Archiwizacja danych

- element planowania kryzysowego
- zabezpieczenie przed mikro-awariami (przypadkowe skasowanie pliku/ów) i prawdziwymi katastrofami (awaria dysku, włamanie, pożar)
- wykonywana na mediach ładowalnych, tzn. takich, które wyciąga się z komputera w czasie pracy, co zabezpiecza przed takimi awariami, jak awaria kontrolera dysku (wyłączającego nie jeden ale wiele dysków), jak również przed jednoczesną utratą lub uszkodzeniem wszystkich dysków
- wymaga planowania:
  - cykl nagrywania i rotacja nośników
  - format danych i oprogramowanie
  - sposób przechowywania i zabezpieczenie
- wymaga testowania — nie wystarczy zaplanować i uruchomić ambitne procedury archiwizacji, a potem spać spokojnie; trzeba testować nie tylko możliwość odczytania poszczególnych taśm, ale pełne scenariusze awarii, i możliwość odbudowania i ponownego uruchomienia całego systemu lub centrum obliczeniowego

Założenia procedur archiwizacji

- dyski, partycje dyskowe krytyczne: muszą być archiwizowane
  - szybkozmienne: np. katalogi użytkowników
  - wolnozmienne: np. systemowe
  - inne, np. pojedyncze zmienione, lub zmieniające się pliki na ogólnie niezmiennej strukturze, np. plik /etc/passwd, lub pliki konfiguracji sieciowej na standardowej instalacji systemu
- dyski, partycje dyskowe niekrytyczne (niekoniecznie wymagające archiwizacji): /tmp, /var/log, /var/spool

Nośniki

- taśma magnetyczna
  - wady: wolny dostęp, podatna na uszkodzenia mechaniczne, magnetyczne, rozproszone promieniowanie elektromagnetyczne
  - czas zachowania danych ograniczony do ok. 2 lat
  - czas użytkowanie ograniczony do ok. 50 przebiegów (zapisu lub odczytu)
  - wysoka pojemność
  - historycznie najczęściej używane i nadal powszechnie stosowane
- inne nośniki magnetyczne (dyskietki, dyski wymienne)
- magneto-optyczne i optyczne
- nośniki jednorazowego nagrywania: WORM, CD-R

  Nowoczesne, wygodne i tanie (ale tylko w cenie za urządzenie, nie za megabajt). Zwykle mają mniejsze pojemności i są głównie stosowane do archiwizacji na mniejszą skalę, np. do zarchiwizowania dyzku prywatnego komputera.

Procedury archiwizacji

- archiwizacja plików systemowych
  - wymaga opracowania planu odbudowy systemu z taśm
    → taki plan musi zawierać szczegółowe komendy do wykonania w czasie odbudowy i musi być przechowywany poza systemem
  - cykl i częstotliwość archiwizacji muszą być dostosowane do częstotliwości zmian w konfiguracji systemu
    → w razie konieczności zachowania logów systemowych (`/var`) częstotliwość ich archiwizacji może być inna niż pozostałych plików systemowych
- archiwizacja plików użytkowników i baz danych
  - częstotliwość archiwizacji z reguły wyższa niż systemu, może być wymagana np. wiele razy dziennie
    → często celowa jest archiwizacja przyrostowa na zmianę z całościową
  - konieczne jest zapewnienie poufności danych

Plan i harmonogram archiwizacji

- określenie częstotliwości archiwizacji poszczególnych struktur, oraz wyboru archiwizacji pełnej i/lub liczby poziomów archiwizacji przyrostowej
- przypisanie konkretnych taśm do konkretnych dysków lub systemów plików, i nadanie taśmom etykiet (nalepek na szpulach lub kasetach)
- określenie sposobu recyrkulacji taśm, tzn. które taśmy i kiedy można nagrywać ponownie

Narzędzia archiwizacji

- archiwizacja plików: nagrywanie oddzielnych plików w postaci archiwum z oddzielnymi plikami, nagłówkami, atrybutami, itp.
  → tar, cpio, backup
- archiwizacja struktur: nagrywanie całych struktur (możliwe jest nagrywanie struktur z wybranymi plikami)
  → dump (ufsdump)
- kompresja danych archiwalnych
  - zalety: mieści więcej danych na jednym nośniku, co może mieć kapitalne znaczenie jeśli uda się dzięki temu zmieścić całe archiwum na jednym nośniku, a w pozostałych przypadkach tylko znaczenie organizacyjne i ekonomiczne
  - wady: utrudniony dostęp do poszczególnych plików w archiwum, trudna do oszacowania pojemność nośnika, w przypadku uszkodzenia nośnika konsekwencje mogą być poważniejsze niż utrata danych jednego pliku

Przykładowy schemat archiwizacji

Wykorzystujemy 10 kompletów taśm, 6 wystarczających do zapisania kompletnego archiwum, i 4 mieszczące dzienne archiwa przyrostowe.

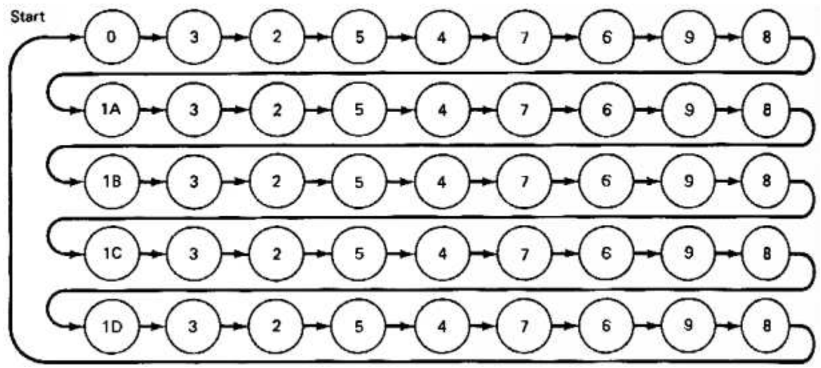
Taśmy etykietujemy: Poniedziałek, Wtorek, Środa, Czwartek (taśmy archiwizacji przyrostowej poziomu 1), i Piątek 1, Piątek 2, Piątek 3, Miesiąc 1, Miesiąc 2, Miesiąc 3 (taśmy archiwizacji pełnej poziomu 0).

Cykl rozpoczynamy w piątek wykonując pełną archiwizację na taśmie (taśmach) Piątek 1. We wszystkie dni powszednie: poniedziałek, wtorek, środę, i czwartek wykonujemy archiwizację przyrostową na odpowiednich taśmach.

W kolejne piątki nagrywamy pełną archiwizację na taśmach: Piatek 2 i Piątek 3, a co czwarty piątek wykorzystujemy taśmy: Miesiąc 1, itd.

Ten schemat zapewnia codzienną (dni powszednie) archiwizację przez 12 tygodni, z pełną archiwizacją w każdy piątek, po czym taśmy zostają ponownie użyte. Należy śledzić liczbę błędów zapisu (skorygowanych) aby wymieniać taśmy na nowe na długo zanim zaczną pojawiać się błędy niekorygowalne.

Bardziej rozbudowany schemat archiwizacji



Nagrywamy pełne archiwum poziomu 0, po czym wykonujemy archiwa przyrostowe poziomów 2–9 w cyklu podwójnym (każdy plik archiwizowany dwa razy). Po nagraniu ośmiu taśm zaczynamy cykl przyrostowy od nowa nagrywając wszystko od poziomu 0, i cztery razy powtarzamy taki duży cykl używając czterech oddzielnych kompletów taśm poziomu 1, ale jednego kompletu taśm poziomów 2–9. Do archiwizacji poziomu 0 używamy za każdym razem nowych taśm.

Zasady archiwizacji

- nośniki nagrywa się raz, a następnie zabezpiecza przed zapisem
- nagrany nośnik musi być od razu precyzyjnie zaetykietowany, najlepiej przed nagraniem
- nagrane nośniki muszą być przechowywane w bezpiecznym miejscu
- wskazane jest prowadzenie dziennika archiwizacji i ewentualnie spisów zawartości taśm
- archiwizację należy wykonywać na nieczynnych systemach plików (zasada generalnie nieprzestrzegana)
- sprawdzać, sprawdzać, sprawdzać:
  - sprawdzać czytelność nagranego archiwum
  - sprawdzać pracę operatora nagrywającego taśmy
  - sprawdzać poprawność całej procedury archiwizacji i zdolność odtworzenia pojedynczych plików i całego systemu
- sensem archiwizacji jest czarny scenariusz; nie ma żadnego znaczenia, że przy sprzyjających warunkach czasami uda się coś odtworzyć