Bezpieczeństwo systemów komputerowych
Archiwizacja i Backup

Prof. dr. hab. I. Grabowski

2017

Motto dnia:

„Ludzie dzielą się na tych, którzy wykonują kopie zapasowe
i na tych, którzy dopiero będą je wykonywali”.

Kopia bezpieczeństwa

Plan

- Motywacja
- Zdefiniowanie problemu
- Wymagania prawne (czyli Państwo myśli za Ciebie)
- Wymagania techniczne
- Wymagania organizacyjne
- Rozwiązanie problemu

Bezpieczeństwo danych elektronicznych

Zapewnienie bezpieczeństwa systemów komputerowych to ogół działań mających na celu zabezpieczenie danych przechowywanych w systemie komputerowym, tak by nie mogły zostać wykorzystane przez niepowołane osoby czy też narażone na trwałą lub nawet tymczasową utratę.

Bezpieczeństwo jest elementem szerszego kontekstu, nazywanego wiarygodnością systemu komputerowego.

System wiarygodny =

- dyspozycyjny (available) = dostępny na bieżąco
- niezawodny (reliable) = odporny na awarie
- bezpieczny (secure) = zapewniający ochronę danych

Filary bezpieczeństwa systemów komputerowych

- Poufność - Zespół wszystkich działań mających na celu zapobieganie by informacja zastrzeżona nie dostała się w niepowołane ręce
- Integralność - Mechanizm, gwarantujący, że kluczowe dane nie zostaną zmodyfikowane przez nieautoryzowanego użytkownika
- Dostępność- Nieprzerwany dostęp do zasobów lub informacji oparty m.innymi na autoryzowanym dostępie do tychże danych
- Niezawodność - Pewność, że system będzie działał stabilnie w oczekiwany przez użytkowników sposób,(odporny na awarie)
- Autentyczność - Weryfikacja tożsamości i autentyczności zasobów

Motywacja

Informacja: W obecnych czasach, rzetelna i co najważniejsze dostępna informacja, niejednokrotnie stanowi o możliwości efektywnego działania i właściwego rozwoju organizacji.

Widać wyraźnie, że w dzisiejszych czasach wykorzystanie informacji jest związane z funkcjonowaniem instytucji/firmy/przedsiębiorstwa ściślej niż kiedykolwiek wcześniej.

Informacja -> dane -> zapis -> …. -> nośniki elektroniczne (obecnie)

Motywacja

Dane „produkowane” w firmie:

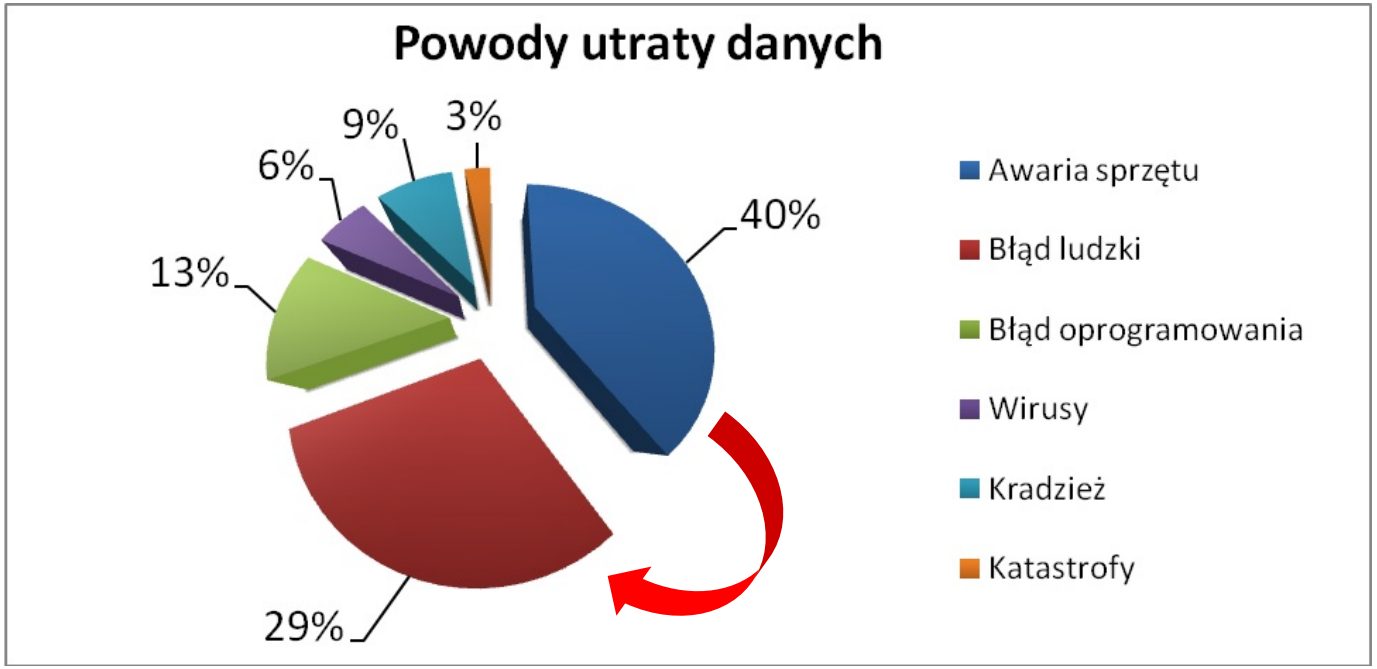
Około70 procent dokumentów, jakie powstają obecnie w firmach lub są tworzone przez użytkowników prywatnych w Polsce i na Świecie, ma postać cyfrową.

Większość bo około 90 procent z nich nigdy nie zostanie jednak wydrukowanych i pozostaną one jedynie w wersji elektronicznej (komputerowe pamięci masowe: HDD, PenDrive, CD/DVD, Internet / chmura, serwery plików, serwery poczty e-mail, serwery WWW, itp…)

A przecież informacja i dane to nie tylko dokumenty (bazy danych, struktury, konfiguracja systemów zarządzania i sterowania, zdjęcia, itp.)

Główne przyczyny utraty danych (duże zmiany w kolejnych badaniach)

- Błąd człowieka
- Awaria sprzętu
- Działanie złośliwego oprogramowania
- Błędy aplikacji
- Klęski żywiołowe np. powodzie.



Kopie bezpieczeństwa (backup) to konieczny element zabezpieczenia danych przed utratą

Backup

To ochronna kopia na bieżąco przetwarzanych danych (a także systemu operacyjnego i zainstalowanych aplikacji) z serwera lub stacji roboczych. W razie wystąpienia awarii pomaga w krótkim czasie przywrócić system informatyczny do stanu z momentu wykonania ostatniego lub ostatniego poprawnego backupu.

- Wykonywane dla zapewnienia bezpieczeństwa danych (systemu).
- Decyduje administrator systemu
- Niezbyt długi czas przechowywania
- Konieczne dane (w zależności od rodzaju kopii bezpieczeństwa.)

Kopie bezpieczeństwa (backup) to konieczny element zabezpieczenia danych przed utratą

Archiwum

Odbywa się wyłącznie na potrzeby użytkowników systemu informatycznego. W procesie tym do archiwum trafiają stare i rzadko przetwarzane dane, dzięki czemu mogą być przeniesione na wolniejsze od dysków twardych, ale tańsze (*i bezpieczniejsze*) nośniki.

- Gromadzenie danych (np. z pewnego okresu)
- Decyduje użytkownik
- Długi okres przechowywania.

Motywacja - jest źle!!!

Przeprowadzone przez Instytut Rozwoju Technologii Informatycznych badania, wśród kadry zarządzającej małymi i średnimi przedsiębiorstwami pokazały że:

67% respondentów nie archiwizuje regularnie kluczowych danych.

81% spośród tych, którzy zadeklarowali, iż regularnie archiwizują dane, przyznało się, iż zdarza się im nie wykonać archiwizacji przez okres dłuższy niż 30 dni!

Jako najczęstszą przyczynę takiego stanu rzeczy podają brak czasu i nadmiar obowiązków i inne powody („*przecież nic się nie stało przez ostatnie 10 lat działania firmy*”).

Test TAK NIE NIE WIEM

właściciel/dyrektor/dziekan/pracownik/student/osoba prywatna

1. Czy w Twoje firmie /w domu „produkowane” / gromadzone/przetwarzane są dane, dokumenty, które istnieją tylko w wersji elektronicznej?
2. Czy wiem ile takich danych jest codziennie „produkowanych”?
3. Czy wśród tych danych są takie, których utrata spowoduje duże kłopoty finansowe lub takie dane, których nie będziemy w stanie odtworzyć?
4. Czy w Twojej firmie wdrożono politykę bezpieczeństwem danych elektronicznych?
5. Ile czasu Twoja firma może „wytrzymać” bez systemu komputerowego?
6. Czy w firmie, w której pracujesz, którą zarządzasz, wykonywane są kopie bezpieczeństwa danych?
7. Czy wiesz jak zorganizowany jest ten proces (jak często, kto jest odpowiedzialny, jak i gdzie przechowywane są kopie bezpieczeństwa)?
8. Czy kiedykolwiek skorzystałeś z wykonanej wcześniej kopii bezpieczeństwa?
9. Kto w Twojej firmie potrafi odtworzyć utracone dane elektroniczne?
10. Kiedy ostatnio wykonywałeś kopie bezpieczeństwa swoich danych (prywatnych)?
11. Czy uważasz, że jesteś wystarczająco zabezpieczony?
12. Czy pozwolisz mi na prosty „siłowy test” twoich zabezpieczeń?

Testu ciąg dalszy – czyli prosty – siłowy test uświadamiający

1. Pozwól mi skasować dowolny plik w Twoim komputerze.
2. j/w tylko w na komputerze firmowym.
3. Pozwól mi „zniszczyć” dysk twardy w Twoim serwerze/komputerze firmowym.
   (oddam Wam taki sam lub lepszy dysk, tylko „czysty”)

Test u ciąg dalszy – czyli prosty – siłowy test uświadamiający

1. Pozwól mi skasować dowolny plik w Twoim komputerze
2. j/w tylko w na komputerze firmowym
3. Pozwól mi „zniszczyć” dysk twardy w Twoim serwerze/komputerze firmowym

(oddam Wam taki sam lub lepszy dysk, tylko „czysty”)

Jeśli mi NIE pozwoliłeś, lub się zawahałeś, tzn., że nie jesteś odpowiednio zabezpieczony!

Test u ciąg dalszy – czyli siłowy test uświadamiający

1. Pozwól mi skasować dowolny plik w Twoim komputerze
2. j/w tylko w na komputerze firmowym
3. Pozwól mi „zniszczyć” dysk twardy w Twoim serwerze/komputerze firmowym

Jeśli mi nie pozwoliłeś, tzn., że nie jesteś odpowiednio zabezpieczony!

Jeśli jesteś właścicielem/prezesem/kierownikiem/pracownikiem firmy – zapytaj o punkty 1,2,3 swojego informatyka/firmę informatyczną/pracownika odpowiedzialnego za kopię bezpieczeństwa.

Test u ciąg dalszy – czyli siłowy test uświadamiający

1. Pozwól mi skasować dowolny plik w Twoim komputerze
2. j/w tylko w na komputerze firmowym
3. Pozwól mi „zniszczyć” dysk twardy w Twoim serwerze/komputerze firmowym

Jeśli mi nie pozwoliłeś, tzn., że nie jesteś odpowiednio zabezpieczony!

Jeśli jesteś właścicielem/prezesem/kierownikiem/pracownikiem firmy – zapytaj o punkty 1,2,3 swojego informatyka/firmę informatyczną/pracownika odpowiedzialnego za kopię bezpieczeństwa.

Jeśli nie wiesz co zrobić – zapytaj fachowca!

Kopia bezpieczeństwa - zagadnienie jest złożone

- Znaczne skomplikowanie i złożoność obecnych infrastruktur informatycznych - np. infrastruktura rozproszona, dostępna tylko czasowo, …
- Kopiowanie nie obejmuje tylko danych, które są składowane na serwerze, ale często mamy do czynienia z kilkoma serwerami, z różnymi systemami operacyjnymi, licznymi stacjami roboczymi oraz komputerami przenośnymi, tabletami czy smartfonami.
- Często istnieje także potrzeba kopiowania „w locie” danych z otwartych baz danych np. SQL , MySQL, Oracle, przetwarzanych plików
- Im większa firma tym trudniej zaplanować, wdrożyć i realizować procedury.
- Tworzenie kopii bezpieczeństwa to proces długotrwały, praktycznie ciągły (problem zaniechania, zaniedbania, braku procedur, ale też zmian w technologii informatycznej - **dark data**) .

Rozwiązanie

- Jeśli nie masz żadnej kopii bezpieczeństwa danych - wykonaj natychmiast backup (jakikolwiek i gdziekolwiek (!))
- Wykonaj audyt bezpieczeństwa w swojej firmie
- Zaplanuj wykonywanie kopii bezpieczeństwa
- Oszacuj potrzeby
- Policz koszty
- Wybierz najlepsze dla twojej firmy rozwiązanie
- Opracuj, wdróż i przestrzegaj procedur (element polityki bezpieczeństwa danych elektronicznych)

- Jeśli nie wiesz co zrobić – zapytaj fachowca!

Jak zaplanować?

- Które dane wymagają archiwizacji?
- Ile jest tych danych?
- Czy dane są „tajne”?
- Ile mamy komputerów, serwerów, …
- Ile danych możesz stracić? (RPO (Recovery Point Objective) - określa moment w przeszłości, w którym po raz ostatni została wykonana kopia danych i do którego momentu działalności będzie można wrócić. Czyli ile możemy odtworzyć z wydruków, pamięci pracownika?)
- Ile możemy „wytrzymać” bez danych/systemu) RTO (Recovery Time Objective) - określa maksymalny czas po awarii potrzebny do przywrócenia działania aplikacji, systemów i procesów biznesowych.
- Kto będzie odpowiedzialny za wykonywanie kopii bezpieczeństwa?
- Koszty …

Jak zaplanować?

RPO i RTO pozwolą zdecydować czy potrzebujemy także kopii w czasie rzeczywistym, czy wystarczy wykonywanie kopii bezpieczeństwa okresowo na zewnętrznych nośnikach.

Aby zabezpieczyć system w czasie rzeczywistym i uchronić się np. przed awarią dysku twardego, należy zastosować technikę RAID (Redundant Array of Inexpensive Discs) – np. Mirroring. Często obecnie realizacja poprzez tzw. macierze dyskowe, ewentualnie zaplanować dyski wymienialne „na gorąco” (hot swap)

Warto wykorzystać nowe technologie i możliwości Internetu/sieci. (chmura, zewnętrzne firmy, specjalne oprogramowanie,…)

Backup tradycyjny - pamięci masowe - przykłady urządzeń

Do grupy urządzeń umożliwiających backup danych na nośnikach zewnętrznych należą:

1. Napęd taśmowy (streamer) i jego rozbudowane wersje:

a) Zmieniacz taśmowy (autoloader) – prosta automatyzacja

b) Biblioteka taśmowa – kilka magazynków, robotyka

2. Dyski twarde (zewnętrzne)
3. PenDrive
4. Napędy optyczne (CD/DVD/…)
5. Macierze dyskowe
6. NAS (Network Attached Storage) - metoda podłączenia urządzeń pamięci masowych bezpośrednio do sieci lokalnej/ internetowej

Pamięci masowe - klasyfikacja urządzeń

Do grupy urządzeń umożliwiających rejestrację (czyli zapis i odczyt) informacji na nośnikach zewnętrznych należą:

1. Napęd taśmowy (streamer) - urządzenie umożliwiające zapis i odczyt danych na taśmie magnetycznej. Charakteryzuje się z reguły długim czasem dostępu, ale za to dość szybkim transferem. Wykorzystywany głównie do wykonywania backupu. Taśmy magnetyczne są obecnie najbardziej pojemnymi nośnikami.
2. Zmieniacz taśmowy (autoloader) - urządzenie wyposażone w jeden napęd taśmowy, magazynek na kilka taśm i prosty automat do zmiany taśm. To najtańsza forma zautomatyzowania procesu backupu.
3. Biblioteka taśmowa - urządzenie z napędami taśmowymi (jednym lub więcej), zaawansowaną robotyką obsługującą co najmniej jeden magazynek taśm, zazwyczaj wyposażone w czytnik kodów kreskowych. Bibliotekę, w przeciwieństwie do zmieniacza, można rozbudowywać o kolejne napędy lub magazynki (w zależności od modelu).

4. Biblioteka optyczna/magnetooptyczna - urządzenie z napędami i nośnikami optycznymi lub magnetooptycznymi służące do archiwizacji i udostępniania danych. Może być wyposażone w kilka napędów i magazynków (w zależności od modelu). Cenione za swoją pojemność i szybki dostęp do danych.

5. Macierz dyskowa - urządzenie wyposażone w kilka dysków twardych służące do zapisu na nich danych modyfikowanych na bieżąco. Dyski są połączone w tzw. Macierz RAID (zwykle poziomu 1 lub 5), zwiększającą bezpieczeństwo zapisanych danych poprzez wykonywanie ich lustrzanej kopii (w RAID 1) lub duplikowanie wg specjalnego algorytmu (w RAID 5).

6. NAS (Network Attached Storage) - metoda podłączenia urządzeń pamięci masowych bezpośrednio do sieci lokalnej. Mogą to być pojedyncze napędy optyczne lub dyski twarde ze specjalnym kontrolerem NAS. Jako NAS można też podłączać wiele macierzy dyskowych.

Wymienny dysk twardy (lub drugi dysk w komputerze – w ostateczności!!!), pendrive

- Duża szybkość zapisu
- Widziany przez OS jako kolejny napęd
- Duża pojemność
- Mała awaryjność
- Relatywnie niska cena
- Kłopoty z organizacją tworzenia kopii (osobne nośniki na osobne dni)
- „Czułość” na wstrząsy, uszkodzenia mechaniczne.

Streamer

- Streamer, czyli napęd taśmowy to najczęściej stosowane urządzenie do tworzenia kopii zapasowych. Za pośrednictwem streamera dane zapisywane są na taśmach magnetycznych.
- Dostęp do danych zapisanych na taśmie możliwy jest za pośrednictwem odpowiedniego oprogramowania.
- Zaletą zastosowania streamera jest duża pojemność taśmy magnetycznej, co umożliwiająca wykonanie pełnej kopii wszystkich dysków serwera.
- Sekwencyjny dostęp do danych – taśmę należy przewinąć

Napęd CD/DVD-RW / BlueRay

- Umożliwia jedno/wielokrotny zapis i odczyt na dyskach CD/DVD-RW.
- Dyski CD/DVD-RW nadają się także do długotrwałego przechowywania danych.
- Specjalne oprogramowanie.
- Duża prędkość odczytu
- Stosunkowo mała prędkość zapisu (4x, 8x, ...)
- Wygodne (np. do przesyłania, przenoszenia danych)
- Widziane jako osobne dyski

Co wybrać?

Wybór konkretnego rozwiązania pamięci masowych jest determinowany przez rodzaj zabezpieczanych danych oraz czasy RPO i RTO.

W kilkunastoosobowej firmie, gdzie dane w niewielkich ilościach są trzymane na jednym serwerze, wystarczy zwykły wewnętrzny napęd taśmowy obsługiwany przez administratora. Dane księgowe może np. archiwizować księgowa na wbudowanej w jej komputer nagrywarce CD (ilość tych danych z reguły nie przekracza kilkuset MB).

W sieciach z kilkudziesięcioma komputerami, gdzie funkcjonuje kilka serwerów, warto pomyśleć o instalacji zmieniacza taśmowego. Urządzenie to przejmuje odpowiedzialność za sterowanie dostarczaniem taśm do napędu taśmowego, ułatwia wykonanie backupu w przypadku, gdy nie mieści się on na jednej taśmie. Eliminuje też ryzyko wystąpienia błędu człowieka np. Przy wybieraniu odpowiedniej taśmy podczas wykonywania backupu.

Największe rozwiązania z reguły są oparte na bibliotekach taśmowych, optycznych i dużych macierzach dyskowych. Są drogie, ale umożliwiają zgromadzenie ogromnych ilości danych i zapewniają szybki do nich dostęp.

RPO i RTO

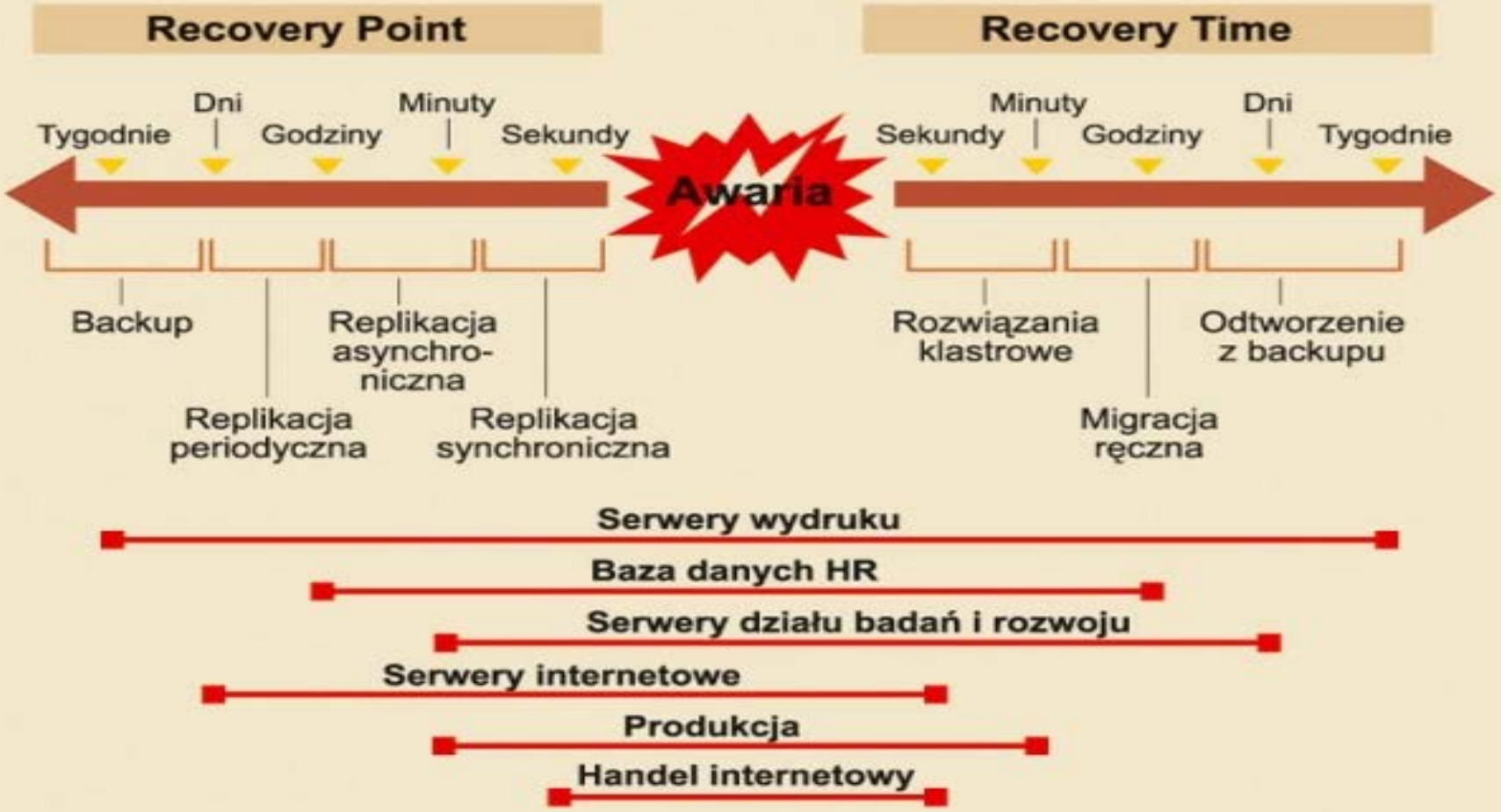
RPO (Recovery Point Objective) - określa moment w przeszłości, w którym po raz ostatni została wykonana kopia danych i do którego momentu działalności będzie można wrócić. (ile możemy stracić – w szczególności ile możemy odtworzyć?)

Czas ten jest różny i zależy głównie od charakteru działalności - jednym wystarczy kopia sprzed tygodnia, inni zaś potrzebują danych sprzed kilkunastu sekund.

RTO (Recovery Time Objective) - określa maksymalny czas po awarii potrzebny do przywrócenia działania aplikacji, systemów i procesów biznesowych. (ile możemy wytrzymać bez danych/systemu)

Określając parametr RTO, należy doprowadzić do kompromisu między potencjalnymi stratami a kosztami rozwiązania umożliwiającego jak najszybsze odtworzenie stanu sprzed awarii.

RPO i RTO



Recovery Point Objective określa moment w przeszłości, w którym po raz ostatni została wykonana kopia danych i do którego momentu naszej działalności będziemy mogli wrócić

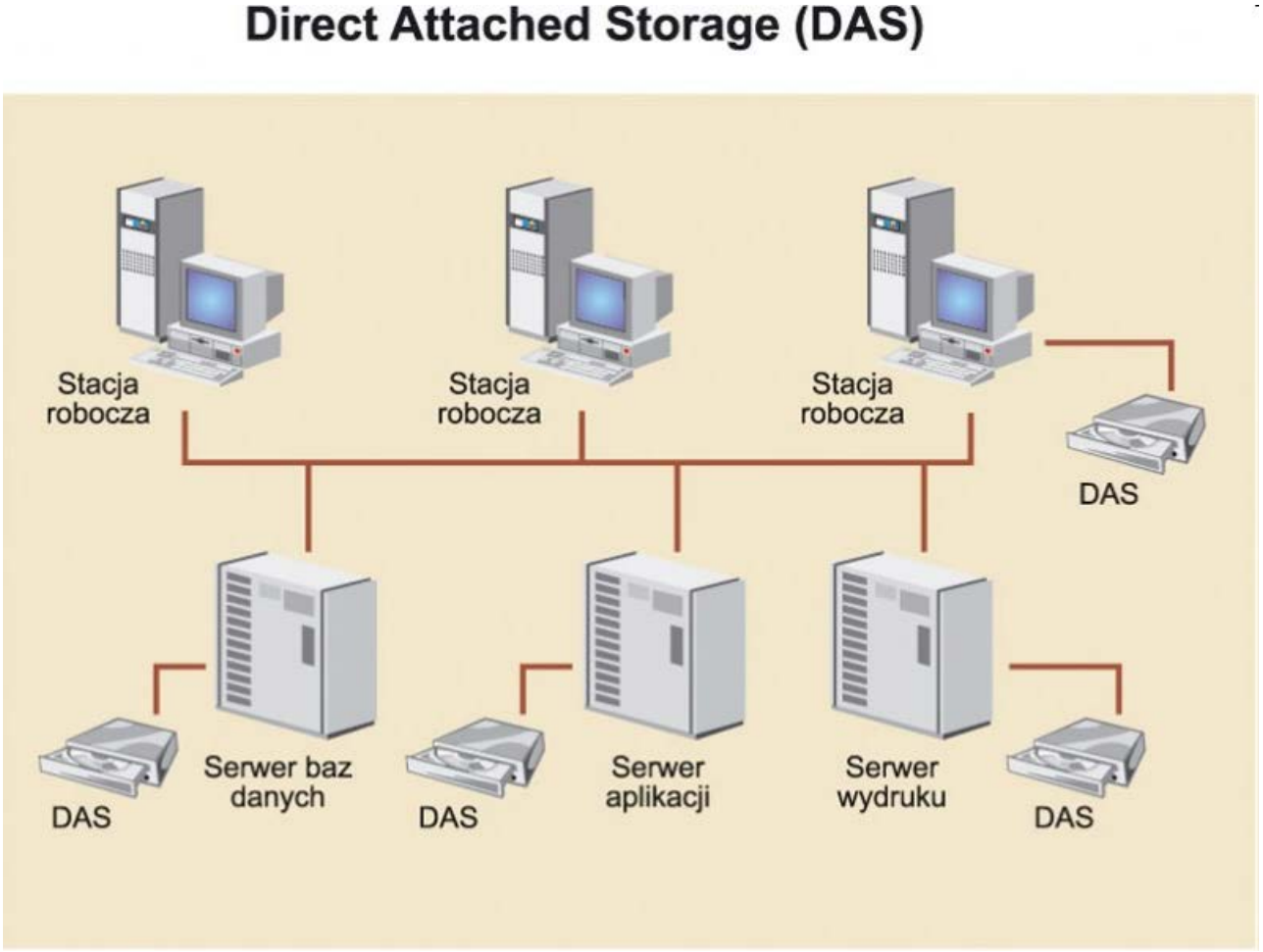
Recovery Time Objective określa maksymalny czas po awarii potrzebny do przywrócenia działania aplikacji, systemów i procesów biznesowych

Direct Attached Storage (DAS).

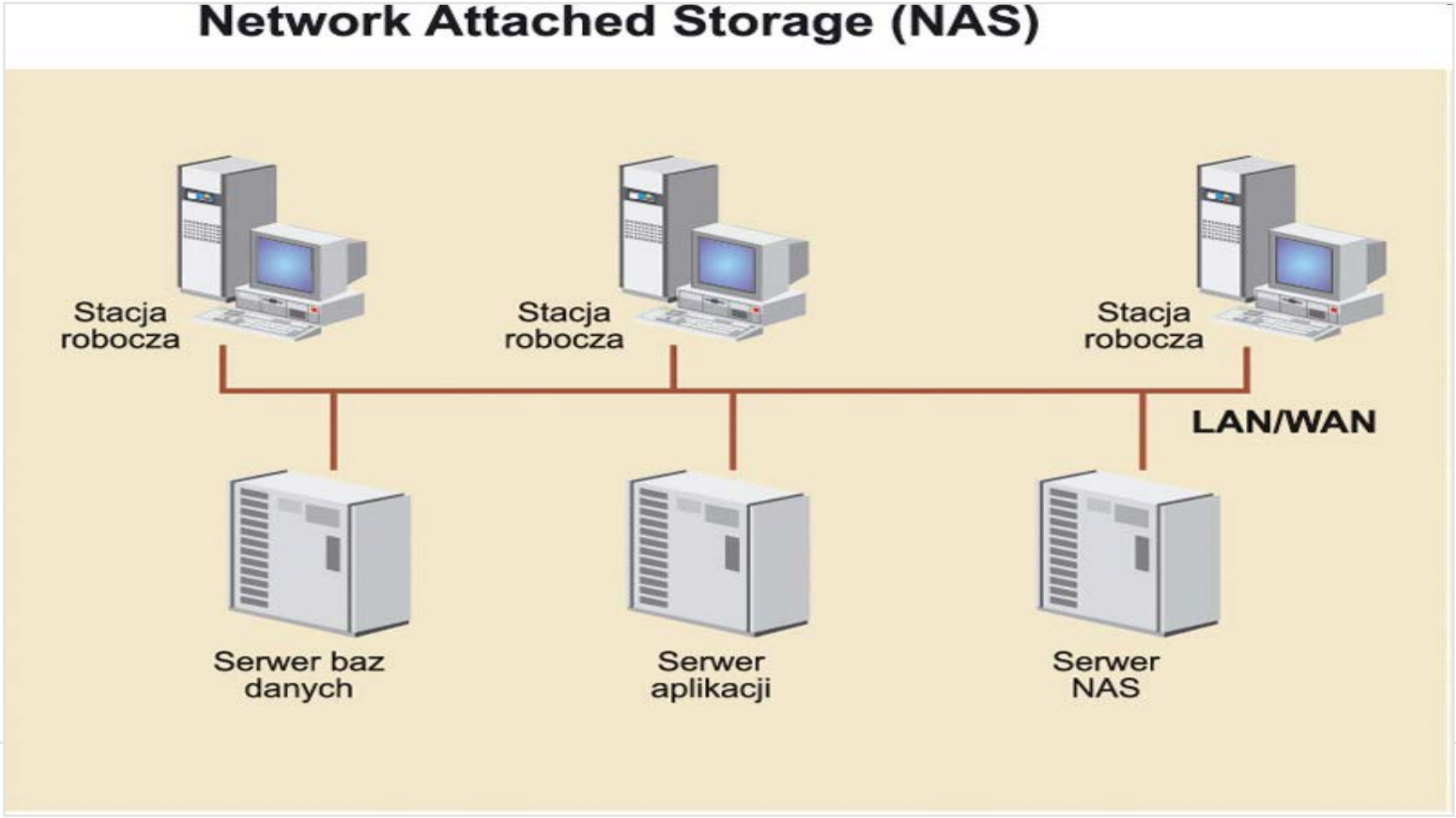
Network Attached Storage (NAS),

Storage Area Network (SAN).

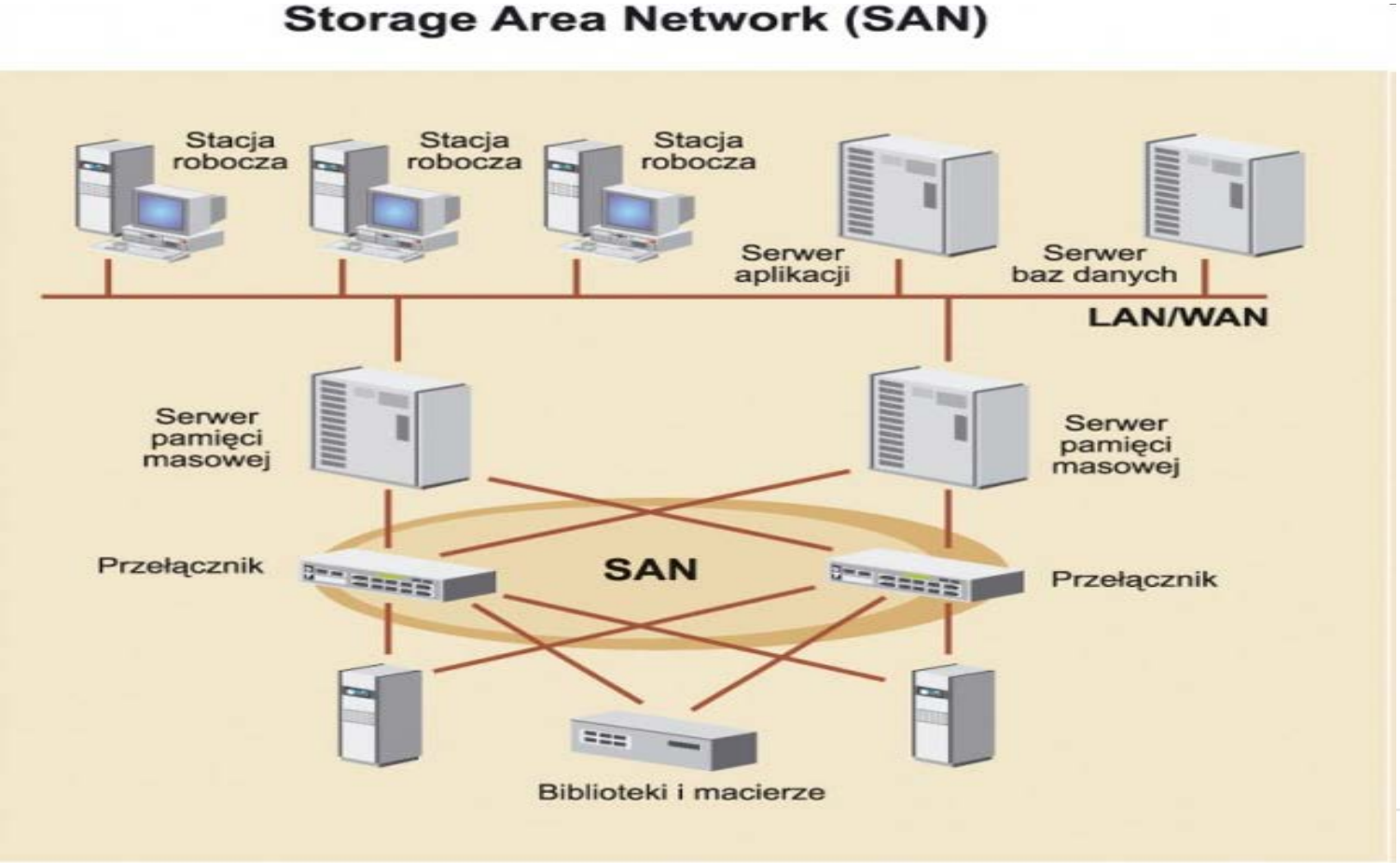
podłączanie pamięci masowej bezpośrednio do gromadzącego dane komputera, czyli Direct Attached Storage (DAS). Taka decentralizacja jest o tyle niekorzystna, że w praktyce uniemożliwia efektywne zarządzanie zasobami pamięci masowych.



Network Attached Storage (NAS), czyli urządzenie pamięci masowych (macierz dyskowa lub serwer pamięci optycznych) podłączone bezpośrednio do sieci Ethernet, które pozyskuje i udostępnia swoje zasoby wskazanym przez administratora użytkownikom. Dyski w podłączonej w ten sposób macierzy dyskowej mogą być połączone np. w RAID poziomu 5,



Storage Area Network (SAN). Umożliwiają podłączenie praktycznie nieograniczonej liczby pamięci masowych. Na sieć SAN, oprócz urządzeń pamięci masowych, składa się także cała dodatkowa infrastruktura (przełączniki, serwery itp.). Trudniejsze jest także zarządzenie całą siecią. Jej zastosowanie jest jednak konieczne w momencie, gdy ilość przetwarzanych i archiwizowanych danych przekracza setki gigabajtów.



Co archiwizować?

1. Kopia danych
2. Kopia systemu

Kopia danych

- Wykorzystywanie narzędzi wbudowanych w oprogramowaniu użytkowym.
- Stworzenie własnych narzędzi (przy wykorzystaniu np. możliwości systemu operacyjnego)
- Kompresja danych – programy kompresujące (pakujące)

  Pkzip, winzip, rar, winrar, arj, tar (Unix)
  - W zależności od rodzaju danych nawet kilku, lub kilkunastokrotna kompresja danych
  - Większa szybkość przesyłania
  - Możliwość dzielenia (np. w celu przenoszenia) zbiorów na mniejsze jednostki
  - Dodatkowy czas potrzebny na archiwizacje i testowanie
  - Większa awaryjność skompresowanych zbiorów.

Kopia systemu

Takie zabezpieczenie danych (nie tylko użytkowych), aby w razie awarii np. serwera, lub systemu sieciowego można było jak najszybciej rozpocząć pracę.

- Konta użytkowników
- Pliki konfiguracyjne
- Ustawienia systemu i sieci
- Struktura katalogów, danych
- Narzędzia do serwisowania
- Źródła oprogramowania, systemu

Wykonywana codziennie, lub po wprowadzeniu zmian w elementach systemu (nowe ustawienia, użytkownicy)

Konieczne specjalistyczne oprogramowanie do tworzenia „backup'u".

Odpowiednia konfiguracja oprogramowania

Możliwość i umiejętność odtworzenia danych w razie wystąpienia awarii, niezależność od systemu.

Backupy są najczęściej wykonywane wg ściśle określonych schematów rotacji nośników (np. „Dziadek/Ojciec/Syn” czy „Wieże Hanoi”). Określają, w których dniach tygodnia jest wykonywany tzw. Backup pełny (wówczas są zapisywane wszystkie, konieczne do zabezpieczenia dane), a kiedy backup przyrostowy lub różnicowy (wtedy są zapisywane tylko te dane, które pojawiły się lub się zmieniły w ostatnim czasie).

Oszczędność czasu, pieniędzy i nośników, które się zużywają

Rodzaje backup'ów

a)pełny, (full backup) polegającą na tworzeniu kopii zapasowych wszystkich plików w systemie

+ łatwość wyszukiwania dowolnych danych (*wszystkie znajdują się na jednym nośniku*)
+ Odtworzenie systemu można przeprowadzić bardzo szybko
- Nieefektywne wykorzystanie nośników - cały czas są backupowane dane rzadko ulegające zmianom
- Długi czas wykonywania operacji

b) różnicowy, (differential backup) składowanie tylko tych plików, które uległy zmianie od czasu ostatniej pełnej archiwizacji danych

+ łatwy sposób wyszukiwania dowolnych danych (*do odnalezienia dowolnego zbioru potrzebne są maksymalnie dwa nośniki*)
+ Odtworzenie systemu przeprowadza się stosunkowo szybko
+ Czas przeprowadzenia backupu dużo krótszy niż w przypadku backupu całościowego
- Nieefektywne wykorzystanie nośników; nadmiarowość backupów, dane które nie uległy zmianie, są cały czas backupowane
- Dłuższy czas wykonywania operacji niż backupów przyrostowych

c) przyrostowy, (incremental backup) Zapisywane są tylko te dane, które powstały lub uległy zmianie od czasu przeprowadzenia ostatniego całościowego lub przyrostowego backupu;

+ Czas przeprowadzenia backupu jest bardzo krótki
+ Efektywne wykorzystanie nośników
- Trudność wyszukiwania danych (*do odnalezienia zbioru są potrzebne wszystkie nośniki z backupami przyrostowymi oraz ostatni nośnik z backupem całościowym*)
- Długi czas odtworzenia systemu

Problem – jak określić, które pliki uległy zmianie?

- sprawdzenie daty aktualizacji pliku (musimy znać datę poprzedniej archiwizacji)
- porównywanie plików (zawartości)
- bit archiwizacji (A- archive) – zmiana tego bitu (atrybutu) po wykonaniu kopii (opcjonalnie)

Schematy rotacji nośników

Nazwa	Liczba kaset	Horyzont backup'u
Syn	1	1 dzień
Ojciec/Syn	6	2 tygodnie
Dziadek/Ojciec/Syn	19	1 rok
Wieże Hanoi	4	15 dni(4 nośniki) N2-1

horyzont backupu to określenie liczby dni sprzed backupu, z których można odzyskać dane. Horyzont backupu zależy od liczby wykorzystanych nośników w danym schemacie.

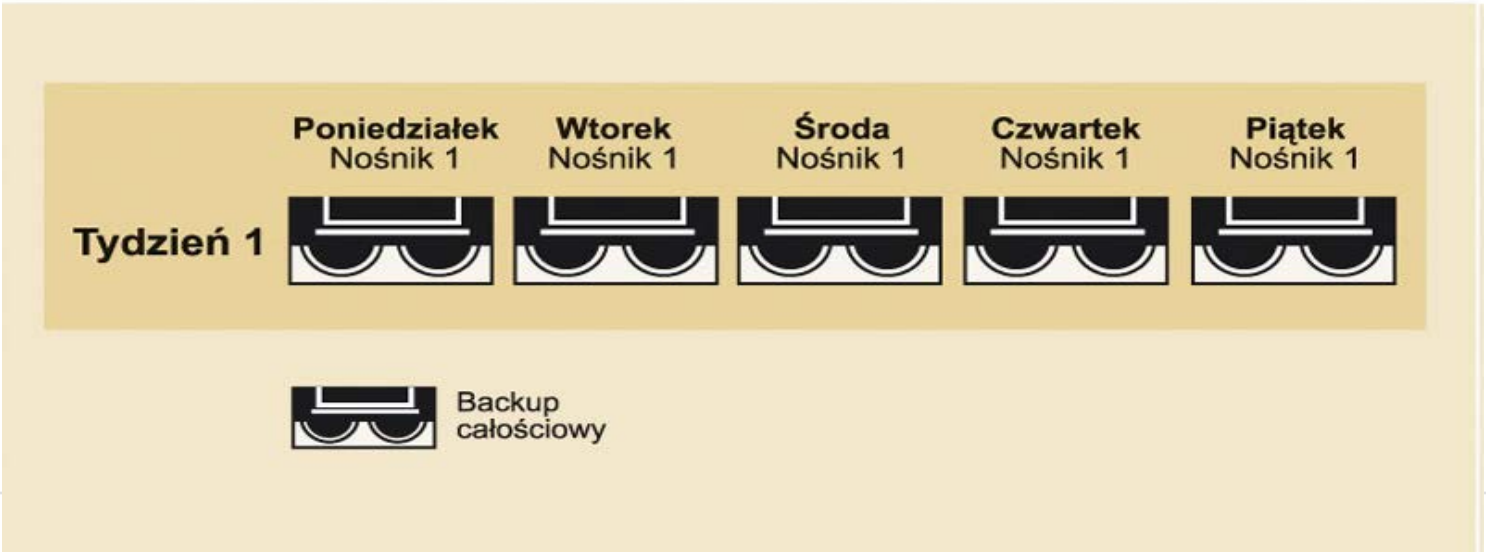
Rotacja typu „Syn”

Wymagana minimalna liczba nośników: 1.

Horyzont backupu: 1 dzień.

Do zaimplementowania tego schematu jest używany tylko jeden nośnik - codziennie jest wykonywany pełny backup. Główną wadą tego rozwiązania jest nadmierne zużycie nośnika. W razie awarii brak jest możliwości odtworzenia jakiejkolwiek kopii danych.

Schemat rotacji typu “Syn”



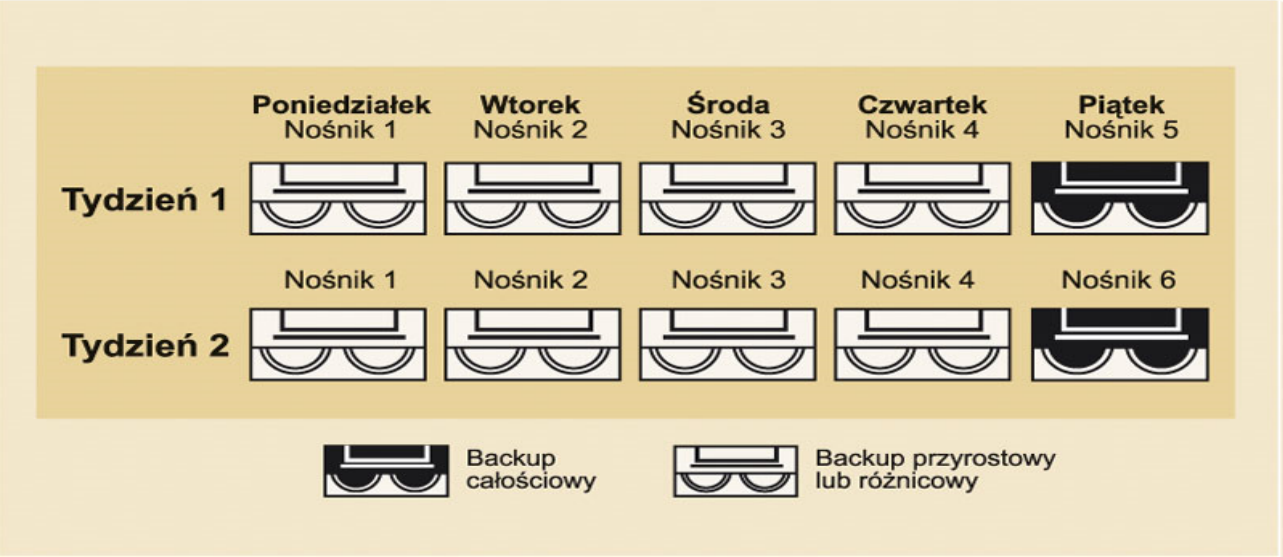
Rotacja typu „Ojciec/Syn”

Wymagana minimalna liczba nośników: 6.

Horyzont backupu: dwa tygodnie.

Jest to kombinacja backupów całościowych i różnicowych/przyrostowych. Od poniedziałku do czwartku jest wykonywany backup różnicowy lub przyrostowy, a w piątek - backup pełny. W drugim tygodniu znów są wykorzystywane nośniki od 1 do 4, a nośnik 5 (spełniający rolę „ojca”) przechowuje wyniki pracy z całego tygodnia. Backupy piątkowe (nośniki 5 i 6) przechowywane są poza główną siedzibą firmy (off line safety copy). Schemat łatwy w implementacji

Schemat rotacji typu “Ojciec/Syn”



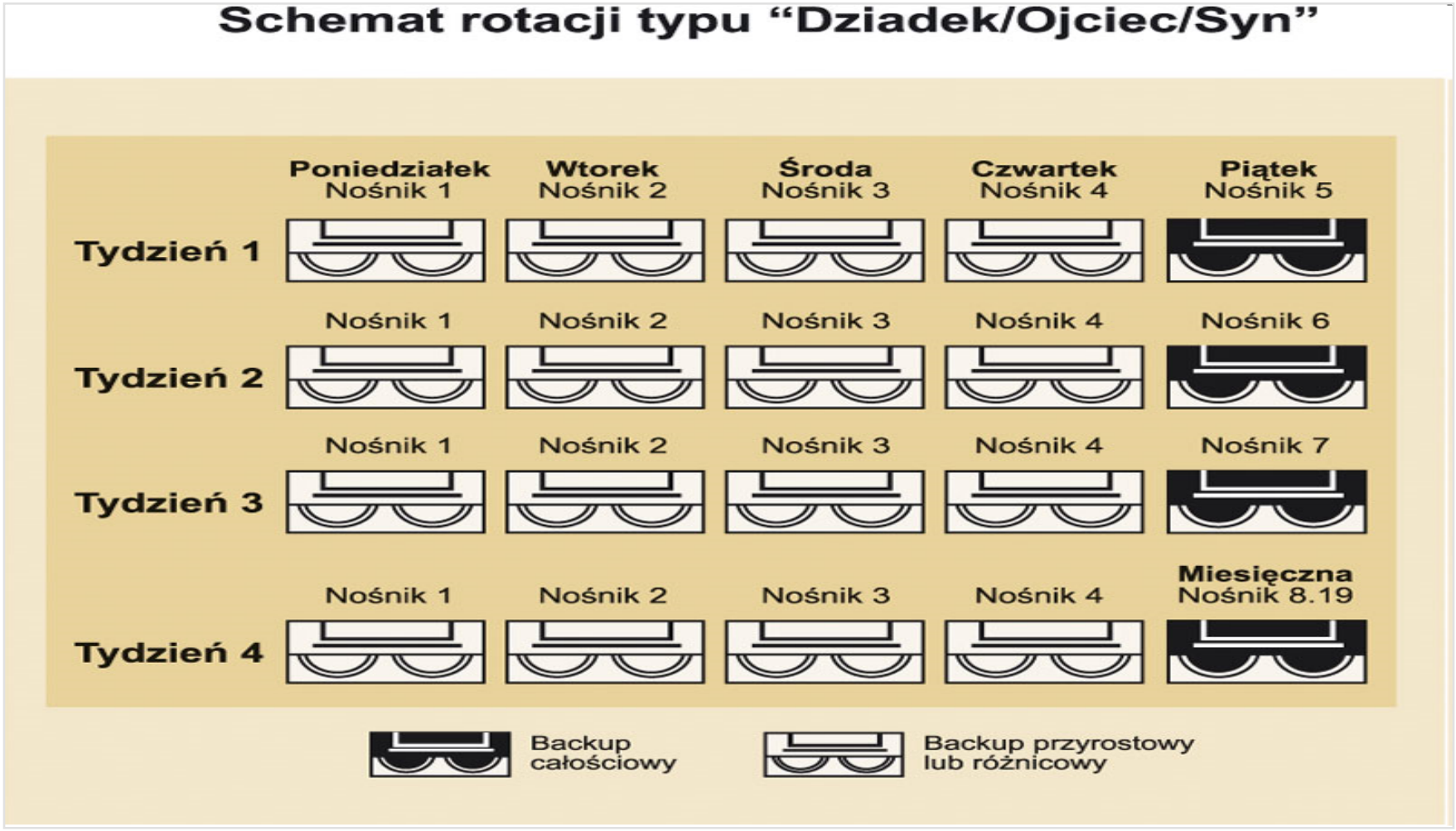
Rotacja typu „Dziadek/Ojciec/Syn”

Wymagana minimalna liczba nośników: 19.

Horyzont backupu: rok.

To najbardziej popularny i powszechnie stosowany schemat. Różnicę w stosunku do poprzednio omówionego schematu stanowi wprowadzenie nośnika, który będzie przechowywał kopię z miesięcznym backupem danych - nośnik ten spełnia rolę „dziadka”. Podobnie jak w schemacie „Ojciec/Syn” od poniedziałku do czwartku są wykonywane backupy przyrostowe lub różnicowe na nośnikach od 1 do 4, w piątek zaś - backupy pełne. Po czterech tygodniach ostatni pełny backup (miesięczny) jest wywożony z firmy, a pozostałe nośniki wracają „na początek” algorytmu. Schemat ten charakteryzuje się dobrym stosunkiem czasu wykorzystania nośników do ich liczby (19 na rok).

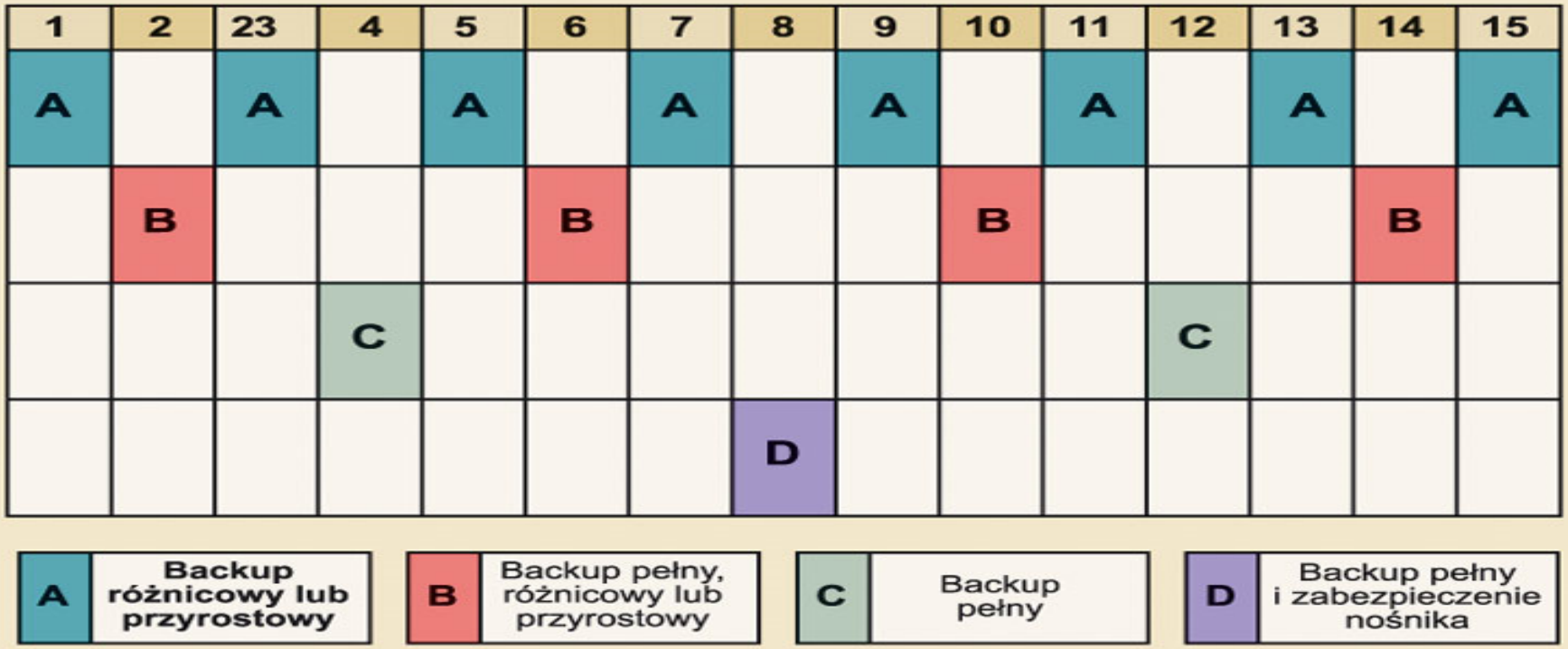
Rotacja typu „Dziadek/Ojciec/Syn”



Wieże Hanoi

- Wymagana minimalna liczba nośników: 4, zależna od przyjętej strategii.
- Horyzont backupu: 15 dni dla 4 nośników

Schemat rotacji typu: Wieże Hanoi



Zabezpieczenie danych w czasie rzeczywistym

Replikacja synchroniczna to zabezpieczanie danych w czasie rzeczywistym. Już podczas tworzenia lub modyfikowania danych jest wykonywana ich kopia, najczęściej w oddalonym o kilka kilometrów zapasowym centrum przetwarzania danych. Awaria serwera głównego od razu pozwala na kontynuowanie działalności, korzystając z danych zapasowych.

Możliwość wykonania replikacji synchronicznej jest ograniczona ze względu na odległość między centrum podstawowym a zapasowym. Dlatego wprowadzono możliwość wykonania replikacji asynchronicznej, w której dane są kopiowane z pewnym, z reguły kilkunastominutowym opóźnieniem.

Pewną formą replikacji asynchronicznej jest replikacja periodyczna. Tu też dane są kopiowane z opóźnieniem, ale o zdefiniowanych z góry porach.

RAID

Rozwiązanie backupu wraz z oprogramowaniem zabezpiecza system w trybie offline. W razie awarii można przywrócić dane dokładnie takie jak podczas backupu.

Aby zabezpieczyć system w czasie rzeczywistym i uchronić się przed awarią dysku twardego, należy zastosować technikę RAID (Redundant Array Inexpensive Discs).

W niewielkich stosunkowo systemach można wykorzystywać kontrolery wewnętrzne RAID SCSI czy choćby RAID IDE, które choć tańsze są mniej wydajne i bezpieczne niż SCSI.

Średnie firmy powinny zbudować system wykorzystujący dyski wymienialne „na gorąco” (hot swap) w zewnętrznej obudowie lub kupić macierz z zewnętrznym kontrolerem RAID. To ostatnie pozwala uniezależnić działanie macierzy od awarii serwera.

Przy wyborze obudowy RAID czy macierzy należy zwrócić uwagę na odpowiedni system wentylacji, monitoringu temperatury wewnątrz i zdalnego zarządzania rozwiązaniem. Ważne są w tym przypadku opcje serwisowe, takie jak czas gwarancji oraz możliwość podpisania umowy na szybką naprawę urządzenia, np. w ciągu 24 godz.

RAID Level 0 – nie stosować do backupu!!

Polega na połączeniu ze sobą dwóch lub więcej dysków fizycznych tak, aby były widziane jako jeden dysk logiczny. Powstała w ten sposób przestrzeń ma rozmiar taki jak suma rozmiarów wszystkich dysków. Dane są przeplecione pomiędzy dyskami. Dzięki temu uzyskujemy znaczne przyśpieszenie operacji zapisu i odczytu ze względu na zrównoleglenie tych operacji na wszystkie dyski w macierzy. Warunkiem uzyskania takiego przyśpieszenia jest operowanie na blokach danych lub sekwencjach bloków danych większych niż pojedynczy blok danych macierzy RAID 0 - ang. stripe unit size.

Korzyści:

* przestrzeń wszystkich dysków jest widziana jako całość
* przyspieszenie zapisu i odczytu w porównaniu do pojedynczego dysku

Wady:

* brak odporności na awarię dysków

W razie awarii jednego napędu wszystkie dane są tracone.

Zwiększenie szybkości widoczne jest przede wszystkim w wypadku dużych, wzajemnie powiązanych plików.

RAID Level 1

Korekcja błędów za pomocą parzystości

RAID Level 1 bywa też nazywany mirroringiem, czyli tworzeniem lustrzanych kopii dysków.

Ta nazwa odzwierciedla sposób działania - wszystkie zapisy odbywają się równolegle na dwóch lub więcej napędach i dlatego każdy dysk stanowi lustrzane odbicie drugiego. Wszystkie dane są dublowane, co zapewnia bardzo wysoki stopień bezpieczeństwa. Nawet w razie całkowitej awarii jednego dysku, pozostają zapisy na drugim/kolejnych.

Korzyści:

* odporność na awarię N - 1 dysków przy N-dyskowej macierzy

Wady:

* utrata pojemności (dokładnie pojemności N - 1 dysków)

RAID Level 5

Macierz składa się z 3 lub więcej dysków. Przy macierzy liczącej N dysków jej objętość wynosi N - 1 dysków.

Sumy kontrolne danych dzielone są na N części, przy czym każda część składowana jest na innym dysku a wyliczana jest z odpowiedniego fragmentu danych składowanych na pozostałych N-1 dyskach.

Korzyści:

* odporność na awarię 1 dysku
* zwiększona szybkość odczytu - porównywalna do macierzy RAID0 złożonej z N-1 dysków

Wady:

* zmniejszona szybkość zapisu z powodu konieczności kalkulowania sum kontrolnych (eliminowana poprzez zastosowanie sprzętowego kontrolera RAID5)
* w przypadku awarii dysku dostęp do danych jest spowolniony z powodu obliczeń sum kontrolnych
* odbudowa macierzy po wymianie dysku jest operacją kosztowną obliczeniowo i powoduje spowolnienie operacji odczytu i zapisu

RAID - programowy

W wypadku programowej macierzy RAID sterowaniem zespołu napędów zajmuje się oprogramowanie działające na procesorze hosta. Niektóre systemy operacyjne są wyposażone w potrzebne składniki. Windows NT obsługuje na przykład RAID 0 oraz RAID 1 i 5 - ten ostatni poziom tylko w wersji serwerowej. Linux obsługuje generalnie poziomy 0, 1, 4 i 5.

Programowa macierz RAID jest więc często najtańszym i najprostszym rozwiązaniem. Można je także łatwo dostosować do rosnących wymagań poprzez wymianę procesora na wydajniejszy. Programowa macierz RAID powoduje jednak duże obciążenie procesora i siłą rzeczy jest związana z platformą i systemem operacyjnym. Poza tym w zwykłych warunkach do dyspozycji jest tylko jedno lub dwa przyłącza do sterowania napędów, co w znacznym stopniu ogranicza możliwości równoległych odwołań do napędów, a zatem i wydajność.

RAID sprzętowy.

Odmienne rozwiązanie to sprzętowa macierz RAID - sterowanie napędów przejmuje wyspecjalizowany kontroler. Korzyść polega na odciążeniu procesora i wzroście wydajności. Dodatkowo kontroler RAID odwołuje się do dysków za pośrednictwem wielu kanałów, co umożliwia odwołania równoległe i zwiększa szerokość pasma.

Jak zwykle w takich wypadkach, rozwiązanie to ma swoją - wyższą - cenę. Sprzętowe macierze RAID pracują w sposób niezależny od platformy, potrzebują jednak również oprogramowania, które musi być dostosowane do systemu operacyjnego.

Pełen system ochrony danych

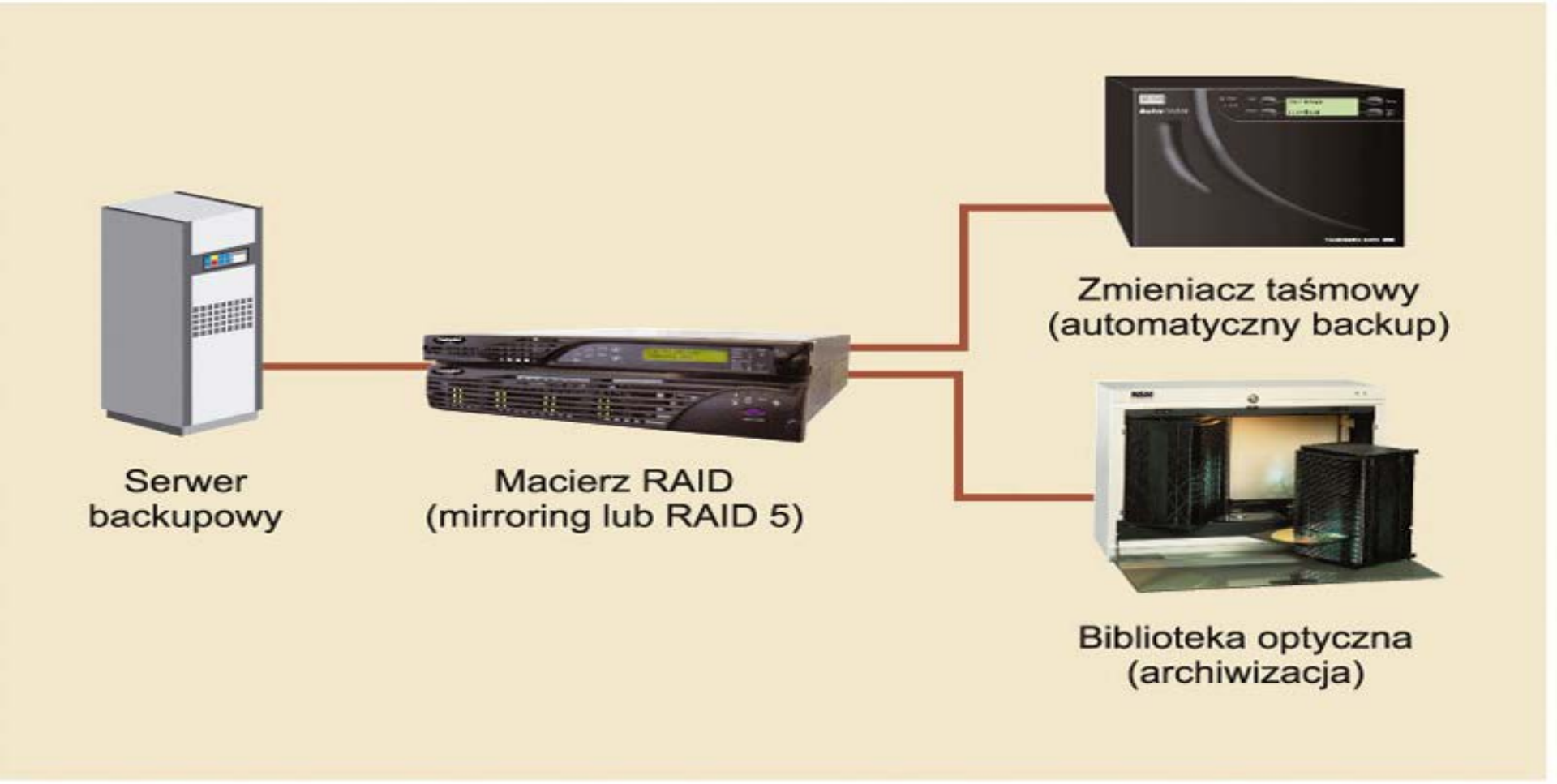
Żaden system mirroringu czy replikacji nie zastąpi backupu. Tylko backup pozwala na spojrzenie wstecz i powrót do wersji danych sprzed określonego czasu. Daje to zabezpieczenie głównie przed wirusami lub błędami ukrytymi w programie bądź strukturze danych. Systemy redundancyjne, przy całej swojej doskonałości, naiwnie duplikują także wirusy czy błędy.

Na pełen system ochrony danych w systemach pamięci masowych składa się kilka elementów:

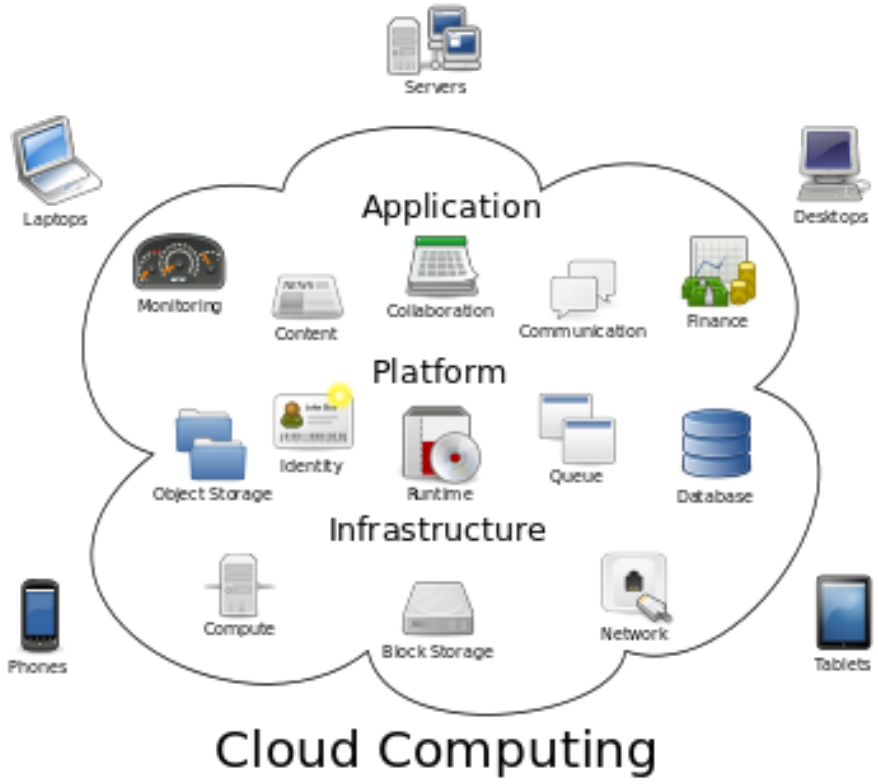
1. macierz RAID zabezpiecza system przed awarią twardego dysku;
2. urządzenie zapisu taśmowego chroni system przed ogólną awarią (w tym całkowitą awarią macierzy RAID), klęską żywiołową (np. powódź, pożar), a także umożliwia archiwizację danych; dzięki przechowywaniu backupów z różnych dni możemy odtworzyć stan np. sprzed awarii, która spowodowała niespójność danych;
3. dane, które nie podlegają zmianom i są rzadko używane, powinny być archiwizowane na urządzeniach optycznych lub magnetooptycznych, których nośniki są obecnie najtańsze w przeliczeniu na 1 GB.

Na system zabezpieczania danych powinna składać się odpowiednia ochrona antywirusowa (z mechanizmem automatycznej aktualizacji kodów wirusów) i elektryczna.

W pełni zabezpieczony system backupu i archiwizacji danych



Nowe trendy - backup w chmurze lub wykupienie usługi automatycznego tworzenia kopii zapasowej



Uwaga na „tajność” danych i wymagania prawne

Podsumowanie: co wybrać?

- Wybór konkretnego rozwiązania wykonywania kopii wymaga każdorazowo analizy konkretnego przypadku. (Wiele czynników ma wpływ na decyzję)
- Podstawa to zauważenie problemu i podjęcie kroków w celu jego rozwiązania.
- Pamiętajmy, że wykonywanie backup'u to proces długotrwały.
- Backup to element polityki bezpieczeństwa danych elektronicznych w firmie
- Konieczne procedury – nie tylko archiwizacji, ale też odtwarzania danych i postępowania w razie awarii
- Automatyzacja procesu
- Inwestować w dobre urządzenia i dobrej jakości nośniki danych.
- Dostęp do kopii bezpieczeństwa ograniczony tylko dla uprawnionych osób.
- Uwaga na administratorów i informatyków – duże uprawnienia – duże zniszczenia
- Oczywiście lepsza jest jakakolwiek kopia niż jej całkowity brak.

Kilka ogólnych uwag o sposobie archiwizacji:

- Archiwizacja powinna być dokonywana na przenośnym nośniku danych (taśmy magnetyczne, przenośne dyski twarde), tak, aby można było przechowywać je w innym pomieszczeniu np. w sejfie.

- Nie należy dopuszczać do sytuacji, gdy jedyna kopia danych znajduje się na tym samym nośniku co dane oryginalne (na tym samym dysku).

- Nie należy używać jednego nośnika do sporządzania kolejnych kopii danych.

- Odpowiednia organizacja tworzenia archiwum – pn, wt, .., sobota, tydzień, miesiąc, ...

- Kopię danych należy sporządzać możliwie najczęściej - najlepiej codziennie.

- Oprócz kopii bieżących, dobrze jest sporządzać w pewnych okresach np. co miesiąc lub przynajmniej przy zamknięciu roku osobne kopie danych i przechowywać je w innym pomieszczeniu.

- Co pewien czas należy sprawdzić czy kopia danych wykonana jest poprawnie. Najlepiej przez próbne odtworzenie bazy danych na innym komputerze.

- Nigdy nie należy kasować starych danych archiwizacyjnych przed sporządzeniem nowej kopii danych.

- Należy sprawdzić swoje umiejętności i możliwości odtworzenia danych z archiwum.

- Zawsze wykonywać kopie systemu przed podjęciem czynności serwisowych

- Możliwe jak najbardziej zautomatyzować i ujednolicić proces archiwizacji.
- Przeszkolić personel w zakresie obsługi urządzeń (np. stramer, Zip) i oprogramowania do tworzenia archiwizacji
- Przygotować szczegółowe, jasne i czytelne instrukcje dla pracowników.
- Nigdy nie zaniechać wykonywania kopii zapasowych.
- Inwestować w dobre urządzenia i dobrej jakości nośniki danych.
- Pamiętać o czytelnym i jednoznacznym opisie nośników, plików, ..

- sprawdzenie poprawności algorytmów wykonywania kopii;
- przechowywanie i regularna rotacja zestawu kopii bezpieczeństwa poza siedzibą firmy - uchroni to dane przed skutkami lokalnych katastrof;
- dostęp do kopii bezpieczeństwa ograniczony tylko dla właściwych osób;
- regularne próby odtwarzania całego systemu na serwerze testowym;
- właściwy opis nośników z kopiami zapasowymi;
- określenie procedury postępowania w razie awarii napędu;
- przygotowanie i aktualizacja awaryjnego zestawu naprawczego zawierającego dysk startowy, oprogramowanie diagnostyczne, wersję instalacyjną systemu operacyjnego, wersję instalacyjną programu do zabezpieczania danych, aktualne wersje sterowników, odpowiednie narzędzia i zapasowe media.

Przykładowy skrypt do archiwizacji

```
#!/bin/bash

# Create a compressed backup of all the directories specified and put the
# resulting file in a directory of your choice.
BACKUP_DIRS="$HOME /etc /var"
BACKUP_FILENAME=`date '+%b%d%Y'`
BACKUP_DEST_DIR="/backups"

# Uncomment the following line for GZipped backups, comment for
# BZipped backups
#tar cvzf $BACKUP_DEST_DIR/$BACKUP_FILENAME.tar.gz $BACKUP_DIRS

# We do a BZipped backup here...
tar cvjf $BACKUP_DEST_DIR/$BACKUP_FILENAME.tar.bz2 $BACKUP_DIRS
```

Zaplanować archiwizację + koszty

Mała firma

- 10 komputerów stacjonarnych
- 3 przenośne
- LAN
- Internet
- Serwer plików Linux/Samba
- Aplikacja centralna + Office + pliki pracowników
- 2 GB danych – aplikacja
- Przyrost roczny ok. 1 GB
- Informatyk „dochodzący”

Co prawo mówi o archiwizacji

- Obowiązek tworzenia i zabezpieczania kopii zapasowych oraz ich przechowywania poza miejscem eksploatacji narzuca przedsiębiorcom ustawa o ochronie danych osobowych (art. 31).

- Z kolei ustawa z 29 września 1994 r. o rachunkowości stanowi, iż w przypadku prowadzenia ksiąg rachunkowych przy użyciu komputera ochrona danych (księgi handlowe, dokumentacja inwentaryzacyjna, sprawozdania finansowe)
  powinna polegać na stosowaniu odpornych na zagrożenia nośników danych, doborze stosownych środków ochrony zewnętrznej oraz systematycznym tworzeniu rezerwowych kopii bezpieczeństwa danych zapisanych na nośnikach komputerowych.

- Kolejnym warunkiem, który należy spełnić, jest zapewnienie trwałości zapisu informacji systemu rachunkowości przez czas nie krótszy od wymaganego do przechowywania ksiąg rachunkowych (tzn. przez 5 lat).

Ustawa nakazuje zadbać także o ochronę programów komputerowych i danych systemu informatycznego rachunkowości poprzez stosowanie zabezpieczeń programowych i organizacyjnych, chroniących przed nieupoważnionym dostępem lub zniszczeniem.

Owe zabezpieczenia organizacyjne to np. umieszczenie kopii danych w innym lokalu (np. poza miejscem prowadzenia działalności), powierzenie firmie przechowującej dane, ale także umieszczenie w skrytce bankowej.

Takie polecenie zawiera również rekomendacja Generalnego Inspektora Nadzoru Bankowego, dotycząca zarządzania ryzykiem towarzyszącym systemom informatycznym i telekomunikacyjnym używanym przez banki:
„Konieczne jest stosowanie kopii bezpieczeństwa (tzw. backup) i bieżącego dziennika zdarzeń (tzw.log), które w przypadku utraty danych powinny pozwolić na odtworzenie zasobów. Kompletne zapisy kopii zapasowych (backup) powinny być przechowywane oddzielnie, w odpowiednim oddaleniu od systemu informatycznego, dobrze zabezpieczone fizycznie i środowiskowo"