

PAMIĘĆ MASOWA

Awaria głowicy (head crash) – kontakt głowicy z powierzchnią dysku.

Sterownik macierzysty (host controller) – sterownik używający szyny do komunikacji ze sterownikiem dysku w napędzie lub tablicy pamięci masowej.

Planowanie dostępu do dysku (disk scheduling)

Czas dostępu (random-access / positioning time) zależy od:

- czasu wyszukania (seek time) – przemieszczenie ramienia tak, aby głowica trafiła do cylindra zawierającego potrzebny sektor. Zależy od odległości.
- opóźnienia obrotowego (rotational latency) – czas zużywany na obrót dysku aby potrzebny sektor trafił pod głowicę.

Przepustowość, szerokość pasma (*bandwidth*) - iloraz łącznej liczby przesłanych bajtów i czasu od pierwszego zamówienia obsługi do zakończenia ostatniej transmisji.

Metoda FCFS (first-come, first-served)

Pierwszy zgłoszony, pierwszy obsłużony. Metoda sprawiedliwa, ale nie zapewnia najszybszej obsługi.

Metoda SSTF (shortest seek-time first)

Najkrótszy czas przeszukiwania – wybiera się zamówienie znajdujące się najbliżej głowicy. Może powodować głodzenie, nie jest optymalny.

Metoda SCAN

Omiatanie. Przesuwanie ramienia od jednej krawędzi dysku w kierunku przeciwległej i obsługując po drodze zamówienia. Po dotarciu do skrajnego cylindra zmienia kierunek i kontynuuje. Dysk nieustannie jest przeszukiwany tam i z powrotem.

Inna nazwa: algorytm windy (elevator algorithm). Średnia gęstość zamówień: mała przy zmianie kierunku, duża przy przeciwległym końcu.

Metoda C-SCAN (circular SCAN)

Omiatanie cykliczne. Odmiana metody SCAN. Przesuwa głowicę od jednej krawędzi dysku do drugiej, obsługując zamówienia. Po dotarciu do krawędzi wraca do miejsca początkowego nie obsługując żadnych zamówień. Cylindry są traktowane jak lista cykliczna – po ostatnim jest pierwszy.

Daje równomierniejszy czas oczekiwania niż SCAN.

Metoda LOOK, C-LOOK

Obieg z obserwowaniem przedpola. Modyfikacje metody SCAN i C-SCAN, w których głowica nie dochodzi do krańca dysku, tylko do najdalej wysuniętego zamówienia.

Wybór algorytmu

Sprawność algorytmów zależy od liczby i typu zamówień, metody przydziału bloków pliku. Algorytm powinien być osobnym modułem systemu operacyjnego, aby można było go zastąpić innym. SSTF i LOOK – dobre na algorytmy domyślne. Najpowszechniejsze jest SSTF. W systemach z dużym zapotrzebowaniem na operacje dyskowe lepiej działa (C-)SCAN – nie powoduje głodzenia.

Zarządzanie dyskiem

Formatowanie dysku

Formatowanie niskiego poziomu, fizyczne (low-level formatting, physical) - podzielenie dysku na sektory, które sterownik dysku potrafi czytać i zapisywać.

Przygotowanie dysku do przechowywania plików:

- podział dysku na grupy cylindrów, SO może traktować te strefy (partycje - partition) jako osobne dyski,
- formatowanie logiczne, tworzenie systemu plików – zapamiętanie na dysku początkowych struktur danych systemu plików (np. mapa wolnych i przydzielonych obszarów, początkowy pusty katalog).

Blok rozruchowy

Program rozruchowy (bootstrap) – ustawia stan początkowy wszystkich elementów systemu, następnie uruchamia system operacyjny. Przechowywany w pamięci ROM (read only memory).

Bloki uszkodzone

Oszczędzanie sektorów, metoda sektorów zapasowych (sector sparing) – metoda radzenia sobie z uszkodzonymi blokami. Przy formatowaniu niskopoziomym zostawia się zapas sektorów niewidocznych dla SO – wtedy możliwe jest zastępowanie uszkodzonego sektora za pomocą sektora ze zbioru sektorów zapasowych.

Zarządzanie obszarem wymiany

Pamięć wirtualna korzysta z pamięci dyskowej jako rozszerzenia pamięci głównej. Obszar wymiany (swap space) umożliwia najlepszą przepustowość systemowej pamięci wirtualnej. Może służyć do przechowywania obrazów całych procesów lub stron usuniętych z pamięci głównej. Może być położony w osobnej strefie dyskowej lub w ramach zwykłego systemu plików.

Mapy wymiany (swap maps) – służą do doglądania użycia obszaru wymiany.

Niezawodność dysku

RAID, nadmiarowa tablica niezależnych dysków (redundant array of independent disks) – technika zwiększenia niezawodności dysku dzięki pamiętaniu nadmiarowych danych (zwiększa średni czas do awarii (mean time to failure)).

Paskowanie dysku (disk stripping) – traktowanie grupy dysków jako jedną jednostkę pamięci. Każdy blok danych jest podzielony na kilka podbloków, z których każdy jest pamiętany na innym dysku.

Schematy RAID:

- odbicie lustrzane (mirroring), tworzenie cienia (shadowing, RAID 1) dubluje każdy dysk.

- paskowane odbicia lustrzane (RAID 1+0), lustrzane paski (RAID 0+1) zapewniają wysoką sprawność i niezawodność.

- parzystość w przeplacie bloków (*block interleaved parity*, RAID 4, 5, 6) – jeden dysk (spośród np. dziewięciu) wykorzystywany jest do przechowywania bloków parzystości. W razie awarii jednego z bloków można obliczyć na nowo bity na podstawie innych bloków danych i bloku parzystości.

RAID w obrębie tablicy pamięciowej może zawieść, jeśli zawiedzie tablica, więc typowe jest automatyczne zwielokrotnianie (replication) danych między tablicami.

Żelazny zapas (hot-spare) - nieprzydzielone dyski do automatycznego zastępowania uszkodzonego dysku i odbudowywania na nich danych.

System plików ZFS (system Solaris uwzględnia sumy kontrolne (checksums) wszystkich danych i metadanych, co umożliwia kontrolę poprawności obiektu i wykrywanie jego zmian.

Realizacja pamięci trwałej

Pamięć trwała – potrzebna do pisania z wyprzedzeniem. Realizacja:

- zwielokrotnienie potrzebnych informacji na wielu urządzeniach pamięci, niezależnych od siebie pod względem awaryjności.
- aktualizowanie informacji w kontrolowany sposób, zapewniając, że potrafimy odtworzyć dane po awarii wynikłej w czasie przesyłania danych lub rekonstrukcji.

Pamięć trzeciorzędna

Główne cechy: niski koszt, zazwyczaj nośniki wymienne (removable media).

Dyski wymienne

- dysk magnetoptyczny (magneto-optic disk) – dane są zapisywane na sztywnej płycie pokrytej materiałem magnetycznym. Zapisywanie danych: ciepło promienia lasera wzmacnia punktowo pole magnetyczne wytwarzane przez cewkę napędu, co umożliwia zapisanie bitu. Odczyt: efekt Kerra (odbicie światła lasera).
- dysk optyczny (optical disk) – stosuje się materiały zmieniające właściwości pod wpływem działania światła laserowego.
- dysk zmiennofazowy (phase-change disk) – wykonany z materiału zmieniającego fazy – bezpostaciowy bądź skryształizowany – różnice stanu dają różną przezroczystość. Usuwanie danych poprzez topienie nośnika.
- dysk WORM (write-once read-many-times) – odporne na uszkodzenia, niezawodne. Wykonanie: cienka folia aluminiowa obłożona z dwóch stron szkłem/plastikiem – wypala się maleńkie otwory w folii, więc każdy sektor może zostać napisany tylko raz. Informacja może zostać zniszczona, ale nie zmieniona.

Taśmy magnetyczne

Większa pojemność niż dysków, znacznie dłuższy czas losowego dostępu, stosowane głównie do składowań danych (np. kopii dysków).

Biblioteki taśm z automatycznymi ich zmieniającymi – staker (niewielka liczba), silos (tysiące taśm).

Zadania systemu operacyjnego

Zarządzanie urządzeniami fizycznymi oraz tworzenie abstrakcji maszyny wirtualnej na użytek aplikacji.

Abstrakcje dysków twardych:

- urządzenie surowe (raw device) – tablica bloków danych.
- system plików (file system) – kolejkovanie, planowanie przeplatania wykonania.

Interfejs aplikacji.

SO traktuje na ogół dyski wymienne tak jak dyski stałe – należy je sformatować i otrzymuje się pusty system plików, z którego korzysta się tak samo jak z tego na dysku twardym.

Taśmy są traktowane jako surowy nośnik pamięci. Program nie otwiera pliku na taśmie, ale cały przewijak jako urządzenie surowe. Przewijak jest zazwyczaj rezerwowany na wyłączny użytek danej aplikacji. Ponieważ każda aplikacja określa własne zasady organizacji taśmy, wypełniona danymi taśma może być na ogół użyta tylko przez program, który ją utworzył.

Operacje napędu dysku: czytaj, pisz, szukaj.

Operacje przewijaka taśmy: znajdź (locate) zamiast szukaj – dokładniejsze, ustawia taśmę w miejscu występowania określonego bloku logicznego, a nie ścieżki; pozycja czytania (read position) – zwraca numer bloku logicznego; odstęp (space) – zmiana położenia względem bieżącego miejsca.

Nazywanie plików

SO nie rozwiązuje problemu przestrzeni nazw (*name space problem*) na nośnikach wymiennych. Aby ustalać dostęp do danych i ich interpretację, trzeba polegać na aplikacjach i użytkownikach. Są jednak ustandaryzowane nośniki – np. CD.

Zarządzanie pamięcią hierarchiczną

Rozszerzenie hierarchii pamięci o pamięć trzeciorzędną, zazwyczaj implementowaną za pomocą robota kasetowego (robotic jukebox) z taśmami lub dyskami wymiennymi.

Pamięć trzeciorzędna służy zazwyczaj jako rozszerzenie systemu plików – małe, często używane pliki pozostają na dysku, a duże rzadko używane są archiwizowane przez roboty kasetowe.

Szybkość

Przepustowość, szerokość pasma (jednostka bajt/sekunda):

- przepustowość stała (sustained bandwidth) – szybkość zaobserwowana w trakcie długich przesyłań danych,
- przepustowość efektywna (effective bandwidth) – średnia z całego czasu trwania operacji wejścia-wyjścia – łącznie z czasem odnajdywania danych oraz przełączania kaset przez roboty.

Opóźnienie dostępu – czas potrzebny do zlokalizowania danych. Znacznie krótszy dla dysków niż taśm.

Niezawodność

Dysk zamontowany na stałe ma większą niezawodność niż dyski wymienne (działanie warunków środowiska). Kaseata optyczna jest bardziej niezawodna niż dysk magnetyczny czy taśma. Awaria głowicy dysku twardego niszczy dane zupełnie, a w przypadku taśmy czy dysku optycznego dane często pozostają nienaruszone.