użyciu jednego systemu, przeprowadzając ją tak samo jak w przypadku systemu produkcyjnego. Jeśli na przykład za pomocą programu zamierza się wykonywać zarówno zdalne, jak i lokalne kopie zapasowe, należy sprawdzić obie, ale w ograniczonej skali.
- Należy podjąć próbę symulacji każdego błędu, który może wystąpić w systemie.
  - Wyjęcie woluminu w trakcie operacji tworzenia kopii zapasowej.
  - Zablokowanie woluminu przed zapisem.
  - Ponowne uruchomienie systemu w czasie trwania jego archiwizacji.
  - Przerwanie połączenia sieciowego i odłączenie zasilania napędu dyskowego.
  - Zapoznanie się z systemem i błędami, pod kątem których się go testuje. Symulowanie każdego błędu w celu sprawdzenia odpowiedniej części systemu.
- Testy należy przeprowadzać na niewielkiej liczbie systemów (preferowane jest uwzględnienie aktualnie wybranej metody).
- Gdy zmianę wdraża się na wszystkich systemach, zdecydowanie należy to zrobić równolegle. Jedna z umożliwiających to metod polega na umieszczeniu wszystkich kopii zapasowych na jak najmniejszej liczbie woluminów, a następnie użyciu pozostałych napędów do równoległego utworzenia nowej kopii. Choć osobom odpowiedzialnym za sieć może się to nie spodobać, naprawdę jest to jedyna metoda pozwalająca przeprowadzić równoległą konwersję z prawdziwego zdarzenia. Gdy dokonywałem migracji na mój pierwszy komercyjny produkt archiwizacyjny, przez niemal rok korzystałem z tej metody.

- Ze starej metody powinno się zrezygnować tylko po przetestowaniu i dokładnym udokumentowaniu nowego systemu. Trzeba pamiętać o trzymaniu w pobliżu dokumentacji i programów niezbędnych do odtworzenia danych ze starego systemu do momentu, aż w nowym systemie znajdzie się zawartość wszystkich starych woluminów.
- Pod uwagę należy również wziąć starsze woluminy archiwizacyjne. Czy jeśli dysponuje się woluminami mającymi pięć lat, będzie można je odczytać za pomocą rozwiązania archiwizacyjnego nowego dostawcy? Czy będzie w ogóle możliwe odczytanie tych woluminów przy użyciu wersji 14. obecnie stosowanego oprogramowania, jeśli firma zaczęła tworzyć archiwalne woluminy za pomocą wersji 2. narzędzia? Czy sam nośnik będzie nadawał się jeszcze do odczytania?
- Tym sposobem przeszliśmy do zupełnie innego zagadnienia, czyli trwałości nośnika. Jeśli nawet nadal teoretycznie można odczytać woluminy po pojawieniu się 12. wersji oprogramowania, prawdopodobnie są one złe. Jeżeli dysponuje się archiwami przechowywanymi długoterminowo, z określoną częstotliwością trzeba wykonywać ich świeże kopie. Operacja polega na kopiowaniu zawartości starszych taśm na nowsze.

Niepowiązane ze sobą różności

Początkowo chciałem nadać temu podrozdziałowi tytuł „Ojej i przy okazji", lecz wydał mi się naprawdę dziwny.

Należy dbać o własną karierę

Jednym z powodów małej popularności kopii zapasowych jest to, że ludzie obawiają się zwolnienia, jeśli źle je wykonają. Ludzie mają kłopoty, gdy nie uda się operacja odtwarzania danych. Jednak poniższe sugestie pomogą uchronić się przed czymś, co można określić jako zwolnienie na skutek nieudanego odtwarzania.

Instynkt samozachowawczy — dokumentować i jeszcze raz dokumentować

Czy Czytelnik kiedykolwiek próbował wyjechać na wakacje? Jeśli jest jedyną osobą, która rozumie proces odtwarzania danych lub organizację nośników, może być pewny, że zostanie wezwany, gdy pojawi się konieczność przeprowadzenia dużej operacji odtwarzania. Kopie zapasowe są tą dziedziną administracji systemami, w przypadku której niewłaściwa dokumentacja może naprawdę przysporzyć kłopotów. Trudno wybrać się na urlop, uzyskać awans lub zrobić cokolwiek innego, co sprawi, że można będzie uwolnić się z tej magicznej sfery znanej tylko nam. Procesy archiwizowania i odtwarzania powinny być udokumentowane w stopniu umożliwiającym dowolnemu administratorowi systemu wykonanie ich krok po kroku pod naszą nieobecność. Okazuje się, że konieczność skorzystania z dokumentacji przez kogoś innego to dobra metoda sprawdzania jej przydatności.

Oczywiście przeciwieństwem dobrej dokumentacji jest zła lub nieistniejąca dokumentacja. Zła dokumentacja to najpewniejszy sposób, aby musieć szukać nowej pracy. Komu uda się wybrać na prawdziwe wakacje, podczas których nie będzie nosić pagera, sprawdzać skrzynki głosowej lub pocztowej, ten, mówiąc wprost, musi uważać, ponieważ prawo Murphy'ego również dotyczy urlopów. Jest pewne, że gdy będzie na wczasach, jego współpracownicy będą świadkami

poważnego przestoju systemów. Jeśli doprowadzą do awarii lub pożaru, ponieważ nie zostawiono wytycznych, jak wykonywać proces odtwarzania, po powrocie z urlopu będą nas szukali. Administrator, który powrócił z wypoczynku, nie będzie się cieszył sympatią, a ponadto może po prostu być zmuszony do przeglądania ofert pracy.

Dokumentacja jest też ważnym sposobem pozwalającym innym pracownikom firmy zorientować się, co robi administrator. Jeśli na przykład pominie się określonego typu pliki, napędy lub systemy plików, dobrze będzie poinformować o tym inne osoby. Pamiętam przynajmniej jedną bardzo długą rozmowę z użytkownikiem, który naprawdę nie chciał słyszeć, że nie archiwizowałem katalogu */tmp*. „Nie miałem pojęcia, że *tmp* jest skrótem od *temporary* (tymczasowe)!" — wołał.

Można biec, lecz nie można się schować

W kilku sytuacjach brak dokumentacji spowodował, że traciłem czas. Pamiętam jedne wakacje, w czasie których codziennie spędzałem przy telefonie dwie lub trzy godziny. Pamiętam przesiadywanie całymi nocami w pomieszczeniach komputerowych, ponieważ nikt nie wiedział, który przycisk wcisnąć jako następny. Jednak żadne z tych wspomnień nie jest tak silne jak wspomnienie chwili, kiedy narodziła się moja wspaniała córka Nina. Pewnie czytając to, Czytelnik powie: „Oj, jakie to słodkie". Zgadza się, Nina i moja druga córka Marissa dały mi całkiem inny powód, aby z ochotą wstawać każdego dnia. Jednak nie o tym jest ta ramka.

Szpital, w którym rodziła moja żona, był oddalony o dwie przecznice od budynku mojego biura. Wiedziałem o tym ja i moi współpracownicy (wiedział o tym każdy, kto wyjrzał przez okno!). W dniu, w którym urodziła się Nina, straciliśmy ważny system plików. Wiedziałem, że system ten znajdował się na woluminie archiwizacyjnym, a także to, że miałem *wolny dzień*. Zostawiłem w domu mój pager, który zwykle nosiłem przy boku. Nie zadzwoniłem do pracy. Wiedziałem, że proces był udokumentowany. Problem polegał na tym, że w pracy nie czytano dokumentacji. „Zadzwoń do Curtisa!". Byłem w szpitalu, w pokoju mojej żony, i rozmawiałem o naszym wspaniałym dziecku, gdy zadzwonił telefon. Współpracownicy szukali mnie i zadzwonili do szpitala! Poprosili mnie, abym przyjechał. Ponieważ wiedziałem, że system był dobrze udokumentowany, moja odpowiedź była *negatywna* (myślę, że w rzeczywistości odłożyłem słuchawkę, nie mówiąc nic ponadto)! Jest to przykład tego, co ludzie potrafią zrobić, żeby znaleźć administratora, gdy *nie udostępniono* odpowiedniej dokumentacji lub nie wyjaśniono, jak z niej korzystać.

Strategia — kopie zapasowe należy wykonywać w ramach procesu instalacyjnego

Gdy zostanie zakupiony nowy komputer, ktoś zadba o podłączenie go do zasilania. Ktoś inny będzie odpowiedzialny za podłączenie komputera do sieci, przypisanie mu adresu IP, dodanie do konfiguracji NIS i zainstalowanie odpowiednich poprawek. Wszystko to ma miejsce, ponieważ bez tego nic nie zadziała. Niestety, nikt nie zauważy tego, że serwera nie dodano do listy archiwizowanych komputerów. Oczywiście, taka sytuacja się utrzyma do momentu, aż serwer się zepsuje i trzeba będzie coś odtworzyć. Zadanie jest trudne i polega na sprawieniu, że coś „nieistotnego", takiego jak kopie zapasowe, stanie się po prostu tak naturalne jak konfigurowanie połączenia sieciowego.

Nowy komputer, który pojawi się w firmie, zwykle najlepiej nadaje się do przeprowadzenia pełnych testów procesu odtwarzania lub powielania serwera. Niewiele osób zauważy brak komputera, którego jeszcze nie posiada.

Będzie tak jedynie wtedy, gdy administrator kopii zapasowych bardzo zaangażuje się w cały proces. Być może Czytelnik jest młodą osobą, która nigdy nie uczestniczy w spotkaniach dotyczących planowania, ponieważ nie rozumie, o co w tym wszystkim chodzi. Być może Czytelnik jest *zorientowany*, lecz po prostu *nie cierpi* brać udziału w takich spotkaniach. Tak jest w moim przypadku. Jeśli ktoś nie chce uczestniczyć w każdym spotkaniu, wystarczy, że się upewni, czy będzie na nim ktoś, kto śledzi interesujące go kwestie. Być może będzie to były operator archiwizacji chodzący na spotkania i pozytywnie do niego nastawiony. Z taką osobą trzeba się spotykać, aby przypominać jej, jak ważne jest podnoszenie na spotkaniach kwestii związanych z archiwizowaniem lub informowanie o wszystkich planowanych nowych komputerach. Od czasu do czasu samemu należy się udać na spotkanie i upewnić się, że ludzie wiedzą o istnieniu operatora archiwizacji i kopii zapasowych. Można mieć nadzieję, że będą o tym pamiętali, gdy następnym razem pomyślą o podłączeniu nowego komputera bez poinformowania o tym operatora kopii zapasowych. Jednak nigdy nie należy liczyć na coś takiego. Trzeba cały czas gorliwie szukać nowych komputerów wymagających archiwizowania.

Nowe instalacje nie są jedyną rzeczą, która może mieć wpływ na archiwizację. Z punktu widzenia tworzenia kopii zapasowych istotne jest również pojawienie się nowej wersji systemu operacyjnego, nowych poprawek i nowej wersji bazy danych. Większość administratorów systemu instaluje i uruchamia nową wersję systemu operacyjnego lub bazy danych na nowym lub testowym komputerze, zanim zostanie umieszczona w środowisku produkcyjnym. Trzeba mieć pewność, że programy archiwizujące działają również na nowej platformie. Na myśl przychodzi mi kilka sytuacji, w których nowe wersje spowodowały problemy z archiwizowaniem. Oto kilka przykładów:

- Gdy pojawił się system HP-UX 10, obsługiwał pliki o wielkości przekraczającej 2 GB. Jednak w dokumentacji narzędzia *dump* napisano, że nie archiwizuje systemu plików z tak dużymi plikami.
- Baza danych Oracle zmieniała się kilka razy. Czasami zapewni zgodność wstecz, a czasami nie.
- Początkowo po pojawieniu się systemu plików EFS (*Encrypting File System*) system Windows nie mógł być archiwizowany przez niektóre systemy archiwizujące.
- Wersje systemu Mac OS nowsze od wersji Mac OS X zawsze wydają się nieznacznie odbiegać od „krzywej archiwizacji". Metody stosowane w poprzedniej wersji systemu do wykonania kopii zapasowej lub obrazu w nowej wersji nie są już przydatne. Daje to pole do popisu wielu nieoficjalnym wersjom narzędzi, które faktycznie się sprawdzają.

Im dłużej o tym myślę, tym więcej historii przychodzi mi na myśl. Jeśli Czytelnik zajmuje się archiwizowaniem od jakiegoś czasu, z pewnością też ma kilka własnych opowieści. Wystarczy powiedzieć, że aktualizacje systemu operacyjnego i aplikacji przysparzają problemów operatorowi kopii zapasowych i nadal będą. Dlatego należy testować i jeszcze raz testować!

Trzeba zdobyć środki finansowe niezbędne do archiwizowania

Ten podrozdział absolutnie nie ma nic wspólnego z kopiami zapasowymi, lecz z polityką, budżetowaniem, pieniędzmi i uzasadnianiem kosztów. Wiem, że chwilami można odnieść wrażenie, iż moim zdaniem kopii zapasowych nie szanuje się. Być może Czytelnik będzie pracować w firmie Utopia S.A., w której pracownicy przede wszystkim myślą o kopiach zapasowych. Z kolei reszta operatorów archiwizacji musi walczyć o każdy wolumin, napęd i oprogramowanie pozwalające wykonać coraz trudniejsze zadanie umieszczenia danych na woluminie archiwizacyjnym.

Uzyskanie środków finansowych wymaganych do wykonania powierzonego zadania czasami może być bardzo trudne. Jak poinformować odpowiedni dział, że niewielki zewnętrzny napęd archiwizujący dołączony do wartego milion złotych komputera, który właśnie dostarczono, po prostu do niczego się nie przyda? Ile wyrzeczeń kosztowało firmę wydanie miliona złotych na ten komputer? Jaki jeszcze wydatek ją czeka?

Trzeba być gotowym

Po pierwsze, trzeba być gotowym uzasadnić swoje wymagania. Trzeba mieć następujące informacje:

- Statystyki dotyczące przeprowadzonych operacji odtwarzania.
- Wszelkie dostępne wartości liczbowe określające możliwy koszt poniesiony przez firmę na skutek przestojów i utraconych danych, z uwzględnieniem wszystkich wymagań oddziałów biznesowych dotyczących docelowego czasu odtwarzania i docelowego punktu odtwarzania (bardziej szczegółowo pojęcia te przedstawiono w rozdziale 24.).
- Wartości pokazujące, jak dokonany zakup pomoże zredukować koszty obsługi.
- Wartości demonstrujące, jak na obecnie używany system archiwizowania negatywnie wpływa rozwój lub nowe aplikacje.
- Porównanie jednorazowego kosztu związanego z nabyciem nowszego systemu magazynowania z ciągłym kosztem robocizny wynikającej z wymiany woluminów każdej nocy (trzeba być przygotowanym na wyjaśnienie, w jaki sposób większy automatyczny zmieniacz lub wirtualna biblioteka taśmowa zmniejszą ryzyko błędu ludzkiego i jaki będzie to miało pozytywny wpływ na funkcjonowanie firmy).
- Udokumentowane argumenty na poparcie twierdzenia , że każdy kolejny gigabajt danych generuje dodatkowe, konkretne koszty.

Dobrze przygotowana prezentacja powinna uwzględniać to, co będą zapewniały tworzone kopie zapasowe, i szybkość, z jaką będzie można odtworzyć dane (choć nie należy deklarować uzyskania nierzeczywistych czasów odtwarzania, gdy nowy system może je znacząco poprawić, należy się tym pochwalić!).

- Przygotowane pismo, z którego szef będzie zadowolony, wyjaśniające bardzo rzeczowo, czego firma może się spodziewać, gdy nie zapewni żądanych środków finansowych.

Przeprowadzenie oficjalnej prezentacji

Im kosztowniejsze rozwiązanie, tym bardziej jest istotne, aby przeprowadzić oficjalną prezentację, zwłaszcza gdy pracuje się w dużej korporacji. Oficjalna techniczna prezentacja składa się z trzech

części: podsumowania dla zarządu, przeglądu przechodzącego do szczegółów i technicznej specyfikacji przeznaczonej dla *naprawdę* zainteresowanych osób.

Podsumowanie dla zarządu
: Powinno liczyć jedną stronę i bardzo ogólnie wyjaśniać, co przedstawiono w pozostałej części prezentacji. Wskazane są ogólne wartości i opisy. Nie należy zbytnio zagłębiać się w szczegóły. Choć prezentacja jest kierowana do wiceprezesa, który musi złożyć podpis, tego samego dnia może mieć do obejrzenia 20 innych prezentacji podobnych do tej. Przede wszystkim należy przedstawić bieżący problem i opisać jego rozwiązanie.

Przegląd
: W przeglądzie należy przejść do szczegółów. W tej części prezentacji należy zastosować wiele nagłówków sekcji, umożliwiających odbiorcom zapoznanie się z ich zawartością lub pominięcie, jeśli nie okażą się godne zainteresowania. Nagłówki pozwalają też osobom czytającym jedynie podsumowanie dla zarządu odszukać konkretne zagadnienia, które nie są dla nich jasne. Zarys przeglądu powinien być zgodny z podsumowaniem dla zarządu. Można dodać odwołania do innych publikacji, takich jak recenzje z czasopism poświęcone konkretnemu produktowi. Jednak nie należy ich cytować. Jeżeli recenzje są stosowne, ich kopie można zawrzeć w części technicznej. Trzeba zadbać o to, aby było widać, że wszystko jest przemyślane. Na ogólnym poziomie należy porównać wybrane rozwiązanie z innymi dostępnymi opcjami, a także wyjaśnić, dlaczego zdecydowano się na zastosowanie konkretnego produktu. Należy opisać, w jaki sposób i w jakim stopniu wybrane rozwiązanie przyczyni się do przyszłego rozwoju firmy do czasu, gdy trzeba będzie ponownie zastanowić się nad zmianą technologii. Dodatkowo należy wyjaśnić, jakie ma się plany w stosunku do dotychczasowej metodologii i jaką metodę konwersji zamierza się zastosować (na przykład tymczasowe jednoczesne korzystanie z obu rozwiązań). W prezentacji sprawdzą się również tabele. Jeśli można użyć rzeczywistych wartości, będzie to znacznie bardziej skuteczne. Trzeba jedynie je zweryfikować. Jeżeli ktoś na prezentacji będzie przekonany, że wartości są nieprawdziwe, całkowicie zdyskredytuje to raport. Warto spróbować porównać początkowy koszt wybranego rozwiązania z możliwym kosztem związanym z utratą danych.

Techniczna specyfikacja
: Trzeba iść na całość. Jeśli ktoś przetrwa dotychczas omówioną część prezentacji, będzie to oznaczać, że naprawdę się nią zainteresował lub jest prawdziwym komputerowcem, tak jak prowadzący! Jeżeli przedstawiany raport uzasadnia koszt zakupu nowego napędu dyskowego, należy poszukać tabeli porównującej względny koszt megabajta w przypadku różnych opcji. Należy dodać potwierdzone wartości i wszelkie dokumenty dołączone do proponowanego produktu. Jeżeli coś uzna się za odpowiednie, lecz będzie to zbyt długie i nudne, najlepiej będzie dodać to do technicznej specyfikacji.

Powodzenia

Następne rozdziały dogłębnie prezentują różne metody, które można wykorzystać do archiwizowania systemów, zwłaszcza za pomocą darmowych narzędzi. Omówienie większości z wcześniej przedstawionych zagadnień można też znaleźć w dokumentacji odpowiedniego producenta lub zespołu tworzącego oprogramowanie open source. Książka nie ma zastępować takiej dokumentacji. Próbuję w niej wyjaśnić rzeczy, których nie zawarto w dokumentacji,