Uniwersytet w Białymstoku

Wydział Matematyczno-Fizyczny

Instytut Matematyki

Daniel Zasim

Tworzenie kopii zapasowych w systemach komputerowych

Praca dyplomowa napisana
pod kierunkiem
dr Mariusza Żynela

Białystok 2005

Składam serdeczne podziękowania
dr. Mariuszowi Żynelowi
za pomoc i udzielenie cennych wskazówek
i sugestii w czasie pisania tej pracy.

Daniel Zasim

Spis treści

Wstęp

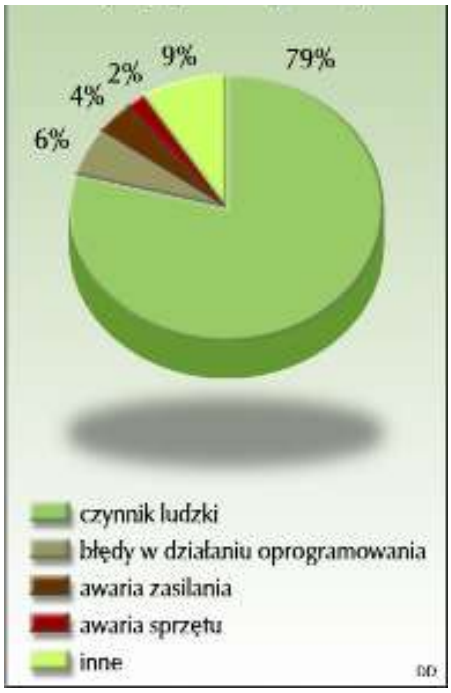
Ilość danych jakimi posługują się współczesne przedsiębiorstwa rośnie lawinowo. Rośnie też nasza zależność od dostępu do tych danych. Tradycyjnie stosowane metody ochrony polegają głównie na wykonywaniu kopii zapasowych i ich użyciu w przypadku utraty danych. Kopie zapasowe (backup), mimo to wciąż są dla użytkowników komputerów tym, czym dla kierowców pasy bezpieczeństwa i poduszki powietrzne - nieznośnymi codziennymi uciążliwościami, które na pierwszy rzut oka niczemu nie służą. Nieraz jednak decydują one o być albo nie być firmy, w szczególności w sytuacji, kiedy np. sprzęt komputerowy odmawia współpracy a do drzwi puka kontrola skarbowa. Należy zdawać sobie więc sprawę, że w przypadku uszkodzenia bazy danych (czy to w wyniku uszkodzenia dysku czy, zdarzającego się rzadko uszkodzenia samej bazy danych) odzyskanie danych może być niewykonalne, bardzo drogie, albo należy pogodzić się z tym, że aktualność bazy danych zostaje cofnięta do ostatniej wizyty administratora.

Celem pracy jest przegląd możliwości i dostępnych na rynku technologii pozwalających na tworzenie kopii zapasowych systemu serwera i danych użytkowych. W opisie uwzględniono takie parametry jak: pojemność nośnika, ilość czasu potrzebna do wykonania i odtworzenia kopii zapasowych, łatwość odtworzenia kopii, wiarygodność oraz możliwość odtworzenia na innej platformie. Dokonano przeglądu i porównania zarówno rozwiązań sprzętowych jak i software'owych.

Rozdział 1

Po co nam backup?

Utrata danych może być wynikiem przypadku lub celowego działania. Według raportu Gartner Group z marca 2005 roku 79% wszystkich problemów na świecie związanych z utratą danych jest spowodowane niedbałą lub nieumiejętną obsługą, złośliwością i wandalizmem. Za 12% przyczyn jest odpowiedzialna technika, tzn.: awarie sprzętu, błędy w oprogramowaniu, wirusy komputerowe oraz różnice w napięciu. Katastrofy i klęski żywiołowe, tj.: pożary, powodzie czy też uderzenie pioruna lub sadza, to zaledwie 9% wszystkich przyczyn. Powodów do obaw jest więc wiele.



Rysunek 1.1: Przyczyny utraty danych.

Z powyższego raportu wynika, iż najczęstszą przyczyną utraty danych są jednak błędy człowieka. O ile najnowsze technologie potrafią wyeliminować różnego rodzaju awarie sprzętu lub oprogramowania, nie zanosi się na to abyśmy szybko potrafili sobie radzić z błędami ludzi. Można próbować ograniczać ich skutki, ale o wyeliminowaniu raczej nie może być mowy. Co więcej skutki takich błędów mogą być o wiele poważniejsze, niż np. awarie sprzętu, szczególnie im większe uprawnienia posiada osoba, która np. przypadkowo usunęła jakieś dane. Paradoksalnie więc, im większe kompetencje danej osoby, tym większe zagrożenie stwarza ona dla bezpieczeństwa systemów informatycznych. Mamy tu do czynienia z zupełnie inną skalą problemu, niż w przypadku zwykłych użytkowników, którzy dzięki mniejszym uprawnieniom są autorami mniej poważnych problemów. Tak czy inaczej, podstawowym zabezpieczeniem systemów informatycznych przed utratą danych spowodowaną działaniem tzw. czynnika ludzkiego jest systematyczne wykonywanie kopii zapasowych. 85% to trochę przerażająca wartość, szczególnie w kontekście faktu, iż w przeważającej ilości firm działających w Polsce polega się wyłącznie na nieomylności administratorów wykonujących kopie zapasowe ręcznie i przechowujących je wg określonych procedur. Jest to również częsty argument producentów bibliotek taśmowych, którzy mówią, że robot (najważniejszy układ automatyki biblioteki taśmowej służący do przekładania taśm) nie zapomni o wykonaniu backupu, nie pomyli taśmy i nie pójdzie na urlop.

Drugą co do wielkości grupą przyczyn braku dostępu do informacji udostępnianej przez systemy informatyczne jest spowodowanych błędami oprogramowania lub awariami sprzętu. Oczywiście nie każda awaria sprzętu czy oprogramowania zmusza nas do odtwarzania danych z kopii zapasowej. Przypadki te ograniczają się raczej do poważniejszych uszkodzeń w strukturze bazy danych, czy uszkodzenia dysku twardego. Nie zawsze jednak w takim przypadku mamy możliwość ponownego ich wprowadzenia, co może mieć katastrofalne skutki dla dalszej działalności firmy. W takich sytuacjach bardzo przydatne są nośniki z ostatnio wykonanymi kopiami danych, które są dla nas jedyną szansą na ich odzyskanie lub ochronę przed poniesieniem określonych kosztów związanych z wprowadzaniem danych od nowa. Istnieją wprawdzie inne metody zapewniające wysoką dostępność do systemów komputerowych (takie jak technologie klastrowe, mirroring etc). Nie można jednak zapominać, że nie są to rozwiązania doskonałe i zawsze może zaistnieć sytuacja zmuszająca nas do szybkiego odtwarzania danych o wartości krytycznej dla przedsiębiorstwa. W szczególności jeśli mamy do czynienia z awarią oprogramowania.

Nie tylko uszkodzenia sprzętu, oprogramowania czy wreszcie błędy użytkownika bądź wirusy mogą stanowić przyczynę utraty naszej pracy. Pojęcie katastrof w systemach informatycznych wiąże się ściśle z angielskim terminem "disaster recovery" (odtworzenie danych po katastrofie) oraz ze wspomnianymi planami awaryjnymi. Jeszcze kilka lat temu dla wielu z nas były to obce pojęcia, jednak takie zjawiska jak powódź, która dotknęła całą południową Polskę, czy miniony 1 stycznia 2000 roku, pozwoliły nam zrozumieć, że katastrofy to

nie tylko ataki terrorystyczne, wybuchy wulkanów i trzęsienia ziemi. Disaster recovery mają zastosowanie nie tylko w takich regionach jak Kalifornia. Nie oznacza to oczywiście że powinniśmy rozważać taki rozwój wypadków, w którym Polska staje się krajem o dużej aktywności sejsmicznej. Istnieje jednak cała grupa innych zagrożeń, być może mniej spektakularnych i przez to nie dostrzeganych, ale znacznie bardziej prawdopodobnych w polskich warunkach. Niosą one ze sobą podobne skutki - zatrzymanie działalności przedsiębiorstwa na kilka lub kilkanaście dni lub tygodni. Są to m.in. pożary, zalania, powodzie, sabotaż (np. złośliwość kolegi czy "zemsta" niezadowolonego pracownika firmy). Wystarczy, że jakiś dowcipniś zadzwoni do centrali banku z informacja, że w budynku znajduje się bomba, żeby sparaliżować pracę na długie godziny, uniemożliwiając pracownikom dostęp do systemów informatycznych. Wtedy właśnie okazuje się, że problem rozwiązuje zlokalizowane zupełnie gdzie indziej centrum zapasowe, w którym znajduje się najnowsza kopia najważniejszych baz danych i w którym do dyspozycji firmowych informatyków moga zostać oddane potrzebne serwery, a najważniejsi pracownicy mogą usiaść przy biurkach, mając dostęp do podstawowej infrastruktury biurowej. Niebezpieczeństw tych może być więcej w zależności od charakteru i lokalizacji naszej działalności. Oczywiste jest, że nie ma zabezpieczenia idealnego. Niemniej jednak gwarancje bezpieczeństwa oferowanego dzięki kopiom danych powinny być naprawdę docenione w obecnych, coraz bardziej zinformatyzowanych czasach.

1.1 Kopia zapasowa czy archiwum?

W mowie potocznej słowo backup jest bardzo często mylone z archiwizacją i nieraz uznawane za synonimy. W ścisłym rozumieniu tych słów i zagadnień z nimi związanych tak nie jest. Bardzo rzadko używa się jako tłumaczenia najbardziej chyba właściwego zwrotu "kopia bezpieczeństwa". Archiwizacja danych ma charakter długoterminowy. Jej celem jest zwolnienie miejsca na dysku i przechowywanie danych "offline", czyli poza systemem komputerowym. Po wykonaniu archiwizacji zapisany nośnik z danymi wędruje do archiwum i tam pozostaje dotąd, dopóki nie będzie potrzebny. Taki sposób składowania jest najtańszy, a jednocześnie zapewnia dostęp do plików nie tylko w przypadku awarii, ale i podczas normalnej, codziennej pracy. Nie daje to jednak możliwości szybkiego odtworzenia stanu systemu w razie wystąpienia awarii, ataku hakera czy innych czynników losowych. Takie zadanie powinno być realizowane przez właściwy backup. Służy on bowiem przede wszystkim do zabezpieczania danych na stacjach roboczych i serwerach przed ewentualną awarią systemu. Ważne dla uzytkownika systemu pliki i katalogi kopiowane są w regularnych, zdefiniowanych wcześniej odstępach czasu. Zapis odbywa sie na wymiennych nośnikach wielokrotnego użytku, a dane są przechowywane znacznie krócej niż w przypadku archiwizacji. Po upływie określonego czasu na tych samych nośnikach zapisywane są nowe, aktualne kopie zapasowe. Dodatkową różnicą

w porównaniu z archiwizacją jest to, że kopie bezpieczeństwa są odtwarzane wyłącznie w sytuacjach awaryjnych.

W praktyce łączone są funkcje backupu i typowo archiwizacyjne. Wielu użytkowników uważa, że w zupełności wystarczy im archiwizacja, a dane, które znajdują się na ich komputerze nie są na tyle ważne żeby w jakikolwiek sposób zabiegać o ich bezpieczeństwo. Jest to rozumowanie błędne, gdyż nawet jeżeli komputer wykorzystywany jest tylko do celów rozrywkowych, to nie ulega wątpliwości, że podczas instalowania jakiegokolwiek programu wykonywaliśmy pracę, która zajęła nam określony czas. Nawet grając w jakąś grę, często chcemy zachować stan naszych aktualnych osiągnięć. Jest to również zapisywane w pliku. Do tego dochodzi jeszcze optymalizacja działania systemu, komputera, zainstalowanych w nim aplikacji itp. Również dysk twardy, na którym wszystkie te informacje są przechowywane jest tylko i wyłącznie mechanicznym urządzeniem i ma pełne prawo któregoś dnia odmówić posłuszeństwa. W takim przypadku, gdy nie istnieje kopia zapasowa zawartości dysku, nie ma żadnej szansy na odzyskanie pełnego stanu sprzed awarii. Nawet jeżeli uda się przywrócić działanie napędu do stanu normalnego, to zwykle jego zawartość jest bezpowrotnie tracona. Można na nowo instalować aplikacje, ale stworzonych dokumentów nic nie przywróci. Zresztą nie tylko awaria sprzętu może spowodować utratę danych. Coraz doskonalsza technologia produkcji dysków twardych, utwierdza nieraz w błędnym przekonaniu, że zanim dysk ulegnie jakiejkolwiek awarii, to zdążymy już zakupić następną generację "twardziela". Takie myślenie nie uwzględnia niestety przykrej sytuacji jak np. kradzież komputera, awaria systemu operacyjnego czy nieszczęścia losowe. Jeszcze bardziej istotne jest odzyskanie zawartości pamięci masowej komputera, gdy jest on wykorzystywany jako narzędzie pracy w firmie czy biurze. Brak danych kontrahentów, brak plików z wykonanymi zamówieniami, czas, jaki pochłonie doprowadzenie komputera do stanu używalności, a także koszt ewentualnych procedur "ratunkowych" może poważnie zakłócić pracę niejednego przedsiębiorstwa. Według badań przeprowadzonych przez UTCRIS (University of Texas Center for Research Information Systems) z firm, które utraciły dane na skutek wypadku, ok. 90% zniknęło z rynku przed upływem dwóch lat od zdarzenia, a blisko połowa (z reszty) nigdy nie osiągnęła stanu sprzed katastrofy. W przypadku, gdy przyczyną utraty danych był np. pożar, kataklizm naturalny czy kradzież, to sprzęt może być ubezpieczony. O dane jednak tworzone przez nas musimy zadbać sami.

1.2 Aspekty prawne

Istotną przyczyną, dla której warto wykonywać systematyczne backupy danych to różnego rodzaju regulacje prawne, które zmuszają nas do ochrony tajemnicy państwowej, tajemnicy służbowej lub danych osobowych. Niektóre z nich są jeszcze w trakcie przygotowywania, inne niekoniecznie są respekto-

wane. Kwestią czasu jest jednak, kiedy ustawodawca zmusi instytucje i przedsiębiorstwa do ścisłego przestrzegania obowiązujących przepisów, a co za tym idzie do zbudowania profesjonalnych systemów zabezpieczania danych. Większość instytucji gospodarczych (firm) ogranicza się do zabezpieczania danych jedynie na poziomie podstawowym. Stosuje najprostrze usługi systemowe czy też programy antywirusowe. Postępowanie takie tłumaczą stwierdzeniem, że dane przechowywane na dyskach nie mają aż takiej wartości żeby stosować bardziej skomplikowane zabezpieczenia, czy tworzyć kopie zapasowe. Podejście to jest korygowane jednak przez odpowiednie przepisy prawa, obligujące do bezpiecznego i efektywnego zabezpieczania danych.

Niemalże w każdej działalności mają zastosowania "Ustawa o ochronie danych osobowych" oraz "Ustawa o rachunkowości", które to między innymi regulują przepisy dotyczące bezpieczeństwa i przechowywania danych. "Ustawa o ochronie danych osobowych" z dnia 29 sierpnia 1997 określa zasady postępowania przy przetwarzaniu danych w systemach komputerowych oraz innych zbiorach ewidencyjnych, chroniąc prawa każdej osoby fizycznej, której dane są przetwarzane (art. 1 i 2). Jedną z głównych zasad zawartych w ustawie jest zasada bezpieczeństwa danych. Nakazuje ona odpowiednie zabezpieczenie danych, uniemożliwiające dostęp osobom nieuprawnionym [1,s.142].

Dane osobowe określane są w ustawie jako "wszelkie informacje dotyczące zidentyfikowanej lub możliwej do zidentyfikowania osoby fizycznej [2]". W przypadku przedsiębiorstw są to dane klientów, dostawców, odbiorców, i pracowników. Przetwarzanie danych to natomiast "jakiekolwiek operacje wykonywane na danych osobowych, takie jak zbieranie, utrwalanie, przechowywanie, opracowywanie, zmienianie, udostępnianie i usuwanie, a zwłaszcza te, które wykonuje się w systemach informatycznych [2]". Przetwarzaniem danych będzie więc przechowywanie na dyskach twardych i innych nośnikach, wspólne korzystanie z danych osobowych (przez aplikacje systemowe, udostępnianie plików, itp.), przesyłanie, czy jakakolwiek operacja na nich. Zabezpieczenie przetwarzanych danych, pod groźbą sankcji karnych, określone jest w Rozdziale 5. W art.36 ustawa zobowiązuje do zapewnienia odpowiednich środków zabezpieczających przede wszystkim przed "ich udostępnieniem osobom nieupoważnionym, zabraniem przez osobę nieuprawnioną, przetwarzaniem z naruszeniem ustawy, zmianą, utratą, uszkodzeniem lub zniszczeniem [2]". Przepis ten wskazuje na fakt, że zapewnienie bezpieczeństwa danych nie jest w organizacji czynnością dobrowolną. W systemie, gdzie następuje przetwarzanie danych osobowych, należy chronić zarówno dane znajdujące się w systemie jak i operacje na danych zapisanych na twardych dyskach lub innych nośnikach. Istnieją także przepisy, które pośrednio chronią przechowywane dane, jak Ustawa z dnia 27 lipca 2001 r. o ochronie baz danych. Chronione jest tu, pod groźbą kar finansowych, prawo do informacji przechowywanych w formie baz danych. Zbiory chronione są przez okres 15 lat przed nieuprawnionym korzystaniem i kopiowaniem [4]. Poprzez art.38 administrator zostaje zobowiązany do zapewnienia kontroli wprowadzania danych (kto i kiedy je wprowadził)

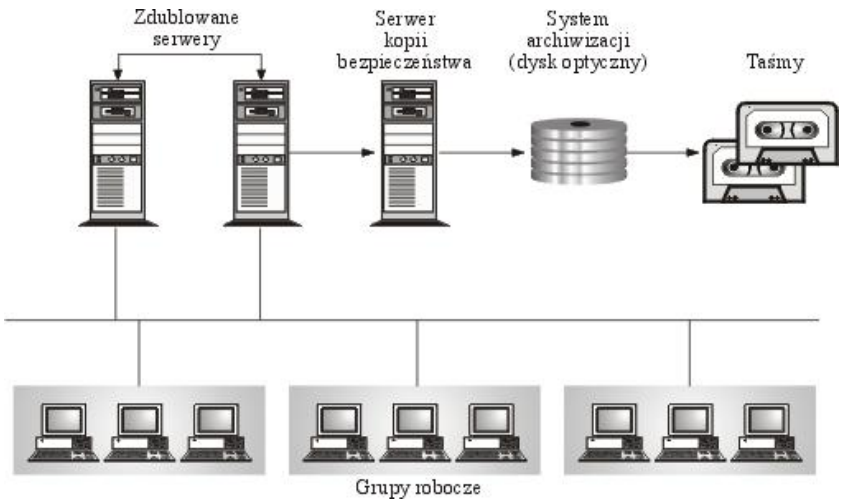
oraz ich przekazywania. Art.37 określa osoby, które mogę mieć dostęp do obsługi systemu informatycznego, w którym dane są przetwarzane (za pomocą określonych w art. 7 ust.2a współpracujących ze sobą urządzeń, programów, procedur przetwarzania informacji i narzędzi programowych). Osoby te podlegają ewidencji prowadzonej przez administratora (art. 39 ust.2).

Prawny obowiązek stosowania mechanizmów zabezpieczających dane ujęty został również w Ustawie o rachunkowości z dnia 29 września 1994. Ponieważ nowoczesna księgowość prowadzona jest z wykorzystaniem programów komputerowych, powstała konieczność ochrony danych choćby na potrzeby kontroli skarbowej. Oczywiście w interesie każdej instytucji leży jak najskuteczniejsze zabezpieczenie danych finansowych, ze względu na groźbę przechwycenia ich przez konkurencję. Art. 10 ust. 1 p. 3 i 4 ustawy (zm. Dz.U. 2000r., Nr 113 poz. 1186) zobowiązuje każdą organizację prowadzącą księgi rachunkowe przy użyciu komputera do posiadania dokumentacji zarówno używanych aplikacji, jak i stosowanych mechanizmów zabezpieczających. Artykuł ten wprowadza obowiązek prowadzenia działań podejmowanych w ramach polityki bezpieczeństwa danych. Art. 71 ust. 2 określa, że przy prowadzeniu rachunkowości przy użyciu komputera ochrona powinna polegać na używaniu "odpornych na zagrożenia nośników danych, na doborze stosownych środków ochrony zewnętrznej, na systematycznym tworzeniu rezerwowych kopii zbiorów danych zapisanych na nośnikach komputerowych (...), przez czas nie krótszy od wymaganego do przechowywania ksiąg rachunkowych, oraz na zapewnieniu ochrony (...), poprzez stosowanie odpowiednich rozwiązań programowych i organizacyjnych, chroniących przed nieupoważnionym dostępem lub zniszczeniem [3]". Przepis ten wymaga kompleksowego zabezpieczenia systemu instytucji, przetwarzającej dane księgowe. Chroniona powinna być dostępność, poufność oraz integralność informacji finansowych przechowywanych i przetwarzanych w systemie informatycznym organizacji. Art. 74 ust. 2 p. 1-8 ustanawia okres przechowywania i ochrony danych na czas od 1 roku do 5 lat w zależności od rodzaju informacji finansowych. Ponadto nakazuje trwałe zabezpieczanie przed wszelkimi niebezpieczeństwami zatwierdzonych rocznych sprawozdań finansowych. W przypadku braku odpowiednich zabezpieczeń lub danych wymaganych ustawą kierownictwo instytucji lub administrator systemu podlega karze 2 lat pozbawienia wolności lub karze grzywny w wysokości od 3.000 zł. do 150.000 zł.

Rozdział 2

Zabezpieczanie danych

Zagadnienie wykonywania backupów jest jednym z najbardziej istotnych problemów współczesnej technologii komputerowej. Jak uważa wielu specjalistów do pełnego rozwiązania problemu jest jeszcze daleko, a świadomość konieczności używania tego typu technologii jest w naszym kraju, szczególnie wśród mniejszych firm nadal niewielka. Również resellerzy mają mieszane uczucia w stosunku do bardziej zaawansowanych rozwiązań backup'owych. Uwarunkowania prawne, przenoszenie w coraz większym stopniu sprawozdawczości do komputerów, rozpowszechnienie internetu oraz lawinowo rosnącą ilość informacji gromadzonych przez firmy powodują, że rynek oprogramowania i sprzętu do backup'ów zaczął się dobrze rozwijać. Dobra baza danych o klientach to nieraz większość majątku firmy, a obowiązkowe sprawozdania (obwarowane odpowiedzialnością karną), księgi rachunkowe, dokumenty dla Urzędu Skarbowego, ZUS-u czy dane pracowników - to też dane, które firma pod groźbą różnych sankcji musi przechowywać latami. Ważna jest również kwestia instytucji operujących wyłącznie na danych takich jak urzędy, banki itd. Firmy dla własnego bezpieczeństwa przechowywują wszelkie dane na sprawdzonych i zapewniających jak największe bezpieczeństwo nośnikach. Papier bowiem zajmuje dużo więcej miejsca niż dane w postaci elektronicznej, które łatwiej jest przechowywać i łatwiej z nich korzystać. Podobnie, zabezpieczenie się przed ewentualnymi pozwami - zachowanie danych, korespondencji, czy innej dokumentacji dla celów dowodowych - skłania firmy do sięgania po specjalistyczne systemy przechowywania danych.



Rysunek 2.1: Hierarchiczny system wykonywania kopii bezpieczeństwa.

Strategię backupu trzeba jednak opracowywać biorąc pod uwagę przypadek każdej firmy z osobna. Należy przeanalizować wewnętrzną strukturę firmy, dynamikę wzrostu, środki finansowe jakimi dysponuje i szereg innych, indywidualnych czynników. Odpowiedzią na ten problem są rozwijane technologie pozwalające na zabezpieczanie danych w systemach informatycznych. Pozwala to na zabezpieczenie się przed katastrofami wyłączającymi z działania podstawowe centra informatyczne, zarówno o charakterze technicznym, jak też spowodowanym przez zjawiska naturalne, czy akty terrorystyczne. Te same rozwiązania technologiczne określane czasem ogólnym hasłem backupu danych wykorzystywane są też w nieawaryjnych zastosowaniach, takich jak na przykład eksport danych do kolejnych aplikacji, czy migracja między systemami.

2.1 Kopie bezpieczeństwa

2.1.1 Backup pełny

Backup pełny (full backup) polega na tworzeniu kopii wszystkich strategicznych danych przechowywanych w danym systemie komputerowym. Oczywistą zaletą tej metody jest łatwość odszukania dowolnych danych, ponieważ wszystkie dane znajdują się na jednym nośniku. Kolejną zaletą jest szybkość odtworzenia danych w razie awarii systemu. Wady to przede wszystkim nieefektywność wykorzystania nośników oraz długi czas wykonywania operacji. Backup pełny to rozwiązanie konieczne, jeśli w ogóle chcemy myśleć o zabezpieczaniu danych. Jednak zbyt częste wykonywanie tego działania może okazać się zajęciem zbyt pracochłonnym, a ponadto, ze wzgledu na duże ilości plików ulegających zmianie podczas codziennej pracy, nieco zbędnym. Z

drugiej strony zbyt okazjonalne archiwizowanie naraża nas na stratę aktualnych informacji. Dlatego w trakcie jednego okresu retencyjnego (czyli czasu, w jakim chcemy mieć możliwość pełnego odzyskania danych) stosuje się zazwyczaj jeszcze kilka backupów wykorzystujących dwie metody: przyrostową i różnicową.

2.1.2 Backup przyrostowy

Backup przyrostowy (incremental backup) wykorzystuje fakt, że w krótkim okresie czasu jedynie nieznaczna część plików znajdujących się w komputerze ulega jakimkolwiek modyfikacjom wprowadzanym przez użytkownika (np. poprawione dokumenty) lub przez sam system operacyjny (np. zmiana ustawień konfiguracyjnych na skutek zainstalowania nowego sprzętu). Gdy stosujemy metodę przyrostową podczas procesu backupu, zapisane zostaną tylko te dane, które były w jakikolwiek sposób zmienione od czasu ostatniej pełnej bądź przyrostowej archiwizacji. Dzięki tamu dzieląc okres retencyjny (np. miesiąc) na podokresy (np. tygodnie) otrzymujemy możliwość odzyskania danych nie tylko z ostatniego okresu (tu: miesiąca), ale także - o ile do każdej archiwizacji przyrostowej wykorzystywany jest inny nośnik - z dowolnego podokresu mieszczącego się w aktualnym okresie retencyjnym. Zaletą przyrostowej metody backupu jest to, że nośniki wykorzystywane są w takim przypadku znacznie efektywniej, ponieważ zapisywane są tylko te dane, które uległy zmianie. Ponadto czas backupu jest bardzo krótki. Wadą tej metody może być natomiast trudność w odnalezieniu właściwych danych. Do odnalezienia określonego zbioru potrzebne są wszystkie nośniki z dotychczas przeprowadzonymi backupami przyrostowymi oraz ostatni nośnik z pełną kopią zapasową. Powoduje to także, że czas ewentualnego odtworzenia systemu znacznie się wydłuża.

2.1.3 Backup różnicowy

W przypadku backupu różnicowego (differential backup) zapisywane są tylko te pliki, które zostały zmodyfikowane od czasu ostatniego pełnego (i tylko takiego) backupu. Nie ma natomiast rozróżnienia, jakie pliki zostały zmodyfikowane od czasu przeprowadzenia ostatniej archiwizacji różnicowej. Oczywiście wybór, jakie pliki były modyfikowane dokonywany jest automatycznie przez oprogramowanie archiwizacyjne. Wykorzystywany w takim przypadku jest fakt, że każdy plik czy katalog w systemie charakteryzowany jest określonymi atrybutami i parametrami. Przykładowo atrybutem wykorzystywanym przez aplikacje do backupu jest atrybut "archiwalny". Jednak równie często jako punkt odniesienia, które pliki były modyfikowane, stosowana jest data ostatniej modyfikacji pliku - ten parametr także jest zapisywany razem z plikiem. Podstawową różnicą pomiędzy backupem różnicowym a przyrostowym jest to, że w przypadku backupu różnicowego nie jest zdejmowany atrybut

"archiwalny" z archiwizowanych plików. W tym przypadku czas konieczny do odtworzenia danych jest najdłuższy gdyż zwykle potrzeba do tego kilku nośników. W razie awarii musimy bowiem dysponować ostatnim pełnym oraz wszystkimi zestawami zapasowych kopii przyrostowych.

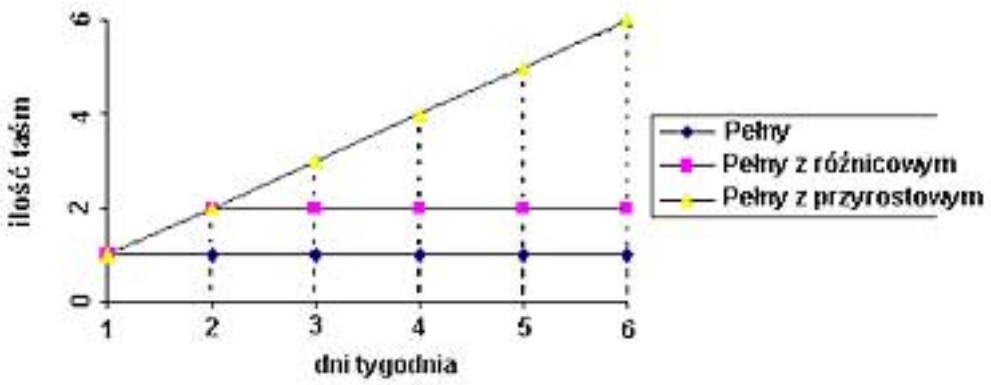
2.1.4 Popularne strategie backupu

Efektywność backupu zależy nie tylko od rodzaju wykonywanych kopii, ale też od przyjętej metody rotacji nośników. Do najpopularniejszych schematów rotacji należą: G/F/S (Grandfather/Father/Son - dziadek/ojciec/syn) oraz Wieże Hanoi.

Pierwsza z wymienionych metod jest najczęściej stosowanym schematem rotacji. Roczny "koszt" tego rozwiązania, przy założeniu pięciodniowego tygodnia pracy to 21 nośników lub ich zestawów w przypadku dużej ilości danych. Cztery z nich oznaczamy: poniedziałek, wtorek, środa i czwartek. Na nośnikach tych sporządzane będą przyrostowe lub różnicowe kopie bezpieczeństwa. Kolejne pięć taśm lub płyt CD-RW należy oznaczyć: tydzień 1, tydzień 2, ..., tydzień 5 i na nich sporządzić kopie w każdy piątek. Pozostałych dwanaście sztuk oznaczamy nazwami kolejnych miesięcy. Na koniec każdego z nich trzeba zapisać pełną kopie zapasową danych, przy czym nośniki z kopiami miesięcznymi powinny byc przechowywane poza miejscem, w którym znajduje sie serwer lub siedziba firmy.

Bardziej wymagającą jest strategia Wieże Hanoi. Wymaga ona bowiem od osoby wykonującej kopie bezpieczeństwa dużej uwagi i w przypadku pracy ręcznej najlepiej rozrysowania na kartce planu działania.

W metodzie tej nowy nośnik (lub zestaw) dodajemy cyklicznie, przy czym dla każdego kolejnego nośnika długość cyklu jest dwa razy dłuższa niż dla poprzedniego. Rozpoczynamy rotacje od nośnika A i używamy go cyklicznie co drugi dzień. Na drugi nośnik (B) zapisujemy kopie w pierwszy wolny dzień, w którym nie jest wykorzystywany A. B jest używany cyklicznie co czwarty dzień. Nośnik oznaczony literą C z kolei jest dołaczany do schematu rotacji w pierwszy wolny dzień, w którym nie jest wykorzystywany ani A, ani B. i jest używany cyklicznie co osiem dni. W ten sposób możemy dołączać kolejne taśmy lub płyty oznaczone D, E itd.



Rysunek 2.2: Zależność ilości taśm od długości okresu backup'u.

Najbardziej aktualne kopie danych znajdują się na najczęściej nagrywanych nośnikach. Im dłuższy cykl zapisu danych, tym starsza kopia danych jest na nim zapisana. Podobnie jak w przypadku poprzedniej metody, nośniki zapisywane najrzadziej należy przechowywać poza firmą.

2.2 Replikacje danych

Tradycyjnie stosowane metody ochrony polegają głównie na wykonywaniu kopii zapasowych i ich użyciu w przypadku utraty danych. Według danych statystycznych około 90% amerykańskich przedsiębiorstw wykorzystuje kombinację kopiowania danych na taśmy i dedykowanego centrum zapasowego, do którego przesyłane są kopie zapasowe. Pozwala to na ograniczenie zakresu danych traconych w przypadku katastroficznej awarii z czasu ok. 24-48 godzin, a okresu powrotu do sprawnego funkcjonowania do poniżej tygodnia. W wielu przypadkach takie rozwiązanie jest wystarczające, jednak ciągle rośnie liczba dziedzin, w których nasza zależność od informacji jest tak duża, iż utrata danych, czy zdolności ich użytkowania nawet na kilka godzin, jest nieakceptowalna. Dlatego też bardzo wygodnym stało się kopiowanie danych między systemami informatycznymi znajdującymi się w rozproszonych lokalizacjach. Dzięki takim rozwiązaniom technologicznym (replikacjom) uległ znacznemu skróceniu czas odtworzenia utraconych danych. Techniki zdalnej replikacji danych są bowiem efektywniejszym odpowiednikiem procesu przewożenia taśm backupowych z jednej lokalizacji do drugiej i odtworzenia tam zapisanych informacji. Rozwój technologii sieciowych, zarówno typowych sieci IP (LAN, MAN i WAN), jak i specjalizowanych sieci takich jak Fibre Channel (FC), czy ESCON pozwala na przesyłanie dużych ilości danych między odległymi centrami komputerowymi. Dzięki temu możliwe jest zrealizowanie za pomocą tych sieci bezpośredniego transferu danych między systemami dyskowymi. Korzyści z takich możliwości są oczywiste: użytkownik zabezpieczony jest

przed utratą danych lub ich dostępności związanych z poważnymi awariami centrum obliczeniowego. Centrum zapasowe ma niemal natychmiastową możliwość przejęcia zadań centrum podstawowego. Stąd już od wielu lat dostępne są na rynku rozwiązania zapewniające taką funkcjonalność. Najczęściej oferowane są one przez producentów zaawansowanych rozwiązań pamięci masowej. Przykładami są PPRC (Peer to Peer Remote Copy) firmy IBM, SRDF (Symmetrix Remote Data Facility) firmy EMC, HORC/HRC firmy Hitachi Data Systems.

W ostatnich latach także firmy specjalizujące się dotąd w rozwiązaniach backupowych i zarządzaniem danymi, jak Legato, czy Veritas wprowadziły na rynek rozwiązania pozwalające na zdalną replikację danych. Jednak wdrożenie systemu zdalnej replikacji nie jest sprawą prostą i wymaga przeanalizowania konkretnych potrzeb i dopasowania rozwiązania do założonych celów i możliwości technicznych i finansowych. Poniżej przedstawione zostały charakterystyczne cechy różnych rozwiązań wraz z ich wadami i zaletami.

2.2.1 Replikacja synchroniczna

Najbardziej zaawansowaną technologicznie, a zarazem najbardziej wymagającą jest realizowana w czasie rzeczywistym replikacja synchroniczna. W tym przypadku system zapisuje blok danych na dysk lokalny i wysyła go do odległej lokalizacji. Po zapisaniu bloku danych na zdalny dysk do lokalizacji podstawowej jest wysyłane potwierdzenie zapisu. Operacja zostaje uznana za wykonaną, kiedy potwierdzenie dociera do serwera z lokalizacji podstawowej. Dopiero wtedy serwer podstawowy może przeprowadzić kolejna operację. Wszystkie zapisy są więc wykonywane synchronicznie w lokalizacji podstawowej i zapasowej. Metoda ta gwarantuje, że w obu lokalizacjach dane są w takim samym stanie.

Zalety replikacji synchronicznej wynikają wprost z faktu, iż oba systemy pamięci masowej "jednocześnie" realizują wszystkie operacje zapisu, czyli zawierają identyczne kopie danych. A zatem centrum zapasowe, dysponując uaktualnianą w czasie rzeczywistym kopią danych, może natychmiast przejąć rolę aktywną, nie tracąc czasu, ani danych.

Replikacja synchroniczna wymaga jednak ogromnych przepustowości łącza - co oznacza jego wysokie koszty. Co więcej liczy się nie tylko przepustowość, ale i opóźnienia wnoszone przez łącze. Ponieważ każda operacja zapisu jest uznana za zrealizowaną dopiero po potwierdzeniu jej przez odległe centrum zapasowe, opóźnienia łącza dramatycznie wpływają na wydajność zapisu. Prowadzi to do efektywnego ograniczenia separacji między centrami dla replikacji synchronicznej do ok. 40 kilometrów.

Replikacja synchroniczna niesie ze sobą także niebezpieczeństwa wynikające wprost z faktu iż oba systemy zawierają "identyczną" informację. W szczególności sytuacje takie jak błąd operatora lub aplikacji natychmiast propagują się na system zapasowy. W takich sytuacjach zabezpieczenie okazuje

się być pozorne. Dla ochrony przed takimi zdarzeniami często stosuje się dodatkowe mechanizmy zabezpieczające - od strony serwera lub systemu pamięci masowej.

Kolejną wada replikacji synchronicznej jest możliwość wystąpienia tak zwanych "rolling disaster": sytuacji w której awaria głównego systemu dyskowego może spowodować utratę spójności danych (a więc i ich dostępności) w centrum zapasowym. Może ono powstać, gdy awaria systemu ujawnia się stopniowo. Zerwanie replikacji poszczególnych dysków logicznych może nastąpić w innych momentach. W efekcie zdalne dyski logiczne moga nie być ze soba zsynchronizowane. Dla wielu baz danych oznacza to utratę integralności systemu i konieczność czasochłonnego odtwarzania danych z kopii zapasowej na taśmach.

2.2.2 Replikacja asynchroniczna

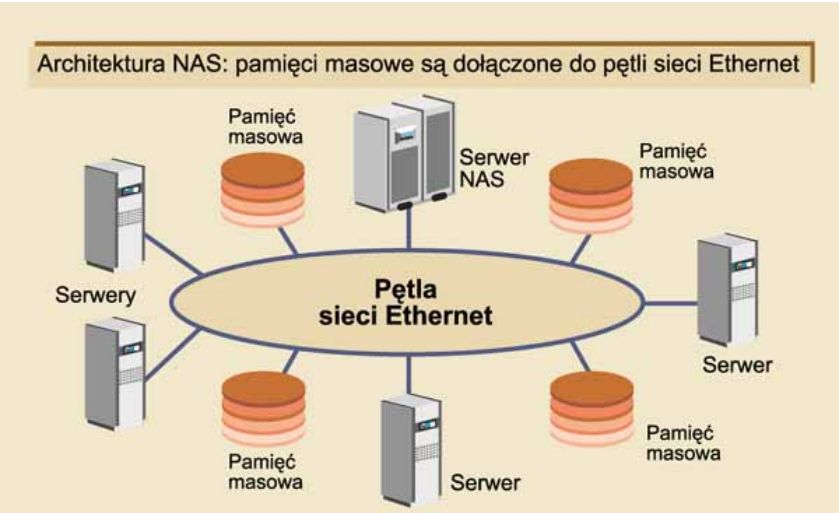
Pomimo opisanych wcześniej wad, dobrze wdrożone rozwiązania replikacji synchronicznej zapewniają najkrótszy czas przywrócenia funkcjonalności systemu i minimalizację utraty danych. Jednak wysokie koszty rozwiązania motywowały poszukiwania rozwiązań mogących zapewnić wysoki stopień bezpieczeństwa przy znacząco niższych wymaganiach i kosztach.

Rozwiązaniem takim jest replikacja asynchroniczna, gdzie centrum główne buforuje operacje zapisu i "we właściwym czasie" przesyła je do centrum zapasowego. W tym przypadku serwer nie oczekuje na potwierdzenie zapisu na zdalnym dysku. Operacja zostaje uznana za przeprowadzoną natychmiast po wykonaniu zapisu na dysku lokalnym w lokalizacji podstawowej. Oznacza to, że replikacja w trybie asynchronicznym minimalnie opóźnia wydajność całego systemu lub w ogóle nie ma na nią wpływu. Brak potwiedzeń zapisu w zdalnej lokalizacji może jednak skutkować różnicą stanu systemów w obu lokalizacjach (część danych jest jeszcze "w drodze" do odległej lokalizacji). W przypadku awarii lokalizacji podstawowej może dojść do utraty ostatnio wykonywanych transakcji. Jeszcze groźniejsza w skutkach może być systuacja, gdy dane są replikowane na bardzo duże odległości (setki kilometrów), a do ich transakcji wykorzystywane są łącza IP o stosunkowo niewielkiej przepustowości, nieraz dodatkowo obciążone innymi zadaniami. W takich sytuacjach może dojść do utraty niektórych przesyłanych pakietów lub zapisania danych w niewłaściwej kolejności. Rezultatem tego jest oczywiście utrata spójności danych, co może być powodem ich uszkodzenia i braku możliwości odczytu danych. Co warto jednak zauważyć, wymagania stawiane przez technologię asynchroniczną są znacząco mniejsze, co prowadzi do radykalnie niższych kosztów bieżących. A luka wielkości kilku, czy kilkunastu minut jest bardzo często w zupełności do zaakceptowania dla użytkowników. Jednocześnie ze względu na brak drastycznych wymagań co do opóźnień wnoszonych przez połączenie pozwala w zasadzie na nieograniczoną separację między centrami. Dodatkowo opóźnienie w transmisji pozwala na zastosowanie skutecznych zabezpieczeń przeciwko

"rolling disasters" i propagacji błędów operatora, czy aplikacji, co jest dodatkowym walorem rozwiązania.

2.2.3 Replikacja w świecie NAS i SAN

Oba tryby replikacji występują również w pozostałych sposobach replikacji, nieobciążających serwerów produkcyjnych. Są to metody wykorzystujące możliwości zdalnego kopiowania danych ze specjalnych urządzeń włączonych do sieci SAN (Storage Area Network) SAN jest terminem zasadniczo bardzo łatwym do opisania. Najprościej rzecz ujmując jest to wydzielona, szybka sieć komputerowa łącząca urządzenia pamięci masowych (storage devices) z heterogenicznymi serwerami, na zasadzie współpracy "każdy z każdym". W efekcie uzyskujemy całkowite wydzielenie pamięci masowych, przechowywania danych z lokalnej sieci komputerowej LAN i przeniesienie ich do osobnej, szybkiej sieci SAN. Istnieje oczywiście wiele innych zbliżonych definicji Storage Area Network. Najbardziej interesująca jest ta, która określa różnicę pomiędzy Storage Area Network (SAN), a Network Attached Storage (NAS). Wydziela ona trzy zasadnicze, centralne elementy całej infrastruktury przetwarzania danych: pamięci masowe, serwery sieci LAN i system plików (file system) łączący urządzenia pamięci z serwerami. Definicja mówi, że mamy do czynienia ze Storage Area Network, jeśli pomiędzy pamięci masowe a system plików wstawimy sieć komputerową. Jeśli natomiast sieć znajdzie się pomiędzy systemem plików a serwerami, otrzymujemy Network Attached Storage (np. Network Appliance).

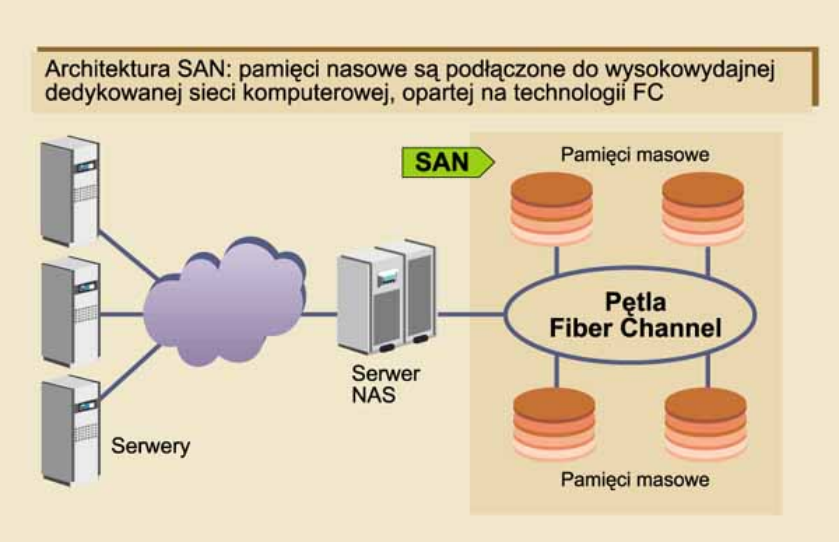


Rysunek 2.3: Architektura sieci NAS.

Zaletą takich metod jest zdolność jednoczesnego replikowania danych z wielu systemów operacyjnych, baz danych i aplikacji funkcjonujących na ser-

werach podłączonych do sieci SAN lub macierzy dyskowych. Jest to możliwe dzięki wykonywaniu kopiowania na bardzo niskim poziomie, w warstwie sprzętowej. Oprogramowanie kopiujące (np. wbudowane w mikrokod macierzy) działa na poziomie dysków logicznych macierzy bez "wiedzy" o tym, jaki system operacyjny korzysta z tego dysku i jakiego typu dane są na nim umieszczone.

Replikacja z sieci SAN oznacza włączenie do tej sieci urządzenia, za pomocą którego wszystkie zapisy przesyłane są na dyski. Urządzenie takie może przesyłać każdy zapisywany blok do analogicznego zdalnego urządzenia przez dowolną sieć IP. Jest to przeniesienie funkcji oferowanej przez niektóre macierze dyskowe na całą sieć SAN i udostępnienie jej w takiej samej postaci wszystkim podłączonym do sieci urządzeniom dyskowym, niezależnie od stopnia ich zaawansowania technologicznego. Ponieważ w sieci pojawia się jednak dodatkowe urządzenie, pośredniczące w komunikacji serwerów z zasobami dyskowymi, należy dokładnie przeanalizować wpływ takiej architektury na wydajność aplikacji. Może się okazać, że nie jest to rozwiązanie dla każdego i nie do dowolnych zastosowań. W stosunkowo mało wymagających środowiskach może to być jednak rozwiązanie idealne, ekonomiczne i skutecznie rozszerzające funkcje nieskomplikowanych i tanich urządzeń o zaawansowane działania replikacyjne.



Rysunek 2.4: Architektura sieci SAN.

Wadą kopiowania sprzętowego na poziomie sieci SAN jest niski stopień integracji warstwy replikującej z obsługiwanymi aplikacjami i bazami danych. Mimo to dostawcy oprogramowania bardzo poważnie traktuja rozwiązania tej klasy z powodu ich dużej popularnościna rynku. Przykładem jest porządkowanie zapisów danych w replikacji asynchronicznej. Przed wysłaniem do zdalnej

lokalizacji wszystkie zapisy macierzy podstawowej otrzymują kolejne numery, według których są w niej szeregowane. Kolejka taka jest zapisywana w zdalnej macierzy tylko we właściwych sekwencjach. W przypadku jakiegokolwiek zaburzenia sekwencji zapis nie nastepuje, a operacja musi byc powtórzona. Dzięki temu można sobie poradzić z problemem zaginionych pakietów, niebezpieczeństwem utraty spójności danych w sytuacji zaburzenia kolejności wykonywania zapisów na dyskach oraz sytuacjach rolling disaster. Pozostaje jedynie problem, którego w trybie asynchronicznym właściwie nie sposób rozwiązać. Jest nim niebezpieczeństwo utraty ostatnich transakcji, które zostały wykonane i potwierdzone w ośrodku podstawowym, a nie zdążyły osiągnąć ośrodka zdalnego z powodu awarii.

2.3 Zdalne kopie zapasowe

Wiele firm zajmujących się backup'em danych oferuje wszystkim posiadaczom komputerów z dostępem do Internetu oryginalną usługę - wykonywanie kopii zapasowych danych znajdujących się na dysku komputera poprzez Sieć. Taśmy z kopiami zapasowymi mogą ulec uszkodzeniu lub zginąć, a do wykonywania kopii na taśmach potrzebne są specjalne urządzenia, zaś aby zrobić backup on-line, wystarczy tylko kilka kliknięć myszy. Takie rozwiązanie nie zaleca się do masowego archiwizowania dużej ilości danych - np. całości zainstalowanego w komputerze oprogramowania czy systemu operacyjnego - a jedynie "krytycznych" plików, wykorzystywanych do pracy przez posiadacza komputera. Objętość jednorazowo archiwizowanych danych nie powinna być większa niż kilkanaście do kilkudziesięciu megabajtów (limit programu wynosi 100 MB), chociażby ze względu na szybkość ich transmisji przez Internet (dla porównania, przeciętna taśma DAT mieści 2 GB danych i zapisywana jest z prędkością kilkuset kilobajtów/s).

Backup on-line może być interesującym rozwiązaniem dla indywidualnych posiadaczy komputerów, dla których nieopłacalny jest zakup streamera do wykonywania kopii na taśmach, a równocześnie chcieliby skorzystać z dobrodziejstwa zabezpieczenia swoich ważnych plików. Aby skorzystać z tych usług, niezbędny jest dostęp do Internetu oraz specjalne oprogramowanie, które można "ściągnąć" z internetowej strony firmy oferującej zabezpieczenie danych przez sieć. Przy pierwszym uruchomieniu programu trzeba zazwyczaj zarejestrować się podając swoje dane. Po dołączeniu w ten sposób do grona zwolenników tego typu rozwiązania przez około pierwszych 30 dni (w zależności od firmy) z usług backupu możemy korzystać za darmo, później pobierana jest niewygórowana opłata wynosząca od 80 do 125 USD rocznie oraz określane jest nazwa konta i hasło, których będziemy dalej używać. Następnie wystarczy w okienku wzorowanym na Eksploratorze Windows zaznaczyć pliki, które chcemy archiwizować, i ustalić, w jakie dni tygodnia i w jakich godzinach program ma wykonywać backup. Program automatycznie uaktywni się o wskazanej porze

(oczywiście komputer musi być w tym czasie włączony, natomiast może znajdować się w stanie "uśpienia", z którego zostanie samoczynnie wyprowadzony przerwaniem zegarowym), zidentyfikuje pliki, które zostały zmodyfikowane od ostatniego kopiowania (wykorzystywany jest tu nie - jak w klasycznych programach do backupu - atrybut Archive w katalogu dysku bądź data i czas modyfikacji pliku, lecz suma kontrolna jego zawartości, co pozwala na wiarygodne stwierdzenie faktu rzeczywistego zmodyfikowania pliku), połączy w razie potrzeby z Internetem (jeżeli korzystamy z połączenia modemowego) i rozpocznie kopiowanie plików. Można również w dowolnym momencie zainicjować archiwizację ręcznie - możliwość ta jest przydatna np. dla podróżujących posiadaczy komputerów przenośnych.

W razie uszkodzenia lub utraty któregoś ze zarchiwizowanych plików wystarczy tylko wybrać jego nazwę z przechowywanej przez program listy i wybrać opcję "Restore", a plik zostanie natychmiast odtworzony. System "pamięta" nie tylko ostatnią wersję każdego pliku, ale wszystkie poprzednie, które były archiwizowane. Pozwala to na bezproblemowe odtworzenie wersji pliku np. sprzed kilku dni, czy tygodnia, w razie omyłkowego zastąpienia ich inną zawartością. W przypadku utraty większej ilości danych, jeżeli odtwarzanie ich poprzez sieć byłoby zbyt czasochłonne, można zamówić w firmie (za dodatkową odpłatnością) CD-ROM zawierający nasze zarchiwizowane dane, który dostarczany jest przesyłką kurierską następnego dnia.

Archiwizowane dane magazynowane są w chronionym centrum komputerowym z całodobową obsługą techniczną, wyposażonym w serwery, dysponujące nośnikami danych (dyskami i taśmami) o łącznej pojemności ok. 1000 TB (terabajtów). Dane te kopiowane są na dodatkowe taśmy, przechowywane w bezpiecznym miejscu poza centrum. Rzecz jasna, poza fizycznym bezpieczeństwem niezbędne jest również zachowanie poufności danych użytkownika. Wszystkie dane archiwizowane przez sieć szyfrowane są przed opuszczeniem komputera użytkownika i przesyłane oraz przechowywane są cały czas w postaci zaszyfrowanej. Klucz do szyfru zna tylko ich właściciel.

W podstawowej wersji - jak już wspomniano powyżej - korzystanie z usług "backup on-line" wiąże się z pewnnymi kosztami rocznymi. Wiele firm oferuje również całkowicie bezpłatna wersja np. Backup Lite,z limitem 10 MB archiwizowanych danych. Wiele firm dla zachęcenia klienta przewiduje również szereg zniżek i promocji: np. wśród klientów, którzy namówią nowe osoby do skorzystania z usług backupu on-line rozlosowywana jest dożywotnia, darmowa licencja na korzystanie z usługi.

2.4 Nośniki

Popularność nagrywarek CD-R/RW stwarza możliwość wykorzystania płyt CD-R i CD-RW w charakterze nośników do backupu. Często pomagają w tym producenci sprzętu i oprogramowania do nagrywania, oferując w zestawach ra-

zem z urządzeniami odpowiednie aplikacje czy oprogramowanie umożliwiające np. nie tylko wykonanie pełnego backupu wszystkich plików znajdujących się na dysku twardym, ale też utworzenie uruchamialnej płyty startowej, dzięki czemu całość takiego zestawu nośników spełnia założenie możliwości stosunkowo szybkiego odzyskania sprawności systemu (w przypadku pojedynczego komputera) po jakiejś awarii. Zaletą takiego rozwiązania jest niska cena zarówno nagrywarki, jak i nośników, popularność takiej metody przenoszenia i dystrybucji danych i wiążąca się z tym znajomość obsługi urządzeń. Wreszcie nagrywarki to nie sprzęt specjalizowany do backupu, a jedynie medium umożliwiające przeniesienie na wymienny nośnik 700 MB danych. Wadą jest mała pojemność tego nośnika. Jednym z kluczowych elementów strategii backupu jest konieczność wykonania tzw. backupu pełnego, czyli wykonania kopii zapasowej całości systemu komputerowego. Przy dzisiejszych pojemnościach dysków twardych (rzędu od kilkunastu do kilkuset gigabajtów) zadanie wykonania backupu na płytach CD-R staje się kłopotliwe i nieopłacalne. Ponadto nośniki CD-R są do jednorazowego użytku, co przy założeniu, że realizujemy podstawową zasadę prawidłowo przeprowadzonego backupu, powtarzalność operacji, stwarza nie lada obciążenie dla naszego portfela. Jeśli będziemy chcieli zaoszczędzić, wykonując pełny backup tylko np. co rok, to w ogóle nie warto zabierać się za tego typu zabezpieczenia. Co bowiem zrobimy, gdy system ulegnie awarii w 364 dniu od czasu wykonania ostatniego pełnego backupu? Wprawdzie istnieją na rynku płyty wielokrotnego zapisu ale charakteryzuja się one znacznie większą awaryjność. Technologia zmiany fazy (phase change), w oparciu o którą przeprowadzany jest zapis na płytach CD-RW jednoznacznie określa niższą, niż w przypadku płyt CD-R, trwałość nośników. Choć teoretycznie archiwizacja na płytach CD-RW jest możliwa, to nie rozwiązuje problemu stosunkowo niskiej pojemności nośnika. Pozostaje jeszcze pamiętać o większej cenie nośników CD-RW i wolniejszym ich zapisie przez nagrywarki. O ile w przypadku nagrywania pojedynczych płyt aspekt szybkości zapisu nie ma tak dużego znaczenia, o tyle w przypadku wykonywania pełnej kopii zapasowej różnice pomiędzy np. 52X dla CD-R a 12X dla CD-RW oznacza odpowiednio więcej czasu spędzonego przy komputerze czy serwerze. Z drugiej strony płyty CD-R, właśnie ze względu na swoją stosunkowo wysoką trwałość, doskonale nadają się do archiwizacji danych. Istnieje też rozwiązanie pośrednie wykorzystujące rejestrację danych na nośnikach optycznych, choć w tym wypadku wymaga to odpowiedniej konfiguracji komputera. Przykładowo chcąc nagrać kopię zapasową na płytce CD-R zapewniającą szybkie odtworzenie systemu w razie awarii, można wspomóc wykorzystanie nagrywarki oprogramowaniem wykonującym skompresowany plik kopii zapasowej. Takim programem może być np. Norton Ghost (Symantec Ghost).

Użytkownicy, którzy myślą o nagrywaniu większych ilości danych (rzędu kilku gigabajtów), powinni raczej zaopatrzyć się w nagrywarkę DVD-RW. Przy dwukrotnie wyższej cenie nagrywarka DVD-RW zapewnia niespełna trzykrotnie wiekszą średnia prędkość zapisu. Co wiecej, backup danych na nośni-

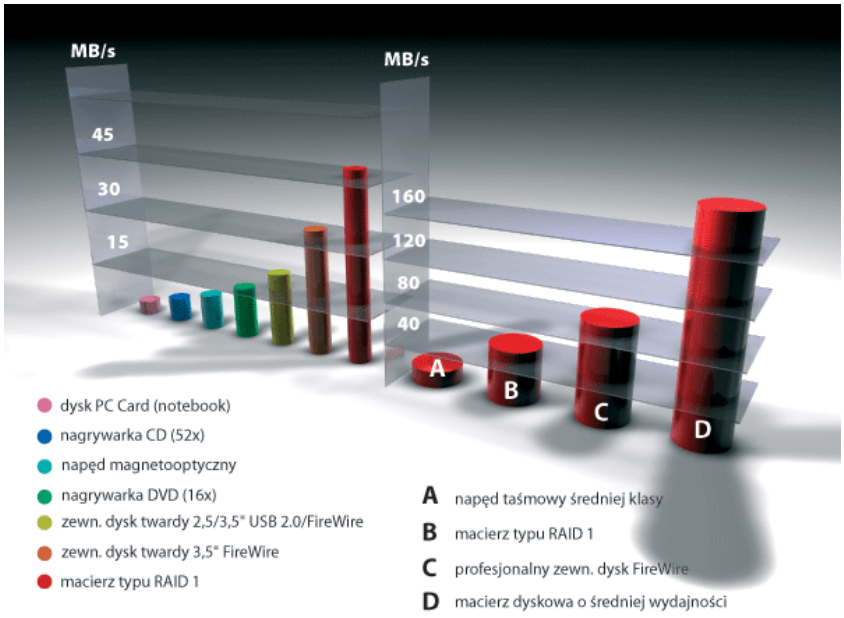
kach DVD-R jest ekonomicznie bardziej opłacalna niżna płytach CD-R. Krążki DVD są tylko o kilkadziesiąt procent droższe, a mieszczą prawie siedem razy więcej danych.

Kolejnym rozwiązaniem, które możemy zastosować zarówno do archiwizacji, jak i do backupu danych są napędy oraz nośniki wykorzystujące technologię magnetooptyczną (MO). Nośniki magnetooptyczne są praktycznie najtrwalsze spośród ogólnie dostępnych rozwiązań. Dyskietki magnetooptyczne zapisywane i odczytywane są bezkontaktowo, czyli głowice zapisujące/odczytujące nie mają bezpośredniej styczności z powierzchnią nośnika, dzięki temu nie ma możliwości, by wystąpiło uszkodzenie powierzchni przez głowice. Ponieważ do zapisu danych wymagany jest udział zarówno pola magnetycznego, jak i energii cieplnej (lasera), bez obaw można umieszczać nośniki MO w silnych polach magnetycznych. Brak jednego z dwóch czynników koniecznych do zapisu danych powoduje, że jakakolwiek zmiana czy skasowanie danych nie są możliwe. Technologia MO to nie tylko trwałość nośników, ale też wytrzymałość mechaniczna napędów. Jako ciekawostkę warto podać, że właśnie ze względu na trwałość technologia MO stosowana jest w amerykańskich myśliwcach bojowych F-16 i helikopterach szturmowych Appache w celu rejestracji aktualnej sytuacji, także w warunkach bojowych. Choć koszt napędu magnetooptycznego i zestawu nośników jest niewątpliwie większy niż w przypadku nagrywarki CD-R/RW i zestawu płyt, to wykorzystując napęd MO omijamy ograniczenie pojemności. Na rynku dostępne są nośniki o pojemnościach do 5,2 GB, a to już w zupełności powinno wystarczyć na wykonanie nawet pełnego backupu pojedynczej stacji roboczej. Kolejna zaleta MO to fakt, że napęd instalowany jest w systemie podobnie jak twardy dysk. Do zapisu nośników nie jest potrzebne żadne dodatkowe oprogramowanie (jak to jest w przypadku nagrywarek).

Najdroższymi ale i najbardziej wydajnymi urządzeniami są z kolei streamery dedykowane. Jednym z najtańszych rozwiązań tego typu jest napęd taśmowy OnStream DI 30, dostępny za ok. 1500 zł. W tym urządzeniu wykorzystywana jest technologia ADR (Advanced Digital Recording - zapis liniowy na taśmie o szerokości 1/4 cala 6,35 mm). Wspomniany model bazuje na interfejsie ATA, co oznacza, że można go bez problemu podłączyć do każdego komputera PC. Pojemność obsługiwanych przez ten model taśm wynosi 15 GB (30 GB z kompresją), co powinno wystarczyć do przeprowadzenia pełnego backupu nie tylko pojedynczej stacji roboczej, ale np. serwera plików w małej sieci. Ciekawostką jest fakt, że dołączane do tego napędu oprogramowanie (pracujące pod kontrolą Windows) umożliwia dostęp do taśmy przez literę dysku, co jest w przypadku streamerów dość rzadko spotykane. Zaletą napędów taśmowych jest, przede wszystkim, olbrzymia pojemność nośników, najmniejsze aktualnie wykorzystywane nośniki mają ok. 4,8 GB pojemności, pojemność największych liczona jest w setkach GB. Zaś wykorzystywane w dużych korporacjach, bankowości itp. biblioteki taśmowe oferują przestrzeń liczoną w TB (terabajtach). Drugą zaletą streamerów jest niski koszt nośnika w stosunku do oferowanej przez niego pojemności. Przykładowo cena nośnika

ADR 15/30 GB (do wspomnianego modelu streamera OnStream) wynosi ok. 250 zł (brutto). W przypadku najpopularniejszej (głównie ze względu na cenę) technologii zapisu taśmowego, DAT/DDS, pojedynczy nośnik np. DDS-4 (pojemność 20/40 GB) BASF/Emtec kosztuje 99 zł. W tym jednak przypadku więcej musimy zapłacić za sam napęd taśmowy, najtańsze napędy taśmowe DAT/DDS (obsługujące nośniki DDS-2 BASF/Emtec o pojemności 4/8 GB kosztują ok. 26 zł/szt.) to wydatek rzędu 2,5-4 tys. zł. Wybór odpowiadającego naszym potrzebom rozwiązania zależy w szczególności od tego, jakie dane zamierzamy chronić (i co za tym idzie, jakich pojemności potrzebujemy), na ile są one istotne dla prowadzonych przez nas prac, jak szybko mamy odzyskać pełną sprawność systemu po awarii i dopiero na końcu zadajemy sobie pytanie: ile za to możemy zapłacić? Wiadomo, że w warunkach domowych priorytetem jest przede wszystkim cena. Czas odzyskania danych ma tu drugorzędne znaczenie.

Inną bardzo wygodną metoda archiwizacji danych jest przenośny dysk w obudowie USB 2.0 lub FireWire. Doskonale sprawdza się on zarówno w domu, jak i w małej firmie, w której znajduje się zaledwie kilka komputerów. Zaletami mobilnego napędu są bardzo łatwa instalacja i duża predkość transmisji danych. Podpięcie "twardziela" do komputera trwa chwilę, a w dodatku nie musimy resetować maszyny. Po zgraniu plików napęd można odłączyć od peceta i schować w bezpiecznym miejscu.



Rysunek 2.5: Wykres prędkości zapisu danych.

Jedynym warunkiem do efektywnego wykorzystania takiego urządzenia jest obecność w PC przynajmniej jednego ze złączy: USB 2.0 (480 Mb/s)

lub FireWire (IEEE-1394a - 400Mb/s, IEEE-1394b - 800Mb/s). Wtedy pliki będą zapisywane na przenośnym dysku z prędkością co najmniej 20-25 MB/s. Port USB 1.X ma zbyt małą szybkość transmisji danych (zaledwie 12 Mb/s), aby myśleć o wydajnej archiwizacji dużych plików, gdyż przegranie jednego gigabajta będzie trwało około pół godziny.

Dodatkową zaletą zewnetrznych dysków twardych jest duży przedział oferowanych pojemności, a co za tym idzie - cen. W zależności od potrzeb i zasobności portfela do dyspozycji mamy napędy o pojemnościach od 20 GB do nawet 2 terabajtów. Najtańsze "twardziele" zewnętrzne kosztują ok. 300 zł (20- lub 40-gigabajtowy napęd w obudowie z portem USB 2.0). Urządzenia te możemy wybierać także pod kątem rozmiarów. Dostepne sa małe i poręczne dyski 2,5" o pojemności maksymalnie do 120 GB, 3,5-calowe zewnętrzne dyski twarde mieszczące do 400 GB czy profesjonalne napędy zewnętrzne z interfejsem FireWire IEEE-1394b o przepustowości 800 Mb/s. Pojemność tych dysków dochodzi do 2 terabajtów. Niestety, ze względu na dość wysokie ceny (dysk o pojemniści 1,6 terabajta kosztuje ok. 7 tys. zł) na profesjonalny naped FireWire o przepustowości 90MB/s stać będzie tylko przynajmniej średnią firmę.

2.5 Kryteria doboru produktów ochrony danych

Wybór rozwiązań informatycznych odpowiadających naszym potrzebom spędza często sen z powiek wielu decydentom, dyrektorom departamentów informatyki, kierownikom projektów itp. Metody wyboru są różne, najczęściej uzależnione od rodzaju poszukiwanego rozwiązania, ale też często od korporacyjnych standardów lub wymagań ustawowych. W praktyce trzeba czasem niektóre procedury wyboru dynamicznie zmieniać (inaczej mówiąc naginać), aby wybrać rzeczywiście najodpowiedniejsze rozwiązanie. Niestety zdarza się też wielokrotnie, że na nasz wybór mają zbyt duży wpływ czynniki marketingowe, które niekoniecznie mają swoje odzwierciedlenie w rzeczywistości, a wybrane rozwiązanie nie spełnia naszych docelowych oczekiwań. Oznacza to, że prędzej czy później przedsiębiorstwo lub instytucja poniesie dodatkowe koszty związane z dostosowywaniem, rozwijaniem lub wymianą nabytych produktów. Aby tego uniknąć warto się zawsze zastanowić, jakie są najważniejsze czynniki, które powinny zadecydować o wyborze odpowiedniego rozwiązania w konkretnym, specyficznym przypadku. Na pewno należy się też nad tym poważnie zastanowić wybierając produkty, które będą zabezpieczały nasze dane, takie jak systemy centralnego wykonywania kopii zapasowych, archiwizacji, wysokiej dostępności. Poniżej kilka podstawowych kroków, które powinniśmy wykonać, aby mieć pewność, że dołożyliśmy starań, aby wybrać najodpowiedniejsze produkty.

Oczywiście podstawową sprawą jest określenie własnych potrzeb i możliwości. To drugie oznacza oczywiście przede wszystkim nasze możliwości finansowe w kontekście kosztów, które ponosimy lub możemy ponieść nie inwestując w nowe rozwiązanie, czy też zysków, które mogą być efektem tej inwestycji. Powraca tu jak bumerang szeroko poruszany temat kosztów przestoju firmy lub wartości jej danych, które należałaby oszacować w każdym większym przedsiębiorstwie czy instytucji. Na temat analizy potrzeb napisano już bardzo wiele i chyba nie ma więc sensu rozwijać tego tematu. Warto natomiast podkreślić, że musimy przewidywać jak nasze potrzeby będą ewoluowały w przyszłości, w jakim kierunku pójdzie rozwój naszych systemów informatycznych, jaka będzie skala tego rozwoju. Należy też bardzo dokładnie zdefiniować, co jest dla nas najważniejsze, np. backup serwerów poczty elektronicznej, systemu, czy też hurtowni danych. Na bazie tych informacji będziemy w stanie określić nasze wymagania dotyczące np. okna czasowego backupu, właściwości i rodzaju nośnika kopii zapasowych, czy wręcz architektury poszukiwanego rozwiązania.

W dalszej kolejności należy dokładnie porównać dostępne technologie. Nie można się przy tym kierować stereotypowymi opiniami (często lokalnymi), które mówią np., że do takich zadań nadaje się tylko taśma, a do innych tylko dysk magnetooptyczny, a jedyne znane oprogramowanie do backupu oferuje firma X i jest ono dobre w każdej sytuacji i w każdym środowisku. Nie ma rozwiązań idealnych i jedynych! Często okazuje się, że rozwój jakiejś technologii poszedł tak do przodu, że zakres jej zastosowań stał się dużo większy niż dotychczas (coraz większe pojemności, coraz szybszy transfer, coraz krótsze czasy dostępu), a w innym przypadku technologia nie wyszła poza zakres obsługi tylko określonych środowisk i mimo, że spisuje się w nich rewelacyjnie, kompletnie nie sprawdza się w innych zadaniach. Co więcej, należy niestety zawsze próbować dowiedzieć się czegoś więcej, niż możemy to znaleźć w specyfikacjach technicznych produktów. Często okazuje się, że po zbudowaniu z kilku produktów całego systemu, żaden z parametrów wydajnościowych nie osiąga wartości podanej w specyfikacji i nie jest to wina urządzenia. Problem leży np. w tym, że serwery NT okazują się być bardziej zaborcze niż kluczowe dla firmy maszyny UNIXowe i zabierają dla siebie większą część wspólnego pasma Fibre Channel. Zamiast spodziewanego wzrostu wydajności kluczowych systemów, mamy jej stagnację lub nawet spadek. A można było to przewidzieć i skonfigurować system inaczej, użyć innych urządzeń i odpowiedniego oprogramowania.

Najtrudniejszym punktem wyboru technologii jest odrzucenie rozwiązań, której nie mają przed sobą przyszłości. Nawet taka firma jak Gartner Group przewidywała kiedyś, że najpopularniejszym systemem operacyjnym na rynku będzie OS/2. Jaki jest tego rezultat wiemy i możemy tylko współczuć tym, którzy zainwestowali w tą technologię na większą skalę. Błędne decyzje w tym zakresie mogą niestety skutkować kompletnym brakiem wsparcia technicznego, problemami z rozbudową systemu, uzupełnianiem go o produkty firm trzecich etc. Mimo że jest to trudne, warto się jednak pokusić o odrzucenie i nie roz-

patrywanie rozwiązań, których szanse na utrzymanie się na rynku są raczej niewielkie.

Decydując się na konkretne rozwiązanie danej firmy w pierwszej kolejności należy sprawdzić jej wiarygodność i oferty. Na dzień dzisiejszy jeszcze stosunkowo niewielu reprezentantów branży "storage" ma jakąś ugruntowaną pozycję w Polsce. W związku z tym powinniśmy sprawdzić, czy dany producent ma silną, udokumentowaną np. odpowiednimi niezależnymi raportami pozycję na rynku w zakresie swojej oferty. Przy czym zdecydowanie warto pokusić się o sprawdzenia, czy jest to tylko odzwierciedlenie pozycji firmy na rynku amerykańskim, czy również w Europie dysponuje ona odpowiednim udziałem w rynku. Cały czas zdarzają się jeszcze przypadki firm, które nieźle wypadają w ogólnoświatowych raportach, natomiast w Europie ich cały support techniczny to dwie osoby. Nie od dziś wiadomo również, że jedną z najlepszych metod oceny konkretnego produktu jest sprawdzenie jego referencji. Jeśli nawet nie ma ich jeszcze w Polsce, to na pewno warto porozmawiać z użytkownikami danego systemu z Europy Zachodniej, szczególnie jeśli należą oni do tej samej co my korporacji. Możemy wtedy usłyszeć kilka słów prawdy na temat atrakcyjności oferowanego nam rozwiązania, a być może przy okazji dowiedzieć się też czegoś o produktach i firmach konkurencyjnych. Jeśli aktualna pozycja producenta na rynku satysfakcjonuje nas, sprawdźmy i oceńmy jego wizję przyszłości, a będziemy wiedzieli, czy za rok, może dwa lata będziemy zmuszeni do porzucenia wybranego systemu i jego wymiany na inny, czy też nie. Bardzo docenia ten fakt m.in. Gartner Group w każdym swoim raporcie oceniając tzw. kompletność wizji i np. przyznając najwyższe noty firmom, które oferują nie tylko oprogramowanie do backupu i zarządzania hierarchicznym składowaniem danych (data management) ale również systemy plików, zarządzanie wolumenami, replikację, systemy wysokiej dostępności (storage infrastructure) oraz zarządzanie w ramach sieci komputerowej zasobami pamięci masowych (enterprise storage resource management). Producent, który jest w stanie zaoferować w tym zakresie kompletną ofertę lub przynajmniej przedstawić wiarygodną wizję jej tworzenia i rozwoju, będzie w przyszłości w lepszej sytuacji, niż ci, którzy borykają się jeszcze z problemami w funkcjonowaniu podstawowych produktów do backupowania.

Kwestia sprawdzania wiarygodności oferenta nie wymaga chyba szczegółowego opisu. Musi to być stabilna firma, która przede wszystkim zapewni odpowiednie wsparcie techniczne. Podkreślić warto chyba jedynie dokładne zbadanie zakresu usług serwisowych (w tym gwarancyjnych). Wbrew pozorom niewielu oferentów może zapewnić odpowiednie czasy naprawy, gwarantując raczej czas reakcji i składając ustne obietnice. W praktyce okazuje się, że wyobrażenia sprzedawcy i klienta dotyczące np. sprzętu zastępczego są bardzo rozbieżne.

Jeśli sprawdziliśmy już dokładanie firmy - producenta i oferenta - przechodzimy do oceny produktów. Absolutnie podstawowym kryterium jest wsparcie dla różnych systemów operacyjnych, baz danych, aplikacji i różnych rozwiązań

sprzętowych. Należy dobrać rozwiązanie, które będzie wspierało jak największą liczbę używanych przez nas systemów (często też używanego przez nas sprzętu). Szczególnie systemów, które okres burzliwego wzrostu i rozwoju mają już za sobą (np. OpenVMS), co oznacza, że mało kto będzie się interesował wprowadzeniem na rynek produktu je wspierającego. W grę wchodzą więc tylko firmy oferujące już takie wsparcie. Teoretycznie jest to wszystko oczywiste, praktyka pokazuje jednak, że np. twórcy oprogramowania do backupu opisują możliwości swoich produktów w tym zakresie w sposób bardzo zawoalowany używając bardzo marketingowych określeń. Na pierwszy rzut oka każdy produkt wspiera praktycznie wszystko, jednak kiedy chcemy się temu przyjrzeć bliżej, możemy mieć trudności z dotarciem do odpowiednich informacji. Jeśli je już jednak zdobędziemy, okazuj się, że szeroko opisywane, znakomite wsparcie dla baz danych Oracle'a oznacza tylko wsparcie dla wersji 7.x i na przykład tylko na platformie Windows NT lub tylko na HP-UX. A to z kolei może zupełnie nie odpowiadać naszym oczekiwaniom. Nie sprawdzając takich informacji narażamy się na dużą stratę czasu poświęconą na rozmowy z oferentem i rozpatrywanie nieodpowiednich produktów. Warto dodać, że nawet jeśli interesujące nas konfiguracje występują w ofercie konkretnego producenta, mogą to być produkty nowe i bardzo "niedojrzałe", co z reguły kończy się dużymi problemami w procesie implementacji rozwiązania, która niekoniecznie musi się zakończyć oczekiwanym przez nas rezultatem i jeszcze większymi problemami jeśli coś przestanie poprawnie funkcjonować już w trakcie użytkowania systemu. Może to np. oznaczać kilkutygodniową przerwę w jego działaniu, podczas której jakieś europejskie lub ogólnoświatowe centrum wsparcia technicznego będzie próbowało (zapewne z pozytywnym skutkiem) rozwiązać problem. Dlatego lepiej trzy razy się zastanowić zanim wybierzemy do zabezpieczania naszych danych zgromadzonych w środowisku w znacznym stopniu UNIXowym, system który jest uważany za najlepszy w środowiskach Novella lub Windows NT, a od pewnego czasu udostępnia również obsługę serwerów UNIX.

Inne elementy produktów, na które należy zwrócić uwagę, to przede wszystkim ich skalowalność i zarządzalność. Są to cechy bardzo ściśle powiązane z architekturą systemu, a o niej z kolei możemy mówić w przypadku zaledwie kilku produktów. Pozostali producenci z reguły dużo mówią o architekturze, ale ich rozwiązania sprowadzają się z reguły do najprostszych konfiguracji z pojedynczym serwerem backupu i brakiem możliwości centralnego zarządzania wieloma instalacjami. W sytuacji gwałtownego wzrostu ilości gromadzonych danych i złożoności sieci komputerowych, wybór takiego rozwiązania może być bardzo ryzykowny i zamiast spodziewanych oszczędności, może spowodować wzrost nakładu pracy i wydatków na jego utrzymanie. Zdecydowanie warto więc przyjrzeć się bliżej dostępnym rozwiązaniom, możliwości ich rozbudowy i centralnego zarządzania zintegrowanego z systemami typu network management (Tivoli, Unicenter, OpenView, BMC), szczególnie pod kątem przewidywanego rozwoju naszej sieci i systemów informatycznych. Jeżeli nadal mamy

do wyboru kilka produktów, należy porównać ich specyficzne cechy funkcjonalne. W tym zakresie większość produktów do wykonywania kopii zapasowych i innych z rodziny storage management wykazuje bardzo duże podobieństwa. Diabeł tkwi jednak w szczegółach. Na tym etapie, mając już do wyboru mniejszą liczbę produktów, możemy się dokładniej tym szczegółom przyjrzeć. Może się okazać, że pewne specyficzne cechy danego produktu pozwalają np. na wykonywanie bardzo szybkich backupów dużych baz danych lub na bardzo szybkie odtwarzanie stanu takiej bazy danych sprzed np. 2 godzin. Z reguły są to jeszcze cały czas rozwiązania oparte na specyficznych technologiach i nie występujące w identycznej formie w ofertach różnych producentów.

Rozdział 3

Praktyczne rozwiązania

Metod sporządzania backupu jest bardzo wiele. Najprostszą i stosowaną (często bezwiednie) przez domowych użytkowników jest ręczne kopiowanie wybranych zbiorów. Decyzja, które są to pliki, podejmowana jest z reguły dość przypadkowo. Nie ulega wątpliwości, że ręczne kopiowanie to zadanie bardzo żmudne i czasochłonne. W sukurs przychodzi nam oprogramowanie. Na rynku dostępnych jest wiele programów i narzędzi znacznie ułatwiających tworzenie backupu. Programy te dysponują bardzo zróżnicowanym zasobem funkcji, dlatego każdy wśród użytkowników znajdzie coś dla siebie.

3.1 Programy dla Windows

Najłatwiejszą metoda wykonywania kopii bezpieczeństwa jest skopiowanie istotnych danych do wybranego folderu. Jeśli ktoś nie ma zamiaru robić niczego więcej w celu zabezpieczenia swoich zbiorów, to w porządku: niech przynajmniej zrobi tyle. Nie potrzeba jednak bardzo przenikliwego umysłu, by stwierdzić, że technika ta ma przynajmniej kilka wad. Wymaga na przykład naszego ciągłego zaangażowania i decydowania, które dane są ważne i gdzie je skopiować. Jeśli ktoś jednak czuje, że to zabawa nie dla niego, niech zainstaluje i skonfiguruje program do backupowania. W ten sposób zapewni bezpieczne schronienie swoim plikom.

Programy dostępne na rynku dysponują bardzo zróżnicowanym zasobem funkcji, dlatego też zapewne każdy znajdzie coś odpowiedniego dla siebie. Dla tych którzy chcą tylko usprawnić sobie kopiowanie danych we właściwe miejsce na dysku, producenci przygotowali aplikacje typu Instant Backup, dzięki którym za niewielka opłata zaopatrzymy się w łatwe w obsłudze i konfiguracji narzędzie. Funkconalność takich aplikacji ogranicza się właściwie tylko do wykonania kopii danego folderu w innym katalogu. Równie prostym programem jest darmowy ABC Backup.

Bardziej zaawansowanymi narzędziemi są Genie Backup oraz Handy Backup. Cena tych aplikacji bywa dość wysoka, ale w zamian otrzymamy progra-

my, które nie tylko przechowuja nasze dane w innym miejscu, ale skompresuja je, zaszyfrują i nagrają na dowolny nośnik. W dodatku wiekszość tych zadań odbedzie się bez naszego udziału, ponieważ mechanizm tworzenia kopii będzie działał w określony przez nas wcześniej odstępach czasu.

Duże archiwa wymuszają użycie większej liczby nośników, na które je nagrywamy. Dlatego bardzo przydatna jest możliwość spakowania backupu. Ostatecznie przyczyni się to do obniżenia kosztów wykonywania kopii bezpieczeństwa, co przy prawidłowo zaplanowanej strategii nie jest bez znaczenia. Programy dzielą się na te, które potrafią skompresować kopię bezpieczeństwa do postaci archiwum (najczęściej ZIP), np. Abacre Backup, AISBackup czy Handy Backup, albo tworzą własny, unikatowy format i system kompresji, np. NBF w NTI Backup NOW!, ZAD w ABC Backupie, SVG w Active Backup Expercie. Wśród nich znajdziemy również takie, które w ogóle nie potrafią "ścisnąć" danych (Instant Backup). W niektórych programach producenci postarali się o to, by użytkownik w zależności od czasu i potrzeb mógł zdefiniować stopień kompresji.

Oczywiście nie wszystkie pliki czy katalogi są dla nas ważne. Musimy pamietać że backup zajmuje miejsce na naszych nośnikach, więc dobrze byłoby, gdybyśmy ustalili które pliki sa dla nas istotne, i zgrywali kopie bezpieczeństwa tylko tych danych. Większość programów pozwala nam na określenie maski, tzn. typu plików które powinny zostać zabezpieczone. Często mamy do dyspozycji przygotowane już przez programistów zestawy rozszerzeń. Z taką sytuacją spotkamy się korzystając z Active Backup Experta. W innych programach reguły te zazwyczaj definiujemy samodzielnie.

Jeśli chodzi o dopracowanie pod względem interfejsu i łatwości użytkowania aplikacji to na wyróżnienie zasługują Genie Backup Menager, Handy Backup oraz NTI Backup NOW!. Korzystając z tych programów użytkownik porusza się po łatwym w obsłudze i przyjemnym dla oka menu.

Jak widać jest wiele programów do tworzenia kopii zapasowych i każdy z nich posiada swoje wady i zalety. Również same systemy Windows oferują swoim użytkownikom możliwość wykonania kopii bezpieczeństwa. Na przykład w wersję 2000 została wbudowana stworzona przez firme Veritas aplikacja Kopia zapasowa 5.0. Jej funkcjonalnośc śmiało można porównać z niektórymi programami biorącymi udział w naszym przeglądzie.

3.1.1 Kopia zapasowa

Narzędzia do tworzenia kopii zapasowych zawsze były częścią systemów operacyjnych Windows i zwykle nazywały się "Kopia zapasowa". Wbrew nieskomplikowanemu wyglądowi i faktowi, że jest ono wbudowane w system, jest dość potężne.

Wraz z pojawieniem się Windows XP szanse na skuteczną ochronę danych przechowywanych na dysku twardym znacznie wzrosły, a to głównie za sprawą narzędzia o nazwie Kopia zapasowa, które standardowo jest instalowane

w Windows XP Professional. Posiadacze wersji Windows XP Home Edition także mają dostęp do programu Kopia zapasowa, lecz nie jest on standardowo instalowany. Aby go zainstalować, należy z płyty instalacyjnej uruchomić plik ntbackup.msi. Znajduje się on w katalogu:

```
\valueadd\msft\ntbackup
```

Za pomocą tego programu możemy zgrać do pliku zawartość całego dysku twardego, natomiast posiadacze Windows XP Professional mogą dodatkowo uruchomić kreator automatycznego odzyskiwania systemu (niestety, niedostępny w wersji Home Edition), co umożliwia przywrócenie komputera (i danych) do stanu używalności nawet po kompletnym „padzie" systemu. Kopia zapasowa jest także składnikiem Windows 2000. Zatem użytkownicy tego systemu także otrzymują w jego cenie przydatne oprogramowanie zabezpieczające dane.

"Kopia zapasowa" w systemie Windows spełnia tzry podstawowe zadania:

- Tworzy kopie zapasowe plików przechowywanych w komputerze w innym miejscu. W skład tej funkcji wchodzi wybór plików, których kopie zapasowe będą tworzone, wybór miejsca, w którym kopia będzie tworzona, ustawienienie dodatkowych opcji i utworzenie harmonogramu regularnego tworzenia kopii zapasowej.

- Przywraca pliki z kopii zapasowej - w skład funkcji wchodzi wybór kopii zapasowej, z której mają zostać przywrócone dane, wybór plików, które mają być przywracane, oraz konfigurowanie zaawansowanych opcji.

- Uruchamia "Kreatora automatycznego odzyskiwania systemu" - jest to szczególny rodzaj kopii bezpieczeństwa, który tworzy awaryjną dyskietkę naprawczą zawierającą konfigurację systemu (umożliwiającą uruchomienie komputera i odzyskanie systemu kopii zapasowej na wypadek awarii systemu) i kopie zapasową wszystkich danych komputera.

Niektóre wersje narzędzia "Kopia zapasowa" w systemach starszych od Windows 2000 obsługiwały jedynie wybrane typy napędów taśm. W systemie Windows 2000 i nowszych, kopie są tworzone jako pojedyńczy plik i mogą być zapisywane na dowolnym, dostepnym napędzie. Obejmuje to napędy taśm, dyski CD-R, CD-RW, napędy Zip, dyski twarde i dyskietki. Kopie zapasowe moga być również zapisane w udostępnionym folderze w sieci, a nawet w folderze w sieci Web.

Ogólnie rzecz biorąc, tworzenie backupu na każdym z tych urządzeń wygląda tak samo. Po prostu należy wybrać napęd, na którym będzie zapisywana kopia zapasowa.

"Kopia zapasowa" dysponuje pięcioma typami backupu:

- Kopia normalna - proces wykonywania kopii zapasowej, w którym kopiowane są wszystkie wybrane pliki. Poszczególne pliki są oznaczone jako te, dla których wykonano kopie zapasowe (atrybut "archiwalny" jest czyszczony).

- Kopia - proces wykonywania kopii zapasowych, w którym kopiowane są wszystki wybrane pliki, jednak poszczególne pliki nie są oznaczane jako pliki, dla których wykonano kopie zapasowe. Jest to przydatne wtedy, gdy zaistnieje potrzeba wykonania kopii zapasowej plików między wykonywaniem normalnych i przyrostowych kopii zapasowych.

- Kopia przyrostowa - proces wykonywania kopii zapasowych, w którym kopiowane są tylko te pliki utworzone lub zmienione od chwili wykonania ostatniej normalnej lub przyrostowej kopii zapasowej. Po wykonaniu kopii pliki są oznaczane jako takie, w których, których backup został wykonany.

- Kopia różnicowa - proces wykonywania kopii zapasowych, w którym kopiowane są pliki utworzone lub zmienione od chwili wykonania ostatniej normalnej lub przyrostowej kopii zapasowej. Po wykonaniu backupu atrybut "archiwalny" nie jest czyszczony.

- Kopia codzienna - proces, w którym kopiowane są wszystkie wybrane pliki zmodyfikowane w dniu, w którym dana kopia jest wykonywana. Po wykonaniu kopii zapasowej atrybut "archiwalny" nie jest czyszczony.

Mając do wyboru pięć różnych typów kopii zapasowych, większość użytkowników korzysta z trzech ulubionych metod . Każda z nich opiera się na regularnym wykonywaniu kopii zapasowej komputera. Pierwsza z nich tworzy codziennie pełne kopie zapasowe systemu, druga tworzy kopie przyrostowe w każdym dniu tygodnia, a ostatnia wykonuje kopię różnicową każdego dnia.

Jeżeli używamy Windows XP lub Windows 2000 i dysponujemy dwoma dyskami twardymi (lub jednym dyskiem podzielonym na co najmniej dwie partycje), możemy w dość prosty sposób zabezpieczyć się przed utratą danych.

Kreator automatycznego odzyskiwania zawarty w programie Kopia zapasowa instalowanym w Windows XP Professional, to rozwiązanie skutecznie chroniące dane nawet w razie awarii uniemożliwiającej w ogóle uruchomienie systemu. Warto używać go w połączeniu z utworzonymi wcześniej kopiami zapasowymi całych dysków/partycji. Dzięki temu łatwo przywrócimy sprawne działanie systemu i nie utracimy żadnych plików, które były zapisane na zarchiwizowanym dysku twardym.

Do backupu i archiwizacji przechowywanych na komputerze danych możemy wykorzystać dodatkowe aplikacje. Doskonałym przykładem programu, dzięki któremu możemy wykonać szybko kopię zapasową dysku lub partycji do pliku, jest Norton Ghost 2002 lub PowerQuest Drive Image 2002. Obydwa

umożliwiają zapis wybranego dysku lub partycji do pliku-obrazu, jak również dokładne skopiowanie (sklonowanie) dysku na drugi dysk. Kopię możemy np. nagrać na płytkach CD. W przypadku obu programów istnieje możliwość korzystania z nagrywarki bez potrzeby ładowania systemu operacyjnego, niezbędne sterowniki sprzętowe zaimplementowano w programach). Wadą tych programów jest to, że... kosztują. Za Norton Ghosta 2002 musimy zapłacić ok. 300 zł, natomiast Drive Image 2002 kosztuje ok. 250 zł.

Wspomnając o nagrywarkach, nie sposób pominąć także funkcji archiwizacyjnych, wbudowanych w program Nero Burning ROM, który jest chyba najczęściej dodawaną do nagrywarek aplikacją do zapisu CD.

3.2 Tar, cpio

W pewnych przypadkach, szczególnie w systemach jednoużytkownikowych, nie jest konieczny wyszukany mechanizm tworzenia kopii zapasowych. Ponieważ administrator i użytkownik to często ta sama osoba, dlatego jest dla niej oczywiste, które pliki są ważne, jak często są zmieniane itd. Polecenia takie jak `tar` i `cpio` mogą w takiej sytuacji okazać się wystarczające do okresowego zabezpieczania ważnych plików na taśmie lub innym nośniku. `Tar` i `cpio` są podobne. Oba programy potrafią zapisywać i odczytywać dane z taśm, do tego potrafią obsłużyć każde medium, z którym potrafi sobie poradzić jądro.

Elementarna znajomość `tar-a` jest niezbędna każdemu użytkownikowi UNIXa lub Linuksa, gdyż archiwa .tar są jedną z podstawowych form rozpowszechniania aplikacji, tekstów źródłowych itp. Stosowany w Linuksie GNU `tar` ma szereg rozszerzeń w stosunku do standardowego, czyli spotykanego w innych systemach UNIXowych. W szczególności dotyczy to możliwości dodatkowej kompresji i dekompresji pliku archiwum oraz tworzenia archiwów wielowarstwowych. Opcje polecenia tar dzieli się na dwie grupy: obowiązkowe i dodatkowe.

Opcje obowiązkowe są następujące:

-A - dołączenie istniejących plików (archiwów) .tar do naszego archiwum,

-c - utworzenie nowego archiwum,

-d - wyszukanie różnic pomiędzy archiwum, a rzeczywistym systemem plikó,w

-t - wyświetlenie zawartości archiwum,

-r - dodanie nowych plików do archiwum,

-u - aktualizacja plików w archiwum,

–**delete** - usunięcie plików z archiwum,

-x - rozpakowanie plików z archiwum.

Wybrane opcje dodatkowe:

-f plik - użycie wskazanego pliku jako archiwum. Wskazany plik może być zwykłym lub specjalnym,

-h - nie zapisuje dowiązań symbolicznych, lecz pliki, na które one wskazują,

-k - przy rozpakowaniu istniejące pliki nie będą nad pisywane,

-M - utworzenie archiwum wieloczęściowego,

-N data - archiwizowanie tylko plików nowszych niż podana data,

-W - weryfikacja archiwum po jego utworzeniu,

-z - kompresja (dekompresja) archiwum programem `gzip`,

-Z - kompresja (dekompresja) archiwum programem `compress`,

-p - dearchiwizacja wszystkich informacji o prawach dostępu.

I tak na przykład następujące polecenie `tar` zapamietuje wszystkie pliki w katalogu /u i w jego podkatalogach i przesyła na domyslne urządzenie polecenia `tar` (zdefiniowane w pliku /etc/default/tar systemu):

```
% tar c /u
```

Standardowym zastosowanie powyższego polecenia jest wykonanie kopii zapasowych w systemach UNIXowych pracujących na mikrokomputerach. Podobny sposób postepowania jest stosowany w przypadku korzystania ze stacji roboczych, gdzie każdy użytkownik jest odpowiedzialny za wykonanie kopii zapasowych własnych plików. Ponizsze polecenie przepisuje wszystkie pliki znajdujące sie w katalogu /u/chavez i jego podkatalogach na napęd taśmowy 1:

```
cd /u/chavez
tar -c -f/dev/rmt1* .[a-z]*
```

Pełną kopię można utworzyć w sposób następujący:

```
tar --create --file /dev/ftape /usr/src
```

W powyższym przykładzie używamy opcji `GNU tara`, oryginalny `tar` obsługuje tylko jednoliterowe opcje. Wersja GNU posiada również inne rozszerzenia: obsługuje bardzo długie ścieżki, podział archiwum na kilku mediach, itd.

Jeżeli tworzona przez nas kopia nie mieści się na jednej dyskietce należy użyć opcji `--multi-volume (-M)`:

```
tar -cMf /dev/fd0H1440 /usr/src
```

Po stworzeniu kopii zalecane jest sprawdzić jej poprawność:

```
tar --compare --verbose -f /dev/ftape
```

Kopia różnicowa może zostać utworzona za pomocą opcji `--newer(-N)`:

```
tar --create --newer '8 Sep 1995' --file /dev/ftape /usr/src --verbose
```

Niestety tar nie zauważa zmian w i-węźle pliku (np. zmiana praw dostępu, lub jego nazwy), można to obejść wykorzystując program find. Skrypty wykonujące te wszystkie operacje są dostępne na wielu serwerach FTP.

W odzyskaniu danych za pomocą polecenia `tar` pomoże nam opcja `--extract (-x)`:

```
tar --extract --same-permissions --verbose --file /dev/fd0H1440
```

Możemy również wyciągnąć specyficzny plik, lub katalog (wraz z podkatalogami i zawartymi w nich plikami):

```
tar xpvf /dev/fd0H1440 usr/src/linux-1.2.10-includes/include/linux/hdreg.h
```

Za pomocą opcji –list (-t) możemy sprawdzić zawartość kopii zapasowej:

```
tar --list --file /dev/fd0H1440
```

Należy pamiętać, że `tar` odczytuje dane sekwencyjnie co przy dużych archiwach jest wolne. Taka metoda działania wynika z budowy taśm (i innych mediów sekwencyjnych). tar nie obsługuje skasowanych plików: jeżeli pełna kopia zawiera plik, natomiast kopia różnicowa nie, plik po odtworzeniu danych będzie istniał. W przypadku niektórych plików może to stanowić poważny problem.

Jeżeli chodzi o `cpio` to przy zastosowaniu opcji `-o` polecenie kopiuje pliki, których ścieżka dostępu jest przekazana na standardowym wejściu (najczęściej za pomocą polecenia ls lub find) na standardowe wyjście. Przekierowanie standardowego wyjścia umożliwia wykonanie polecenia zapisu na wybrany przez nas nośnik. Poniższy przykład pokazuje, w jaki sposób polecenie cpio może być zastosowane do wykonania kopii zapasowej:

```
find . -print | cpio -o > /dev/rfd0
find . -name \*.c -print | cpio -o > /dev/rfd0
```

Pierwsze polecenie kopiuje wszystkie pliki z bieżącego katalogu i jego podkatalogów na taśmę w napędzie 0. Drugie polecenie kopiuje na nosnik wszystkie pliki źródłowe z programami w języku C, znajdujące się poniżej katalogu bieżącego. Niektóre wersje polecenia `find` udostępniaja opcję `-cpio`. Przy jej zastosowaniu analogiczne polecenie wygladałoby następująco:

```
find . -cpio /dev/rmt0
```

Polecenie `cpio` może archiwizować dowolnie wybrany zestaw plików, podczas gdy `tar` jest ograniczone do poddrzew katalogowych. Może również kopiować pliki specjalne. Nadaje się więc do tworzenia kopii zapasowych całego systemu. Upakowuje dane na taśmie znacznie bardziej efektywnie niż `tar`. Użycie polecenia `cpio` zalecane jest wtedy, gdy istotne jest zmieszczenie wszystkich danych na jednym nośniku. Przy odtwarzaniu polecenie to omija uszkodzone miejsca nośnika, podczas gdy tar poprostu przestaje działać.

Poniżej przedstawiony został przykładowy skrypt, który działa na serwerze jednej z firm:

```
#!/bin/sh
#
# Rotating backup
#

BACKUPDIR=/export/home2/backup/auto

PATH=/usr/bin:/usr/sbin:/usr/sfw/bin:/opt/bin:/opt/sbin:
     /opt/cfw/bin
export PATH

rotatebackup () {
  BACKUPFILE=$1
  if test -d $BACKUPDIR
  then
    cd $BACKUPDIR
    if test -s $BACKUPFILE.tgz
      then
        test -f $BACKUPFILE.6.tgz && mv $BACKUPFILE.6.tgz
        $BACKUPFILE.7.tgz
        test -f $BACKUPFILE.5.tgz && mv $BACKUPFILE.5.tgz
        $BACKUPFILE.6.tgz
        test -f $BACKUPFILE.4.tgz && mv $BACKUPFILE.4.tgz
        $BACKUPFILE.5.tgz
        test -f $BACKUPFILE.3.tgz && mv $BACKUPFILE.3.tgz
        $BACKUPFILE.4.tgz
        test -f $BACKUPFILE.2.tgz && mv $BACKUPFILE.2.tgz
        $BACKUPFILE.3.tgz
        test -f $BACKUPFILE.1.tgz && mv $BACKUPFILE.1.tgz
        $BACKUPFILE.2.tgz
        test -f $BACKUPFILE.0.tgz && mv $BACKUPFILE.0.tgz
        $BACKUPFILE.1.tgz
        mv $BACKUPFILE.tgz    $BACKUPFILE.0.tgz
```

```
      fi
   fi
}

gtar czf $BACKUPDIR/mail.tgz -C /var mail
rotatebackup mail

(cd /etc/; find . -type f > /tmp/etcfiles)
gtar czf $BACKUPDIR/etc.tgz -C /etc -T /tmp/etcfiles
rotatebackup etc
rm -f /tmp/etcfiles

gtar czf $BACKUPDIR/cfwetc.tgz -C /opt/cfw etc
rotatebackup cfwetc
gtar czf $BACKUPDIR/magazyn.tgz -C /export/home2 magazyn
rotatebackup magazyn
```

Zadaniem skryptu jest wykonanie kopii zapasowych istotnych dla firmy danych znajdujących się na serwerze. Backup wykonywany jest codzinnie w godzinach nocnych. Kopie zapasowe poddawane są rotacji tak, że możliwy jest powrót do danych do 8 dni wstecz.

Do tworzenia kopii zapasowych używany tutaj jest program `tar` (`gtar` dla określenia GNU tar). Tworzone kopie kompresowane są programem gzip, służy do tego przełącznik -z w programie `gtar`. Kopiowane są pliki konfiguracyjne systemu znajdujące się w katalogach

`/etc`

oraz

`/opt/cfw/etc`,

poczta uzytkoników z katalogu

`/var/mail`

oraz dane programu magazynowego firmy z katalogu

`/export/home2/magazyn`.

3.3 Ufsdump i ufsrestore

`Ufsdump` jest poleceniem służącym do wykonywania kopii zapasowych w systemie Solaris i działa na niskim poziomie systemu plików. Przy jego użyciu jesteśmy w stanie zrobić wierną kopię systemu plików, łącznie z plikami specjalnymi tj. dowiązaniami symbolicznymi, twardymi oraz plikami urządzeń,

z którymi zwykłe programy kopiujące mają problemy. Polecenie to jest dość elastycznym narzędziem. Zapamiętuje pliki niezależnie od tego, gdzie w czasie ich wykonywania były zamontowane systemy plików. Ścieżki dostępu używane przez polecenie są podawane względem katalogu głównego ich własnego systemu plików, a nie całego drzewa katalogowego. Tworzona kopia może być zapisana na taśmie magnetycznej, dyskietce lub dowolnie wybranym nośniku. W trakcie działania `ufsdump`, kopiowany system plików nie może być aktywny, ponieważ utworzona kopia może zostać niewłaściwie stworzona i niemożliwa do przywrócenia.

Ogólna postać polecenia `ufsdump` jest następująca:

```
%ufsdump [opcje] [argumenty] plik-dump
```

gdzie `opcje` stanowią listę opcji użytych do wykonania kopii, `argumenty` listę argumentów odpowiadających poszczególnym opcjom, a `plik-dump` jest blokowym plikiem specjalnym reprezentującym archiwizowany system plików.

Do najważniejszych opcji należy zaliczyć:

0-9 - numery od 0 do 9 wskazują poziom nagrania, które będzie wykonywane przez polecenie `ufsdump`. Dla danego poziomu `n` polecenie to przeszukuje zawartość pliku w celu znalezienia informacji dotyczących czasu wykonania ostatniej kopii zapasowej sporządzonej na poziomie `n-1` lub niższym. Polecenie kopiuje wszystkie pliki zmienione od tego czasu. Jeśli `n` jest równe zeru, polecenie `ufsdump` dokona archiwizacji całego systemu plików. Opcja ta nie wymaga podawania żadnych argumentów.

-u - jeśli polecenie `ufsdump` zakończy się błędnie, opcja ta uaktualni plik /etc/dumpdates i zapisze informacje o przebiegu bieżącego kopiowania,

-f - opcja ta określa, że należy utworzyć kopię na urządzeniu innym niż domyślny napęd taśmy. Jeżeli opcja ta nie została podana, dane zostaną skopiowane na domyślny napęd taśmy.

Poniższy skrypt jest przykładem typowego użycia polecenia `ufsdump`:

```
#!/bin/sh
#
# Backup
#
BACKUPDIR=/export/home2/backup
MAILER="sendmail -t"

PATH=/usr/bin:/usr/sbin:/usr/sfw/bin:/opt/bin:/opt/sbin:
     /opt/cfw/bin:/opt/cfw/sbin
export PATH
```

```
/etc/init.d/apache stop
/etc/init.d/mysqld stop


DAY=‘date ’+%u’‘
if [ "x$DAY" = "x7" ]
then
    LEVEL=0
    rm -f ${BACKUPDIR}/home1-*
else
    LEVEL=$DAY
fi
ufsdump ${LEVEL}uf - /dev/rdsk/c0t1d0s0 | gzip >
${BACKUPDIR}/home1-${LEVEL}.gz
/etc/init.d/mysqld start
/etc/init.d/apache start

cat | $MAILER <<EOF
To: sysadm@solution-4u.com
Subject: Backup na BCMB

Backup na BCMB zrobiony:
‘ls -ltr $BACKUPDIR‘
Dyski:
‘df -kh‘
EOF
```

Jest to skrypt wykorzystujący program `ufsdump`. Skrypt jest dostosowany do konkretnych potrzeb i konkretnej instalacji systemu, bez przeróbek raczej nie będzie prawidłowo działał w innym środowisku.

Zadaniem skryptu jest wykonanie backupu przyrostowego calego wolumenu w systemie plików, gdzie umieszczone są dane bazy MySQL oraz dane związane z serwerem HTTP - dokumenty HTML, logi i statystyki. Skrypt uruchamiany jest codziennie o 2 rano. Przechowywane są poszczególne pliki z backupem, aż do powtórzenia pełnego backupu, który ma miejsce w niedzielę. Jest to najmniej ruchliwy dzień i dopuszczalny jest dłuższy przestój serwera.

Z punktu widzenia zastosowań serwera mozliwe jest zatrzymanie części usług na czas wykonania kopii zapasowych.

Skrypt wykonuje nastepujące czynności:

1. Zatrzymywany jest serwer bazy MySQL i serwer HTTP, tak aby kopia danych była spójna.

2. W zmienej DAY zapamiętywany jest numer dnia tygodnia od 1 do 7. 1 oznacza poniedziałek.

3. Jeśli DAY jest 7 to zmienna LEVEL ustawiana jest na 0 i usuwane są pliki z wcześniej wykonanym backupem. Gdy DAY nie jest 7 LEVEL przyjmuje wartość DAY. LEVEL oznacza poziom wykonywanego backupu, 0 to pelny backup.

4. Wołany jest program ufsdump i kopiowana jest zawartość całego wolumenu /dev/rdsk/c0t1d0s0 do pliku na innym fizycznie dysku. Plik jest kompresowany za pomocą gzip. W nazwie podany jest poziom wykonanego backupu.

5. Uruchamiane są serwer MySQL i HTTP.

6. Do administratora systemu wysyłany jest raport o wykonanym backupie. Raport zawiera listę zgromadzonych plików będących kopiami bezpieczeństwa. Pdana jest też informacja o zajetości wszystkich wolumenów w systemie, tak aby uniknąć za wczasu przepełnienia na którymkolwiek z nich.

Program `ufsdump` zapamiętuje kiedy został wykonany ostatni backup, dla jakiego woluminu i jaki był jego poziom w pliku

`/etc/dumpdates`

```
alpha# cat /etc/dumpdates
/dev/rdsk/c0t1d0s0              0 Sun Jun 12 02:00:06 2005
/dev/rdsk/c0t1d0s0              4 Thu Jun 16 02:00:03 2005
/dev/rdsk/c0t1d0s0              5 Fri Jun 17 02:00:05 2005
/dev/rdsk/c0t1d0s0              6 Sat Jun 11 02:00:04 2005
/dev/rdsk/c0t1d0s0              1 Mon Jun 13 02:00:07 2005
/dev/rdsk/c0t1d0s0              2 Tue Jun 14 02:00:04 2005
/dev/rdsk/c0t1d0s0              3 Wed Jun 15 02:00:06 2005
```

Polecenie `ufsrestore` umożliwia odzyskanie danych z kopii zapasowych tworzonych uprzednio za pomocą polecenia `ufsdump`. Narzędzie to może odtworzyć pojedyńczy plik, katalog lub cały system plików. Aby odtworzyć cały system plików, musimy odtworzyć ostatnie kopie zapasowe każdego poziomu: ostatnią pełną kopię zapasową (poziom 0), ostatnią kopię zapasową poziomu 1 itd. Konieczne jest odtwarzanie kopii każdego poziomu we właściwym porządku, poczynając od poziomu 0. Jeśli tego nie zrobimy, może dojść do zachowania starych wersji pewnych plików w systemie. polecenie `ufsrestore` umieszcza odtwarzane pliki w katalogu bieżącym. Dlatego, aby odtworzyć cały system plików, może zaistnieć potrzeba utworzenia i zamontowania na nowo systemu plików.

Ogólna postać polecenia `ufsrestore` jest następująca:

```
% ufsrestore opcje argumenty [pliki-i-katalogi]
```

gdzie `opcje` stanowią listę opcji, `argumenty` listę argumentów odpowiadających opcjom, a `pliki-i-katalogi` listę plików i katalogów do odtwarzenia z taśmy. Jeśli nie podano żadnych plików, wówczas odtwarzona zostanie cała zawartość kopii zapasowej.

Polecenie `ufsrestore` z opcją `-i` może być uruchomione w interakcyjnym trybie pracy. Wydaje się, że jest to najbardziej wygodny i łatwy sposób jej użycia. A oto inne ważniejsze opcje:

-r - odczytanie i odtworzenie całej taśmy. Jest to bardzo potężne narzędzie i powinno być używane do odtwarzania całego systemu plików, zapisanego na jednej lub kilku taśmach,

-R - wznowienie częściowo wykonanej operacji odtwarzania. Przy zastosowaniu tej opcji `ufsrestore` żąda podania konkretnej taśmy w celu kontynuowania pełnego odtwarzania,

-f - argumentem tej opcji jest nazwa pliku lub urządzenia przechowującego kopię zapasową. Jeśli argument jest pominięty, `ufsrestore` zakłada, że taśma z kopią zapasową znajduje się w domyślnym napędzie. Niewłaściwe użycie tej opcji może zniszczyć system plików.

Rozdział 4

Metody zmniejszenia ryzyka utraty danych

Systemy komputerowe stały się wszechobecne w każdej dziedzinie naszego życia, spotyka się je wszędzie: w domu, w pracy, w banku, w urzędzie, sklepie etc. Zadania, które są im powierzane sprawiają, że w coraz większym stopniu jesteśmy od nich uzależnieni, a jakakolwiek nieprzewidziana przerwa w pracy systemów komputerowych paraliżuje nasze codzienne działania. Czy ktoś jest w stanie wyobrazić sobie na przykład pracę współczesnego banku bez systemu informatycznego?

W celu poprawienia niezawodności i zmniejszenia do minimum możliwości wystąpienia jakiejkolwiek awarii stosuje się różnego rodzaju zabezpieczenia, na różnych poziomach funkcjonwania systemu: np. odpowiednie systemy zasilania, nadmiarowe serwery plików, macierze dyskowe, systemy archiwizacji i backup'u danych.

Bardzo trudno, a nawet niemożliwe jest podanie uniwersalnego modelu systemu zabezpieczania danych ze względu na fakt, iż każdy system komputerowy konkretnego użytkownika jest inny. Sposób podejścia do przygotowania właściwego rozwiązania należy rozpatrywać w różnych, wzajemnie przecinających się płaszczyznach:

- ilość i rodzaj zabezpieczanych serwerów,
- systemy operacyjne podlegające zabezpieczeniu,
- całkowita ilość danych które należy chronić,
- ilości danych do zabezpieczenia rozkładająca się na poszczególne serwery,
- czas jaki można przeznaczyć na wykonanie operacji backup'u,
- wydajność napędów taśmowych i medium transmisyjnego pomiędzy serwerami,

- jakie i na jakich systemach operacyjnych pracują serwery bazodanowe.

Jak widać na podstawie powyżej wymienionych czynników podejście do przygotowania rozwiązania konkretnego przykładu może być różne i może być tak różne ile osób je opracowuje, tzn. można przygotować kilka przykładów rozwiązań systemu zabezpieczania danych, które różnią się znacznie, ale wszystkie spełniają wcześniej postawione założenia.

4.1 Systemy zabezpieczania danych

4.1.1 Typ Small Office

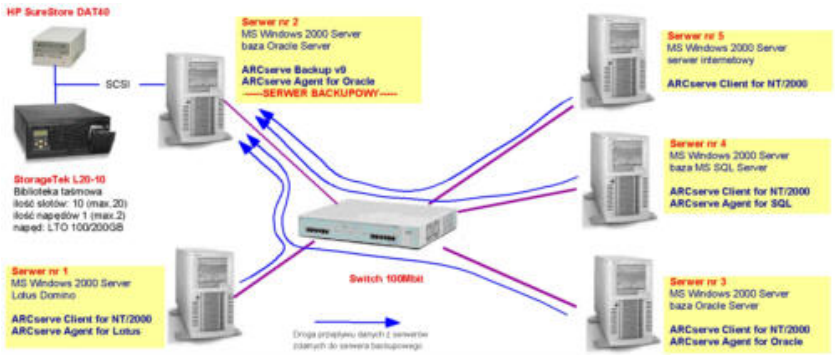
Gdy zabezpieczeniu podlega tylko jeden serwer dobór oprogramowania do backup'u uzależniony jest od systemu operacyjnego serwera:

- Novell NetWare: Backup Exec, BrightStor ARCserve
- Windows NT/2000/2003: Backup Exec, BrightStor ARCserve, Legato NetWorker
- Linux: BrightStor ARCserve, Legato NetWorker
- Unix: BrightStor Enterprise Backup, Legato NetWorker, IBM TSM

W przypadku gdy na serwerze dodatkowo pracuje system bazodanowy np. Informix, Oracle, MS SQL lub pocztowy np. MS Exchange lub Lotus Domino, a chcemy zabezpieczyć te systemy w trybie on-line konieczne jest zainstalowanie dodatkowych modułów tzw. Business Module lub Backup Module koniecznych do wykonywania kopii systemu bazodanowego bez konieczności jego wyłączania (backup w czasie pracy).

4.1.2 Typ Workgroup

Jeżeli do zabezpieczenia mamy większą liczbę serwerów tzn 2 - 5, a środowisko sieciowe jest niejednorodne tzn. współpracują w sieci serwery NetWare, Windows, Linux oraz Unix dobór odpowiedniego oprogramowania nie jest już sprawą tak oczywistą.



Rysunek 4.1: Schemat systemu typu Workgroup.

Wiele zależy od tego w jaki sposób, kiedy i w jakim czasie musimy wykonać kopie bezpieczeństwa systemu informatycznego przedsiębiorstwa oraz od tego jaka platforma sprzętowa zostanie wybrana jako serwer backupowy.

Dobór oprogramowania podobnie jak w przypadku systemów typu Workgroup może opierać się na oprogramowaniu do backup'u i archiwizacji takim samym jak wyżej w Small Office.

4.1.3 Typ Enterprise

Jeśli do zabezpieczenia mamy większą liczbę serwerów, a środowisko sieciowe jest heterogeniczne tzn. współpracują w sieci serwery NetWare, Windows NT i Unix,a system informatyczny przedsiębiorstwa oparty jest o strukturę wielooddziałową sprawa doboru odpowiedniego sprzętu i oprogramowania nie jest już rzeczą najważniejszą.

Aby przygotować system ochrony i zabezpieczania danych oprócz podstawowych parametrów takich jak:

- ilość danych podlegająca ochronie,
- okres wykonywania backup'u,
- czas niezbędny na wykonanie backup'u,
- wybór platformy serwera backupowego,
- dopasowanie urządzeń,
- dopasowanie oprogramowania,

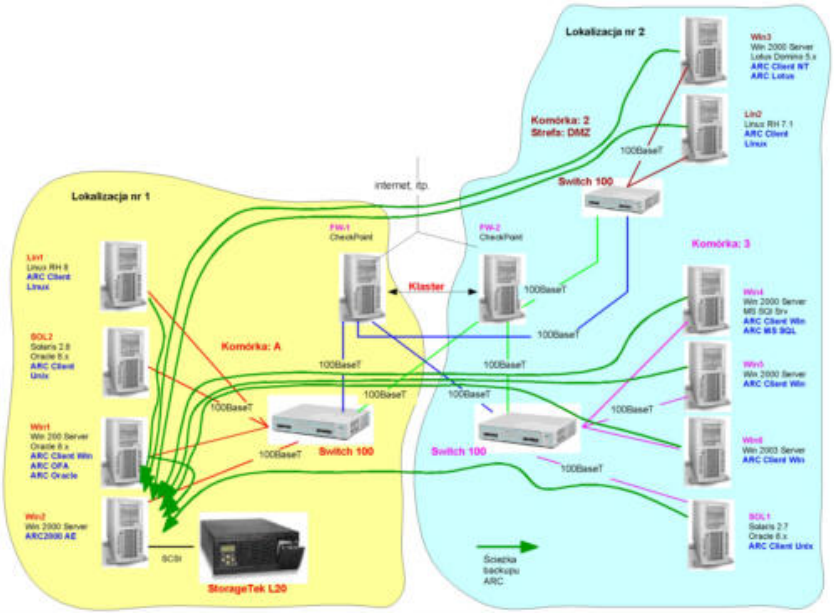
należy brać pod uwagę także inne elementy systemu informatycznego przedsiębiorstwa takie jak:

- topologię sieci informatycznej,

- strukturę sieci i parametry techniczne (ATM, FastEthernet, GigaEthernet, Frame Relay),
- przepustowość sieci (w Mb/s),
- obciążenie sieci,
- centralne zarządzanie operacjami backup'u (konsola zarządzająca).

Zarówno urządzenia do backup'u i archiwizacji (streamery, autoloadery i biblioteki taśmowe) jak i oprogramowanie do zarządzania operacjami backup'u i archiwizacji muszą być dostosowane do potrzeb i wymagań klienta.

Niestety nie wszystkie dostępne na rynku urządzenia i oprogramowanie do backup'u spełniają wymagania klienta w zakresie ochrony i zabezpieczania danych, dlatego dobór odpowiedniego oprogramowania jest bardzo istotny.



Rysunek 4.2: Schemat systemu typu Enterprice.

Wśród narzędzi do obsługi i administracji systemów ochrony danych w systemach Enterprise znajduje się oprogramowanie:

- Novell NetWare: Backup Exec, BrightStor ARCserve
- Windows NT/2000/2003: Backup Exec, BrightStor ARCserve, Legato NetWorker
- Linux: BrightStor ARCserve, Legato NetWorker
- Unix: BrightStor Enterprise Backup, Legato NetWorker, IBM TSM

Zbudowany w oparciu o powyższe wytyczne system ochrony i zabezpieczenia danych Small Office, Workgroup, Enterprise musi charakteryzować się:

- wysoką niezawodnością i funkcjonalnością,
- szybkością działania,
- dużą pojemnością,
- możliwością kompletnego zabezpieczania danych w sieci heteregonicznej wieloserwerowej zarówno teraz jak i w przyszłości,
- możliwością zabezpieczania baz danych on-line (24x7),
- otwartą architekturą umożliwiającą jego rozbudowę w przyszłości,
- pełną automatyzacją procesu ograniczając w ten sposób nadzór ze str.ony użytkownika (administratora) do niezbędnego minimum,
- możliwością zarządzania z dowolnego punktu w sieci.

Najważniejszym jednak elementem całego systemu ochrony i zabezpieczania danych jest urządzenie, które wykonuje centralnie wszelkie operacje backup'u, archiwizacji, odtwarzania i zabezpieczania danych oraz gromadzi i składuje dane z centrali oraz ze wszystkich jednostek organizacyjnych.

W celu zautomatyzowania systemu ochrony i zabezpieczania danych oraz ściśle według założeń przedstawianych przez klienta konieczne jest zastosowanie autoloadera taśmowego (biblioteki taśmowej) automatyzującego proces zabezpieczania i składowania danych co znacznie podnosi bezpieczeństwo systemu eliminując możliwość popełnienia błędu ze strony administratora.

4.2 Dfs

Dfs jest istotną częścią Win2K, ale jego nazwa wprowadza w błąd. Dfs nie jest prawdziwym systemem plików jak NTFS lub FAT, nie oznacza także sposobu organizowania danych na magnetycznych lub optycznych nośnikach pamięci. Jest to raczej mechanizm umożliwiający organizowanie sposobu prezentowania współdzielonych danych użytkownikom oraz utrzymywania replikowanych kopii w wielu lokalizacjach.

Krótko mówiąc, Dfs umożliwia zbudowanie wirtualnych udziałów plików w sieci i przypisanie tych udziałów do fizycznych udziałów na wielu serwerach. Użytkownicy mogą podłączyc się do tych wirtualnych udziałów, tak jak do udziałów fzycznych. Kiedy użytkownicy podłączają się do wirtualnych udziałów, Dfs kieruje ich do kopii danych, która jest dostępna w lokalizacji lub do jednej z kilku replikowanych kopii dostępnych w sieci.

Za magią tego rozwiązania kryje się usługa replikacji plików - FRS (File Replication Service). FRS pojawiło się w Win2K i nie wymaga instalacji (Dfs jest tylko jego komponentem). Rozwiązanie to posiada jednak pewne ograniczenia - nie jest idealnym dla wszystkich potrzeb związanych z replikacją danych, ale jest użyteczne w pierwszej fazie. Przy jego wykorzystaniu należy pamietać jakiego typu dane możemy replikować przy użyciu Dfs. Przykładowo, Dfs działa dobrze z plikami, które użytkownicy otwierają, zmieniają, a następnie zamykają aż do następnego otwarcia. Dfs rozpoznaje, że plik został zmodyfikowany i replikuje nową wersję tego pliku przez sieć, jeżeli jest to wymagane. Pamiętajmy, Dfs replikuje cały plik, a nie tylko zmiany. Dlatego też, jeżeli zmodyfikujemy duży plik, Win2K skopiuje cały nowy plik do wszystkich replik w sieci.

Oprócz przenoszenia plików danych z jednej repliki do drugiej, Win2K replikuje także uprawnienia pomiędzy replikami, które zostały zdefiniowane w strukturze Dfs. Jeżeli dodamy lub usuniemy uprawnienia do folderu lub pliku w jednej lokalizacji, Dfs wykona replikację tych zmian w środowisku.

Jedyną rzeczą której nie replikuje Dfs jest blokada pliku - czyli wskaźnik, którego Win2K używa do określenia czy ktoś inny pracuje z danym plikem. Jest to ważna informacja, która może mieć wpływ na definiowanie struktur Dfs. Ponieważ Dfs nie replikuje blokad plików, dwóch użytkowników może pracować z tym samym plikiem w tym samym czasie, każdy z inną jego kopią i całkowicie bez świadomości istnienia innej kopii. Aby uniknąc takiej sytuacji możemy poinformowac użytkowników, aby zawsze traktowali synchronizowane repliki tylko jako backup udziału, do którego nigdy nie powinni odwoływać sie bezpośrednio.

Jednym ze sposobów synchronizacji replik jako backupu udziałów jest zdefiniowanie udziałów dla poszczególnych ośrodków, a następnie wykonanie backupu tych udziałów w innym ośrodku. Przykładowo, repozytorium dokumentów działu sprzedaży Nowego Jorku może być udziałem o nazwie NYSALESDOCS. Aby replikować te informacje do Londynu, tworzymy katalog, udostępniamy go jako NYSALESDOCS-BACKUP, a następnie definiujemy go jako replikację Dfs. Taka konwencja nazw pomoże nam w identyfikacji poszczególnych udziałów na serwerze w Londynie. Ponieważ replikacja Dfs jest uruchamiana przez zmianę pliku, Dfs nie działa dobrze z plikami baz danych, które sa zawsze otwarte. Pamiętajmy także, że replikacja Dfs jest podobna do dublowania dysków.

Dfs to solidny i niezawodny mechanizm zabezpieczania plików przed katastrofami dla określonych typów plików w organizacji. Zważywszy na to że już zapłaciliśmy za to rozwiązanie, możemy zaimplementować Dfs jako część planu na wypadek katastrofy i użyć go najlepiej jak potrafimy.

4.3 Macierze RAID

Niskie ceny twardych dysków oraz coraz popularniejsze ostatnio płyty główne wyposażone w zintegrowany kontroler macierzowy RAID wręcz prowokują do zbudowania rozsądnym kosztem komputera z wyjątkowo wydajnym lub odpornym na uszkodzenia podsystemem dyskowym. Idea RAID (Redundant Array of Inexpensive/Independent Disks) polega na spięciu ze sobą "twardzieli", tak aby scalić pojemność napędów składowych i uzyskać wzrost prędkości zapisu/odczytu lub utworzyć drugi dysk zwierciadlany, na którym zapisana będzie kopia danych z pierwszego napędu.

Głównym zadaniem macierzy dyskowych jest pierwszy stopień zabezpieczenia danych przed utratą pozwalajacy na nieprzerwaną pracę systemu podczas awarii pojedynczego urządzenia i łatwość odbudowy systemu po awarii podczas jego ciągłej pracy oraz: zwiększenie wydajności całego systemu, uzyskanie bardzo dużej, ciągłej przestrzeni w jednym logicznym woluminie, a także łatwość rozbudowy systemu (powiększenie pojemności macierzy) w trakcie jego ciągłej pracy. Macierze RAID, dość powszechnie stosowane w serwerach, stosunkowo rzadko trafiają do stacji roboczych czy komputerów osobistych. Istnieje kilka typów macierzy dyskowych RAID pozwalających zabezpieczyć dane:

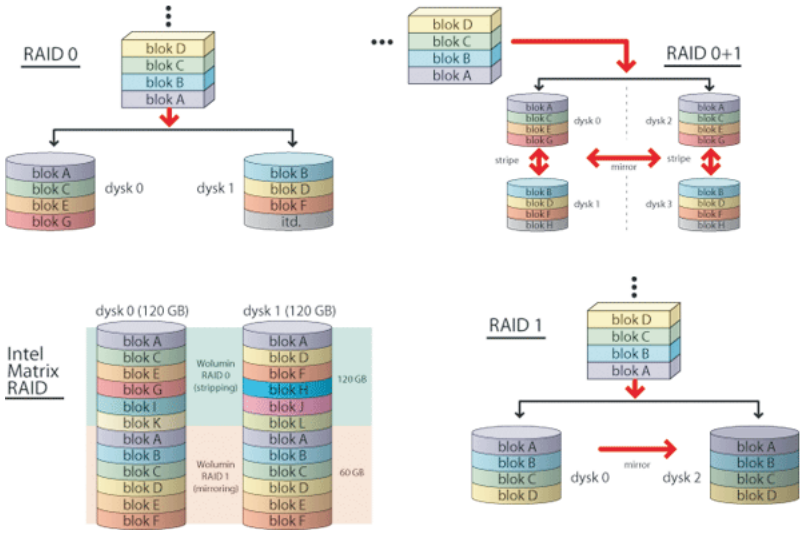
- RAID 0 (stripe) - ten typ umożliwia złączyć dwa lub więcej napędów w jeden dysk logiczny czyli np. z dwóch dysków 80-gigabajtowych zostaje utworzony jeden o pojemnośco 160 GB. Dzięki temu, że zapis i odczyt danych prowadzone są równolegle w obu napedach jednocześnie, wydajnośc macierzy RAID 0 jest niemal dwa razy większa (przy większej liczbie napędów wydajność rośnie proporcjonalnie) niż pojedyńczego dysku twardego. Wadą tego typu macierzy jest ryzyko utraty wszystkich danych w wypadku awarii jednego z dysków. W połaczeniu jednak z innym RAID-em doskonale chroni dane przed utratą.

- RAID 1 (mirror) - dane z dysku nadrzędnego duplikowane są na drugim napędzie zapasowym. W przypadku utraty plików z pierwszego "twardziela" dane są odzyskiwane z drugiego napędu. Pojemność takiej macierzy jest równa objętości jednego dysku. Zaspis w takiej sytuacji zajmuje tyle czasu, ile jednego dysku, natomiast prędkość odczytu rośnie dwukrotnie, bo macierz czyta dwa dyski przemiennie.

- RAID 0+1 - do stworzenia tej macierzy potrzebne będą cztery dyski lub ich większa ale parzysta liczba. Jest to połączenie RAID 0 i 1, czyli uzyskujemy prawie dwukrotny wzrost wydajności, a jednocześnie chronimy dane. Pojemność macierzy jest równa połowie objętości wszytkich napędów.

- RAID 3 - scalanie dysków z jednym dedykowanym dyskiem dla danych (Disk Striping with Dedicated Parity Disk.) Pod tym skomplikowanym

terminem kryje się macierz, która zapewnia co prawda nieco mniejszą wydajność niż RAID 0, ale dobre zabezpieczenie danych. RAID 3 wykonuje scalanie na poziomie bajtów, i zapisywane na osobnym dysku. W ten sposób zapewnione jest odtworzenie danych w przypadku ich uszkodzenia na jednym z dysków.



Rysunek 4.3: Typy macierzy RAID

- RAID 5 - scalanie dysków z bitami parzystości rozłożonymi na wszystkie dyski (Striping with Interspersed Parity). Idea pracy RAID 5 jest podobna do RAID 3. Różnica polega na tym, że dane o parzystości nie są składowane na jednym dedykowanym do tego celu dysku, a są równomiernie rozłożone na wszystkich dyskach macierzy. RAID 5 jest algorytmem wielotorowym zarówno dla zapisu jak i odczytu co pozwala uzyskiwać wyższe wydajności w porównaniu z RAID 3 szczególnie w aplikacjach typu baz danych. Efektywna pojemność macierzy N - 1, gdzie N to liczba dysków.
- RAID Matrix RAID - najnowszy typ macierzy, której pracą steruje wbudowany w układy ICH6R/RW (chipsety Intel 915 i 925X/XE) kotroler RAID. Umożliwia on założenie dwóch macierzy RAID typu 0 i 1 obok siebie zaledwie na dwóch dyskach Serial ATA. Dzięki temu np. połowę objętości napędów możemy przeznaczyć na macierz RAID 0 i zainstalować na niej system operacyjny (wtedy zyskujemy na wydajności i komputer działa szybciej). Drugą połowę pojemności dysków przeznaczamy na macierz RAID 1 i trzymamy na niej ważne pliki, chroniąc je tym samym przed utratą.

Podsumowanie

Rozwiązania zabezpieczające dane przed utratą są kluczowym zagadnieniem dla wszystkich organizacji. Wdrożenia związane z systemem backup'u przeważnie obejmują swoim zasięgiem wiele obszarów systemu komputerowego. Jednym z najlepszych stosowanych rozwiązań są systemy centralnego backup'u gdzie, dane z systemów w bezpieczny sposób docierają do centralnego punktu składowania, w którym są bezpiecznie przechowywane. W wypadku utraty danych w którymś z systemów mogą one być szybko odtworzone z centralnego punktu. Nie mniej istotnym zagadnieniem od zapewnienia szybkiej możliwości odtworzenia danych jest sposób na szybkie przywrócenie systemu do pracy po awarii. Stosowane rozwiązania opierają się na automatycznych rozwiązaniach np. systemie NetBackup, lub tam gdzie to jest niemożliwe przetestowanych procedurach odtwarzania systemu wspartych skryptami. Nie jest to jednak stałą regułą, którą możemy posługiwać się w każdym przypadku. Zapewnienie bezpieczeństwa danych jest bardzo złożonym problemem i zależnym od wielu warunków, które należy za każdym razem oddzielnie rozpatrywać i dopasowywać do danej sytuacji.

Bibliografia

[1] Adamski A., *Prawo karne komputerowe*, Wyd. C.H.BECK, Warszawa 2000.

[2] *Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych*, Dz.U. 1997 r., nr 133, poz. 883.

[3] *Ustawa z dnia 29 września 1994 r. o rachunkowości*, Dz.U. 1994 r., nr 121, poz. 591.

[4] *Ustawa z dnia 27 lipca 2001 r. o ochronie baz danych*, Dz.U 2001 r., nr 128, poz. 1402.

[5] *Enter magazyn komputerowy*, nr 6/2001

[6] *Systemy IT*, Wyd. IT PRESS, Warszawa 2003.

[7] *Chip*, nr 5/2005 Wyd. MULTIMEDIA VISION, Warszawa 2005.

[8] *Dokumentacja techniczna*, SUN MICROSYSTEMS, http://docs.sun.com.

[9] Kopie zapasowe i archiwizacja danych, http://www.robomatic.pl/?id=enchaslo&idh=117.

[10] Kopie zapasowe, http://www.linuxpl.org/SAG/c2107.html.

[11] Replikacje, SAN, NAS, http://integrator.solidex.pl/190_125.html.

[12] Kopie bezpieczeństwa, http://www.bigvent.pl/storge.