Sekundarne memorije

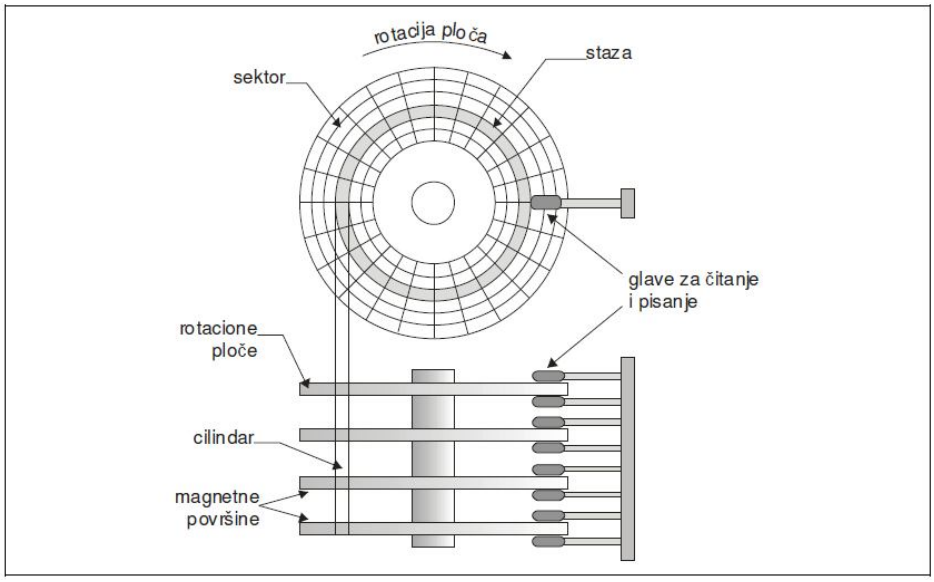
Za razliku od radne memorije, sadržaj sekundarnih memorija ne gubi se nakon isključivanja napajanja. U sekundarnim memorijama smešteni su operativni sistem, programi i podaci koji se obrađuju. Formatiranje diskova na niskom nivou, podela diskova na particije i formiranje sistema datoteka najčešće se izvode pre instaliranja operativnog sistema. Većina diskova koji koji se danas proizvode fabrički su formatirani. Disk se deli naparticije ako na njemu treba da se instalira više operativnih sistema ili u cilju razdvajanja sistemskih i korisničkih podataka. Za razliku od radne memorije, koja je brza, diskovi i optički uređaji znatno su sporiji jer imaju mehaničke delove, pa predstavljaju usko grlo računarskog sistema.

Geometrija diskova

Disk uređaj se sastoji od kružnih ploča koje rotiraju oko zajedničke ose. Površine ploča su presvučene magnetnim materijalom. Svaka površina ima pridruženu galvu za čitanje i pisanje koja čita podatke sa magnetnih ploča ili ih upisuje na ploče. U većini disk uređaja, samo jedna glava može biti aktivna u jednom trenutku. Glave se linearno pokreću pomoću servo-sistema, čime im je, uz rotaciju ploča, omogućen pristup svim delovima magnetne površine. Kompletnim uređajem upravlja kontrolna elektronika diska.

Procesor računara i disk komuniciraju preko kontrolera diska. Kontroleri različitih diskova pružaju isti interfejs ka ostatku računara, čime pojednostavljuju pristup podacima na disku, pa računar ne treba da poznaje način rada elektromehanike diska i ne mora upravljati njenim radom. U ostale funkcije kontrolera spadaju baferovanje podataka koje treba upisati na disk, keširanje diskova i automatsko obeležavanje neispravnih sektora diska.

Površina diska je podeljena u koncentrične prstenove – staze (*tracks*), a svaka staza je podeljena na sektore (*sectors*). Na starijim diskovima, sve staze najčešće imaju isti broj sektora. Spoljašnje staze novijih diskova obično su podeljene na veći broj sektora, čime obezbeđuju jednaku magnetnu površinu za sve sektore. U jedan sektor obično se upisuje 512 bajtova podataka, i to je najmanja količina podataka koja se može upisati na disk ili pročitati sa njega.



Geometrija diskova

Sve površine magnetnih ploča jednako su poreljene na staze i sektore. To znači da se glave za čitanje i pisanje na svim pločama diska u jednom trenutku nalaze na istim stazama. Ekvidistantne staze svih ploča čine jedan cilindar. Datoteke koje nisu smeštene u okviru jednog cilindra fragmentisane su – pomeranje glava s jedne staze na drugu prilikom čitanja ovakvih datoteka unosi kašnjenje. Performanse diska se mogu poboljšati smeštanjem datoteke u jedan cilindar kad god je to moguće.

Geometrija diska je u opštem slučaju određena brojem magnetnih površina (tj glava ua čitanje i pisanje), cilindara i sektora i čuva se na posebnoj memorijskoj lokaciji sa baterijskim napajanjem – CMOS RAM. Operativni sistem čita vrednosti koje opisuju geometriju diska priliko podizanja sistema ili inicijalizacije drajvera. Dalje se trodimenzionalnim adresiranjem „glava, cilindar, sektor“ može pristupiti svim delovima diska da bi se prišlo postojećim podacima ili dodelio prostor za nove podatke. Npr podatak koji je upisan na drugu površinu, u stazu 3, u sektor 5 može se adresirati uređenom trojkom (*head, cylinder, sector*)=(2,3,5).

ATA i SCSI diskovi

ATA i SCSI diskovi su najčešće korišćene klase diskova na PC računarima. ATA uređaji (*Advanced Technology Attachment*), poznati i kao IDE (*Integrated Drive Electronics*), dobili su ime po elektronici integrisanoj na samom uređaju. Ovoj klasi pripadaju relativno jeftini diskovi solidnih performansi – kapaciteta do 160 GB, sa brzinama okretanja ploča od 5400 do 7200 obrtaja u minuti. Kontroleri za ATA uređaje su ugrađeni na matične ploče računara i obezbeđuju interfejs ka računaru pri brzini od 33 do 133 Mbps. Realna brzina čitanja sa magnetnih površina i upisivanja nanjih znatno je manja, tako da na kontroleru postoji bafer u koji se podaci smeštaju pre upisivanja na disk. Na taj način se sprečava da performanse sistema značajno oslabe prilikom rada s diskovima.

Svaki ATA kontroler ima dva kanala – primarni (*primary*) i sekundarni (*secondary*), a na svaki kanal se mogu vezati najviše dva uređaja u odnosu nadređen/podređen (*master/slave*). Uređaji vezani na različite kanale mogu istovremeno da primaju i šalju podatke računaru. Na jednom kanalu, samo jedan uređaj može biti aktivan u jednom trenutku. Svaki ATA uređaj ima preklopnike koje, pre vezivanja na kontroler, treba postaviti u željeni režim rada – nadređen ili podređen. Administrator sistema određuje način na koji će uređaji biti vezani na ATA kontroler – uređaje treba vezati tako da se performanse sistema održe na najvišem mogućem nivou.

SCSI uređaji predstavljaju profesionalni interfejs za široki spektar uređaja – diskova, CD-ROM uređaja, DVD uređaja, traka, skenera itd. Kontroler za SCSI diskove nije integrisan na matičnim pločama i kupuje se odvojeno. Elektronika SCSI kontrolera je komplikovanija, interfejs ka računaru je brži (do 320 Mbps), a na kontroler je moguće vezati od 7 do 15 uređaja, u zavisnosti od kontrolera. SCSI uređaji se ne nalaze u odnosu nadređen/podređen, već se na kontroler vezuju prema prioritetima. Prioritet svakog uređaja određen je njegovim identifikacionim brojem, koji se postavlja preko preklopnika na uređaju. Princip je sledeći: najviši prioritet ima uređaj čiji je ID=0 i treba ga dodeliti sistemskom disku, a najniži prioritet ima uređaj sa ID=15. identifikacioni broj ID=7 rezervisan je za SCSI kontroler.

Priprema diskova za rad

Priprema diskova za rad obuhvata sledeće administrativne postupke:

- Formatiranje diskova
- Izradu particija
- Formiranje sistema datoteka

Formatiranje diskova

Formatiranje diskova je proces kojim se na magnetni medijum upisuju oznake koje predstavljaju granice staza i sektora, čime se uvodi red u magnetni haos neformatirane površine. Disk koji nije formatiran ne može da se koristi. Pod formatiranjem se na DOS i Windows sistemima podrazumeva proces formiranja sistema datoteka. Taj proces se u literaturi označava pod imenom formatiranje visokog nivoa (*high-level formating*). Formatiranje diska na niskom nivou (*low-level formating*) predstavlja pripremu magnetne površine diska za rad.

Pravljenje particija

Da bi se formatirani disk mogao koristiti pod bilo kojim OS, disk se priprema u dve faze.U prvoj fazi disk se mora izdeliti na particije u kojima se može formirati sistem datoteka. Veći broj particija se pravi ukoliko na računar sa jednim diskom treba instalirati više operativnih sistema. Svaki OS koristi svoju pariticiju, a po potrebi može da čita podatke sa drugih sistema ukoliko ima podršku za rad sa tim sistemom. Disk se često deli na particije i na računarima na kojima je instaliran samo jedan OS – na taj način se sistemske datoteke lako mogu razdvojiti od korisničkih.

Informacije o svim particijama diska se čuvaju u prvom logičkom sektoru, tj u prvom sektoru prve staze na prvoj površini diska. Ovaj sektor je poznat pod nazivom ***glavni startni zapis (Master Boot Record, MBR)*** i njemu BIOS pristupa kad god se računar uključuje. MBR sadrži mali program koji očitava particionu tabelu, proverava koja je particija aktivna i očitava prvi sektor aktivne particije tj ***startni sektor (boot sector).*** U startnom sektoru nalazi se mali program čijim pokretanjem započinje ***bootstrap*** rutina, tj učitavanje operativnog sistema u RAM memoriju. Po originalnom konceptu particionisanja diskova na PC računarima, mogle su postojati najviše 4 particije po disku, što se pokazalo kao nedovoljno.

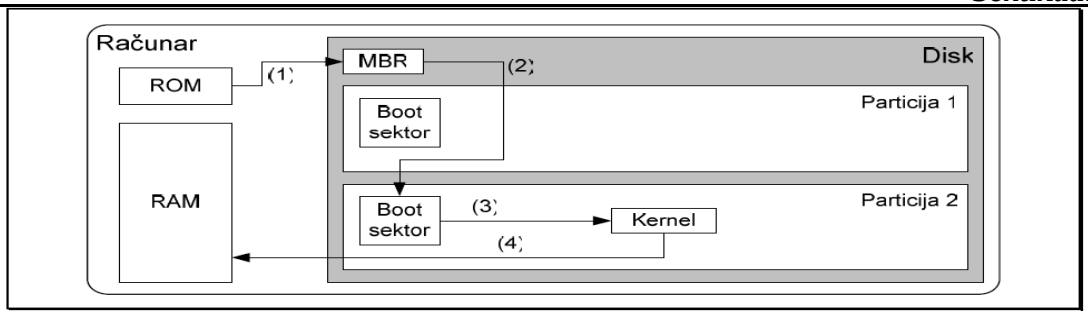
Razlozi su sledeći:nemoguće je na samo 4 primarne particije instalirati više OS-a (naročito ako se zahtevaju dodatne particije kao što su swap i boot), boot manager i odvojiti particije za korisničke podatke. Problem je rešen uvođenjem ***produžene particije (extended partition).*** Ona služi kao okvir u kome se mogu praviti nekoliko logičkih particija. Logičke particije se ponašaju kao primarne, ali se razlikuju po načinu izrade. Informacije o logičkim particijama vode se u startnom sektoru produžene particije, koja se još i naziva ***tabela produženih particija.*** Na disku može postojati samo jedna produžena particija. Nakon podele diska na particije sledi druga faza – formatiranje visokog nivoa, tj formiranje samih sistema datoteka. Programi za logičko formatiranje prave kontrolne strukture pomoću kojih se upravlja sistemom datoteka.

Bootstrap rutina

Bootstrap rutina uključuje tri osnovna koraka:

- ***rutinu POST***
- ***pronalaženje aktivnog operativnog sistema***
- ***učitavanje jezgra u memoriju***

Kada se računar uključi, BIOS izvršava rutinu POST (*Power On Self Test*), tj pokreće seriju testova hardvera nakon čega kreće proces podizanja sistema. Procedura podizanja sistema (*boot*) izvršava se u cilju dovođenja sistema u operativno stanje. Inicijalno podizanje većine operativnih sistema izvršava se kroz više faza, pri čemu je početak koda u ROM memoriji. Taj kôd obično nije dovoljan i teško se menja, tako da se dopunjuje programima na disku. Kôd u prvom sektoru diska (*master boot record, MBR*) najpre identifikuje aktivnu particiju u particionoj tabeli, a zatim izvršava kôd upisan u startni sektor aktivne particije, koji dalje učitava jezgro u memoriju. U startnom sektoru se nalazi mali program koji je zadužen da pokrene učitavanje operativnog sistema u memoriju. Delovi koda kojim se memorija puni u toj ranoj fazi, nalazi se na fiksnim područjima diska, a ne u sistemima datoteka, zato što u toj fazi nema jezgra, pa ni podrške za sistem datoteka. Rana faza podizanja operativnog sistema završava se učitavanjem jezgra.



Bootstrap rutina

Raspoređivanje zahteva za rad sa diskom

Operativni sistem mora da obezbedi efikasno iskorišćenje hardvera, a ulazno-izlazni uređaji predstavljaju usko grlo računarskog sistema po pitanju performansi. Uzmimo npr magnetne diskove – vreme pristupa disku zavisi od sledećih komponenata:

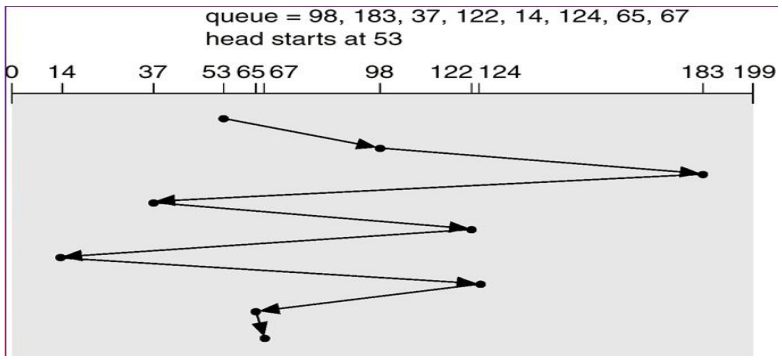
- vreme pozicioniranja glava za čitanje i pisanje sa tekuće pozicije na zahtevani cilindar
- vremena rotacionog kašnjenja tj vremena potrebnog da se ispod magnetne glave postavi zahtevani blok podataka
- brzine prenosa podataka sa magnetnog medijuma, koja zavisi od gustine medijuma i brzine okretanja rotacionih površina diska

Brzina prenosa podataka sa diska ili na disk (*bandwidth*) količnik je ukupnog broja prenetih bajtova i ukupnog vremena koje obuhvata ove tri komponente. U višeprocesnom okruženju, u jednom trenutku postoji veliki broj zahteva za rad sa diskom. Pravilnim raspoređivanjem ovih zahteva (*disk scheduling*), ukupno vreme pozicioniranja ili rotacionog kašnjenja može se skratiti. Svaki zahtev koji je upućen disku sadrži sledeće informacije: da li se zahteva operacija čitanja ili pisanja, adresu bloka na disku, adresu mamorijskog bafera i broj bajtova koje treba preneti. Više zahteva može stići istovremeno, a u jednom trenutku disk može obraditi samo jedan. Postoji više algoritama za raspoređivanje zahteva za rad sa diskovima.

FCFS

Najjednostavniji algoritam FCFS (*First Come First Served*) zahteve prosleđuje po redosledu pristizanja. FCFS je krajnje fer prema prispelim zahtevima, ali daje znatno lošije performanse. Uzimo sledeći primer: neka je glava za čitanje i pisanje trenutno na cilindru 57, a u redu čekanja za disk zahtevi pristižu po sledećem redu: 98,183,37,122,14,124,65,67.

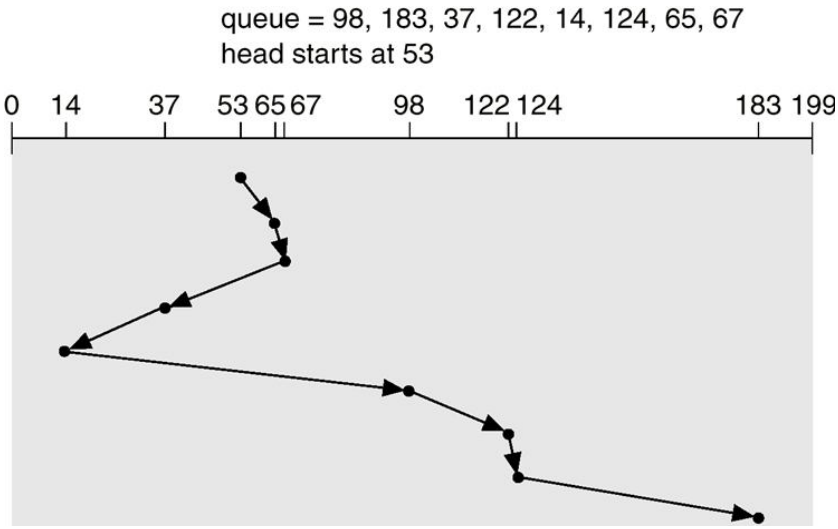
Performanse zavise od ukupnog pomeranja glava za čitanje i pisanje pri opsluživanju zahteva iz reda, a savršeno je jasno da performanse slabe sa većim pomeranjima. Relativne performanse izrazićemo ukupnim brojem cilindara koje glave za čitanje i pisanje prelaze pri opsluživanju zahteva. U slučaju sa slike, ukupni pomeraj glava diska iznosi 640 cilindara.



Algoritam FCFS

SSTF

Algoritam SSTF (*Shortest Seek Time First*) opslužuje zahteve na sledeći način: od prispelih zahteva najpre se uzima onaj koji će izazvati najmanji pomeraj glava. Uzmimo isti preimer koji je korišćen za ilustraciju algoritma FCFS. SSTF će opslužiti zahteve prema sekvenci datoj na slici, a ukupni pomeraj glava diska iznosi 236 cilindara

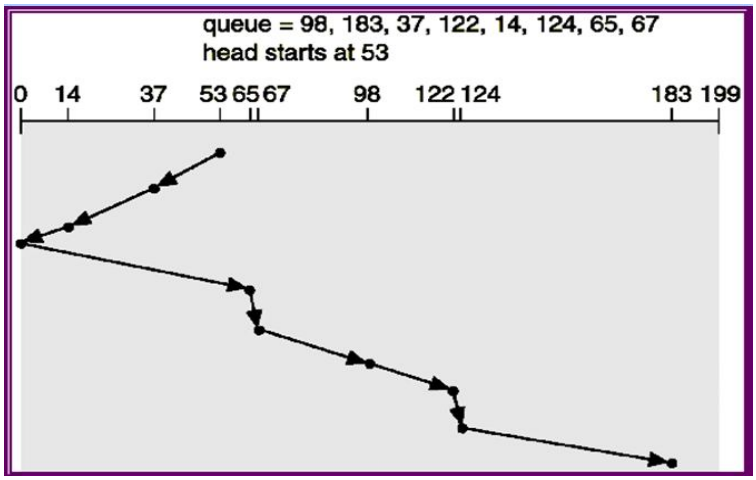


Algoritam SSTF

Algoritam podseća na SJF (*Shortest Job First*) – algoritam za paspoređivanje procesa – i optimalan je u pogledu vremena pozicioniranja. Međutim, pri korišćenju algoritma SSTF, nastaje problem zakucavanja (*starvation*). Glave mogu ostati veoma dugo u jednoj zoni, opslužujući zahteve koji unose male pomeraje, tako da zahtevi čiji su cilindri daleko od tekuće pozicije mogu dugo čekati u redu.

SCAN

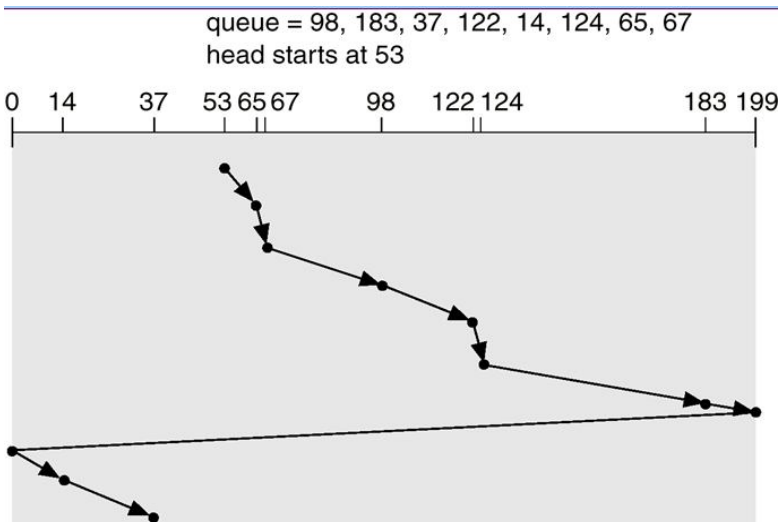
Algoritam SCAN radi kao lift koji se naizmenično kreće od prizemlja do vrha zgrade. Algoritam naizmenično pomera glave od početka do kraja diska i nazad, opsluživajući zahteve koji se nalaze na tekućem cilindru. Na ovaj način se rešava problem zakucavanja. Prilikom obrade zahteva, SCAN daje prednost unutrašnjim cilindrima u odnosu na periferne.



Algoritam SCAN

C-SCAN

Kružni SCAN algoritam (*C-SCAN, Circular SCAN*) je varijanta SCAN algoritma koji rešava problem favorizovanja unutrašnjih cilindara. Izmena se sastoji u tome da se zahtevi opslužuju samo u jednom smeru. Kada glave dođu do poslednjeg cilindra, pomeraju se na početak, ne opslužujući zahteve na tom putu. Posle toga se nastavlja opsluživanje zahteva od početnog do krajnjeg cilindra.

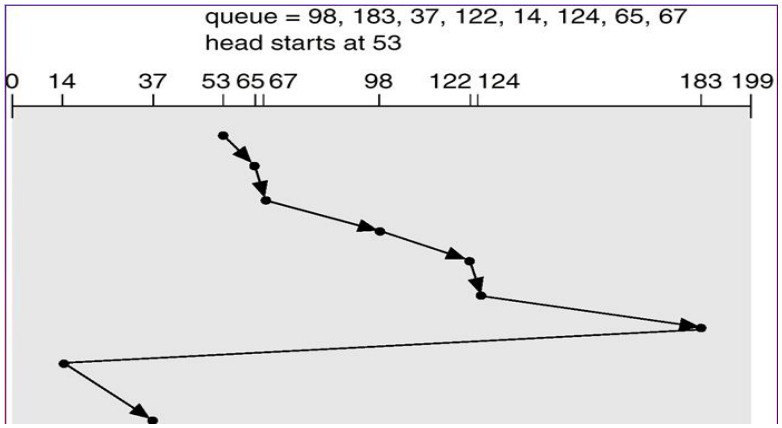


Algoritam C-SCAN

LOOK i C-LOOK

Algoritmi LOOK i C-LOOK modifikacije su algoritma SCAN i C-SCAN. Glave se ne pomeraju do kraja ili početka diska, nego do poslednjeg zahteva koji se nalazi u redu čekanja u tom smeru. Pri tome, LOOK opslužuje zahteve u oba smera, C-LOOK samo u rastućem smeru do poslednjeg zahteva u redu, nakon čega se vraća na zahtev najbliži početku diska.

Algoritmi su dobili ime po tome što „gledaju“ na kom se cilindru nalazi poslednji zahtev u tom smeru i krežu sa tog cilindra. U praksi se umesto algoritma SCAN uvek koriste algoritmi LOOK.



Algoritam C-LOOK

Izbor algoritma za raspoređivanje zahteva za rad sa diskovima

Kako izabrati najbolji algoritma za raspoređivanje zahteva za korišćenje diska?

Jasno je da su svi algoritmi bolji od FCFS-a, ali je teško odrediti koji je najbolji, jer performanse samih algoritama zavise od opterećenja, tj od prispelih zahteva za rad sa diskom. Kružne varijante algoritama SCAN i LOOK imaju mnogo bolju raspodelu opsluživanja i ne izazivaju problem zakucavanja, koji postoji pri radu sa algoritmom SSTF. C-LOOK je najbolje rešenje za jako opterećene sisteme.

Prethodno pomenuti algoritmi su zastareli, a modernije varijante ovih algoritama minimizuju obe mehaničke komponente - i pozicioniranje i rotaciono kašnjenje (pri čemu rotaciono kašnjenje ima dominantan uticaj na performanse savremenih diskova). Jedan takav algoritam je SATF (*Shortest Access Time First*), koji radi na principu algoritma SSTF, ali pri odabiru sledećeg zahteva iz reda, računa obe mehaničke komponente. Najsavremeniji algoritmi uzimaju u obzir i keširanje na samom disk uređaju. Alogoritam C-LOOK, u kombinaciji sa ugrađenim keširanjem diska, daje najbolje rezultate.

RAID strukture – realizacija stabilnih sistema

RAID koncept razvijen je na University of California, Berkeley, sa ciljem da se što bolje iskoriste diskovi malog kapaciteta. RAID tehnika (Redundand Area of Inexpensive Disks) predstavlja različite načine upotrebe diskova radi postizanja veće pouzdanosti i boljih performansi. RAID se deli na nivoe 0-6 koji se mogu realizovati hardverski i softverski.

1. RAID 0

RAID 0 predstavlja RAID konfiguraciju u kojoj je traka na nivou jednog ili više blokova podataka (*block-striping*). To je postupak kojim se podaci ravnomerno raspoređuju na sve diskove u nizu, u cilju poboljšavanja performansi sistema. Ovaj postupak se realizuje deljenjem diskova na trake (stripes) čije veličine zavise od tipa opereativnog sistema i namene niza diskova. RAID 0 ne unosi redundansu, ali se u slučaju otkaza jednog diska svi podaci nepovratno gube.

Disk 1	Disk 2	Disk 3	Disk 4
A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P
↓	↓	↓	↓

RAID 0

2. RAID 1

U konfiguraciji RAID 1 (*disk mirroring*), svaki disk ima svoje ogledalo. RAID 1 ima najveći utrošak prostora i najgore performanse upisa. Za svaki disk u nizu se uvodi jedan rezervni koji u svakom trenutku sadrži sliku originala. RAID 1 uvodi 100% redundanse, zaštita podataka je potpuna, ali je cena sistema duplo veća.

Disk 1	Disk 2	Disk 3 (r1)	Disk 4 (r2)
A	B	A	B
C	D	C	D
E	F	E	F
G	H	G	H
↓	↓	↓	↓

RAID 1 (Disk mirroring)

3. RAID 2

Konfiguracija RAID 2 poznata je pod nazivom RAID sa memorijskim stilom korekcije (*memory style error correcting code ECC organisation*). Memorije imaju ECC algoritam koji za svaki bajt ima 3 ekstra bita, potrebna za detekciju i korekciju jednobitnih grešaka. RAID 2 ima organizaciju deljenja podataka na bit ili bajt nivou, a bez obzira na broj diskova podataka, potrebna su još tri diska za ECC koja mogu sačuvati podatke u slučaju otkaza bilo kog diska. RAID 2 je dobar po pitanju paralelizma, bolji je od RAID 0 po pitanju utroška diskova , ali se praktično ne koristi

4. RAID 3

RAID level 3 je postupak kojim se podaci ravnomerno raspoređuju na više diskova u nizu u cilju poboljšanja performansi, a u cilju povećanja pouzdanosti uvodi se disk za kontrolu parnosti. Ukoliko dođe do otkaza jednog diska svi podaci su i dalje dostupni. Uvodi se parnost za diskove. Ideja je sledeća: za razliku od memorije u kojoj je jako teško odrediti tačnu poziciju grešaka, kod diskova se tačno zna gde je nastupila greška. Za oštećeni bit, dovoljan je jedan bit da čuva parnost i na bazi te parnosti može da se rekonstruiše oštećeni bit. Za sve diskove podataka dovoljan je jedan bit parnosti.

Disk 1	Disk 2	Disk 3	Disk 4 (par)
A	B	C	par (A, B, C)
D	E	F	par (D, E, F)
G	H	I	par (G, H, I)
J	K	L	par (J, K, L)
↓	↓	↓	↓

RAID 3

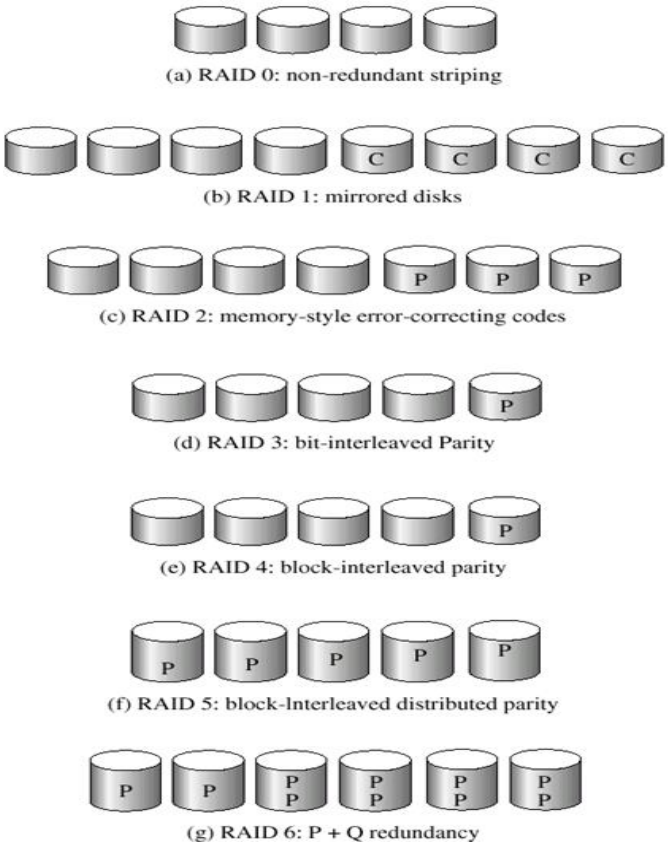
5. RAID 5

RAID level 5 je postupak sličan RAID 3 s tim što se provera parnosti raspodeljuje na sve diskove u sistemu. Podaci su dostupni nakon otkaza jednog diska. Ovaj postupak je poznat pod imenom rotacioni niz parnosti (*Rotating Parity Array*). U ovom slučaju ne postoji jedan disk parnosti , nego svi diskovi predstavljaju i diskove podataka i diskove parnosti. Parnost se upisuje u levom simetričnom rasporedu. Od osnovnih struktura RAID 5 predstavlja najbolju kombinaciju. Poseduje paralelizam, konkurentnost, dobar je za velike upise, a svi diskovi su ravnomerno opterećeni. Jedino ostake problem malih upisa koji se delimično razrešava preko keša na RAID nivou.

Disk 1	Disk 2	Disk 3	Disk 4
A	B	C	par (A, B, C)
D	E	par (D, E, F)	F
G	par (G, H, I)	H	I
par (J, K, L)	J	K	L
↓	↓	↓	↓

RAID 5

RAID 6 predstavlja jedinu RAID kombinaciju koja može razrešiti problem u slučaju otkaza više od jednog diska (ukoliko se koriste prethodne šeme, osim RAID 1, svi podaci se gube ako istovremeno otkažu bar 2 diska). Da bi se postigao takav stepen pouzdanosti, RAID 6 nastaje uvođenjem dodatnog diska (Q) i *Read-Solomon* koda u šemu RAID.



RAID nivoi

Prilikom projektovanja RAID sistema, projektant treba da odgovori na nekoliko pitanja:

- Koji će RAID nivo da se primeni?
- Koliko diskova će biti u RAID sistemu?
- Koliko grupa parnosti će se napraviti?
- Koliko će biti veličina trake?

Potrebno je odabrati kvalitetan RAID kontroler. Većina novojih matičnih ploča ima ugrađene RAID kontrolere koji podržavaju konfiguracije RAID 0,1,0+1 i 5 – ovi kontroleri nisu hardverski, već softverski. Hardverski RAID sam deli podatke u trake, dok u slučaju softverskog RAID-a to obavlja procesor. Uz softverske RAID kontrolere obično se isporučuju drajveri samo za Windows platforme. Uloga kontrolera je da omogući *bootstrap* rutinu, dok je dalji rad – podela podataka na trake – posao procesora.