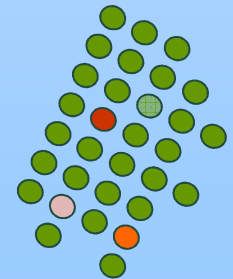
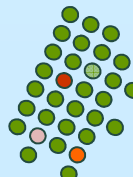

Gestione della memoria di massa

Capitolo 12 -- Silberschatz



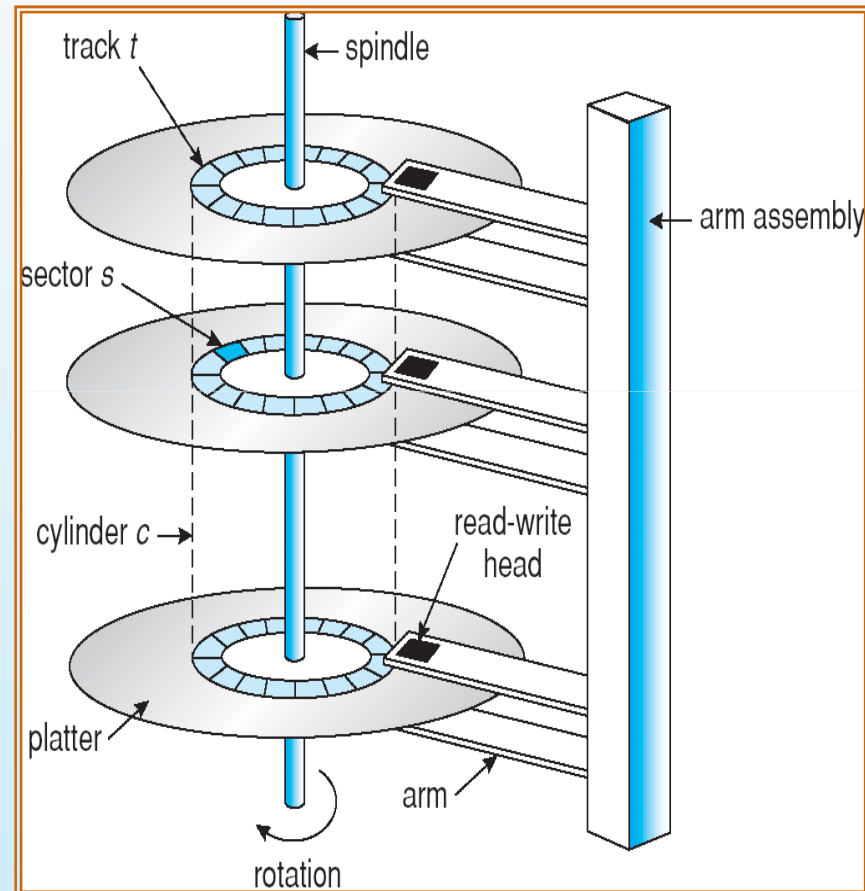
Nastro magnetico

- Usato in passato come dispositivo di memorizzazione secondaria
- Può contenere grosse quantità di dati
- Principalmente usato per backup, memorizzazione di dati usati raramente, trasferimenti dati tra sistemi
- Tempo di posizionamento elevato
 - Accesso diretto ~1000 volte più lento del disco (essendo fisicamente ad accesso sequenziale)
- Quando i dati sono sotto la testina, **le velocità di trasferimento sono comparabili** a quelli dei dischi
- Capacità tipiche vanno da 20 a 200GB
- I nastri sono classificati per larghezza
 - tipologie comuni sono: 4mm, 8mm, 19mm



Disco magnetico

- I dischi magnetici realizzano la memoria secondaria nei calcolatori moderni
 - I piatti del disco ruotano da 60 a 200 volte al secondo
 - **Velocità di trasferimento:** è la velocità con cui i dati vengono trasferiti dal disco al computer
 - **Tempo di posizionamento (o tempo di accesso):** è il tempo per muovere la testina sul settore desiderato (tempo di ricerca + latenza di rotazione)
 - **Caduta della testina:** la testina entra a contatto con la superficie del disco



Disco magnetico

- I dischi possono essere rimovibili
 - I floppy disk sono dischi magnetici economici e rimuovibili costituiti da un involucro di plastica che protegge il piatto flessibile
- Sono collegati al computer tramite **bus di I/O**
 - Più comuni: **EIDE, ATA, SATA, USB, Fibre Channel, SCSI**
 - Un **controller (host controller)** del computer usa il bus per parlare con il **controller (disk controller)** del disco



Struttura logica del disco

- Un disco viene visto come un grande array monodimensionale di *blocchi logici*, dove il blocco logico è la più piccola unità di trasferimento (512 byte di solito).
- L'array monodimensionale di blocchi logici è mappato in settori del disco sequenzialmente.
 - Il settore 0 è il primo settore della prima traccia del cilindro più esterno.
 - La mappatura procede in ordine attraverso la traccia, poi attraverso le tracce restanti nel cilindro, e infine attraverso i cilindri restanti, dal più esterno al più interno.



Mapping logico - fisico

- Alcuni settori possono essere danneggiati
- Il numero di settori per traccia non è costante su alcuni dispositivi
- **Dispositivi CVL** (constant linear velocity) – la densità di bit per traccia è uniforme
 - tracce esterne contengono più settori
- **Dispositivi CAV** (constant angular velocity) – la densità di bit decresce dalle tracce interne verso quelle esterne



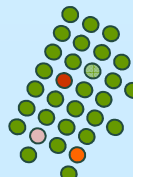
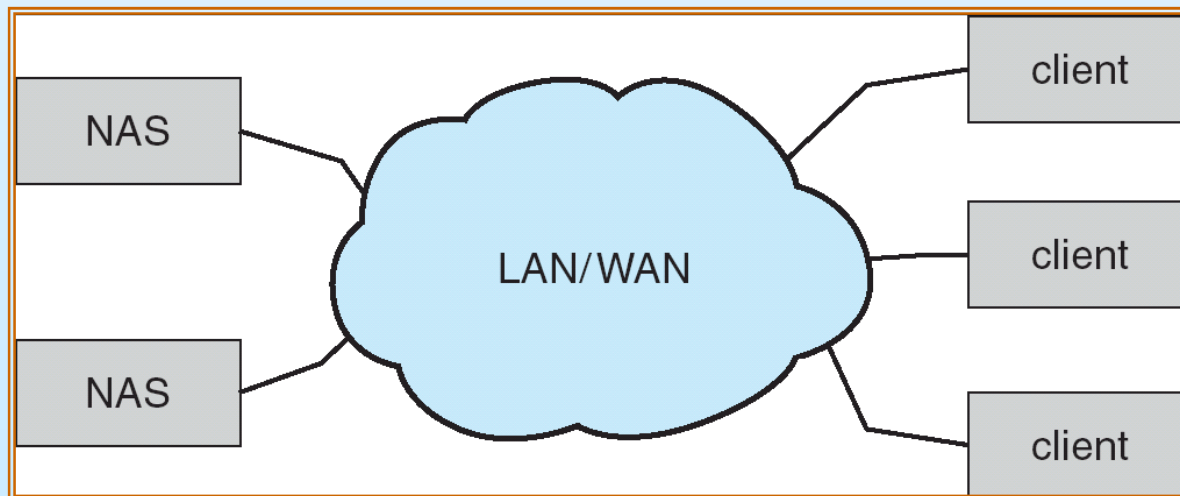
Collegamento dei dischi

- I dischi connessi direttamente al calcolatore sono accessibili tramite porte di I/O che comunicano con i bus di I/O
- **SCSI**: supporta fino a 16 dispositivi su un cavo. Uno **SCSI initiator** (controllore) richiede un'operazione e gli **SCSI target** (15 unità di memorizzazione) la realizzano
 - Ciascun target può avere fino a 8 **unità logiche**
- **FC** (Fiber Channel) è un'architettura seriale ad alta velocità
 - Può avere capacità di commutazione con uno spazio di indirizzamento a 24-bit – la base per **storage area networks (SANs)** in cui molti host sono collegati a molte unità di memorizzazione
 - Variante: a ciclo arbitrato (**arbitrated loop (FC-AL)**) con 126 dispositivi (unità e controller)



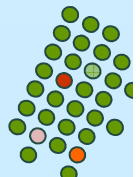
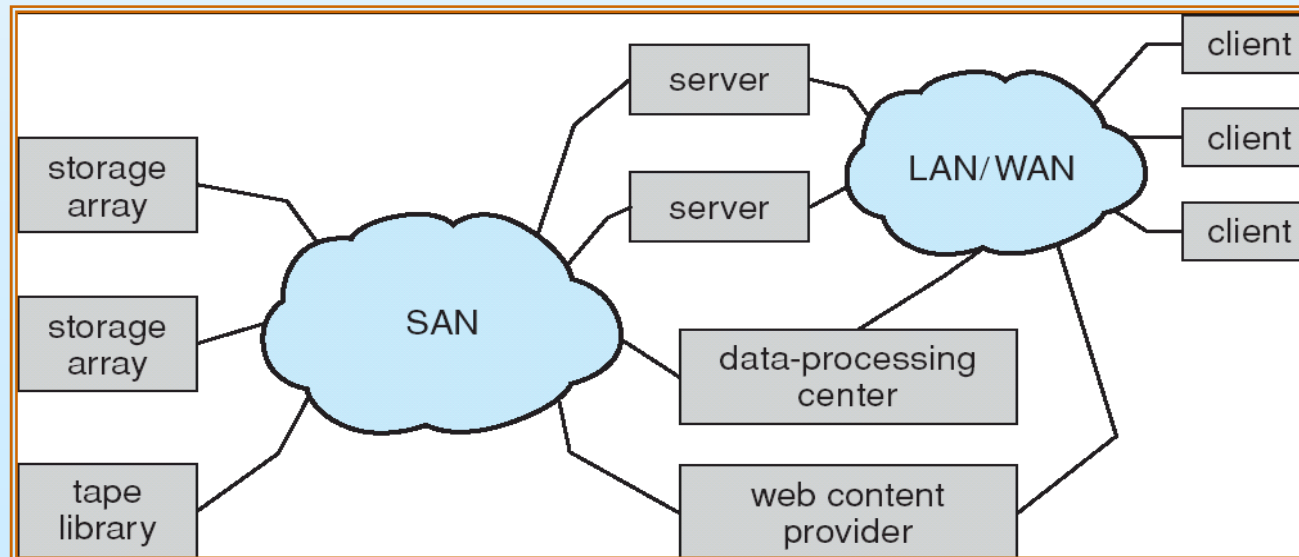
Memorizzazione con collegamento di rete

- La memoria con collegamento di rete (Network-attached storage -**NAS**) è memoria accessibile tramite rete piuttosto che connessioni locali (e.g., bus)
- Implementata tramite chiamate di procedura remota (RPC) tra l'host e il dispositivo di memorizzazione
- Il protocollo iSCSI usa reti IP per portare su rete il protocollo SCSI



Reti di memoria secondaria

- Impiega protocolli specifici per la memorizzazione anzichè protocolli di rete
- Usata in grandi ambienti di memorizzazione (sta diventando sempre più comune)
- Host multipli sono collegati a più array di dischi – schema flessibile



Scheduling del disco

- Uso efficiente dell'hardware per le unità a disco: rapido tempo di posizionamento e ampiezza di banda.
- Il **tempo di posizionamento** ha due componenti principali:
 - Il **tempo di ricerca** (seek time) è il tempo che impiega il braccio del disco a muovere le testine fino al cilindro contenente il settore desiderato.
 - La **latenza di rotazione** (rotational latency) è il tempo aggiuntivo speso in attesa che il disco faccia ruotare il settore desiderato sotto la testina.
 - Minimizzare il tempo di ricerca \approx distanza di ricerca.
- L' **ampiezza di banda** (bandwidth) del disco è data dal numero totale di byte trasferiti diviso per il tempo totale che intercorre fra la richiesta di servizio e il completamento dell'ultimo trasferimento.



Scheduling del disco

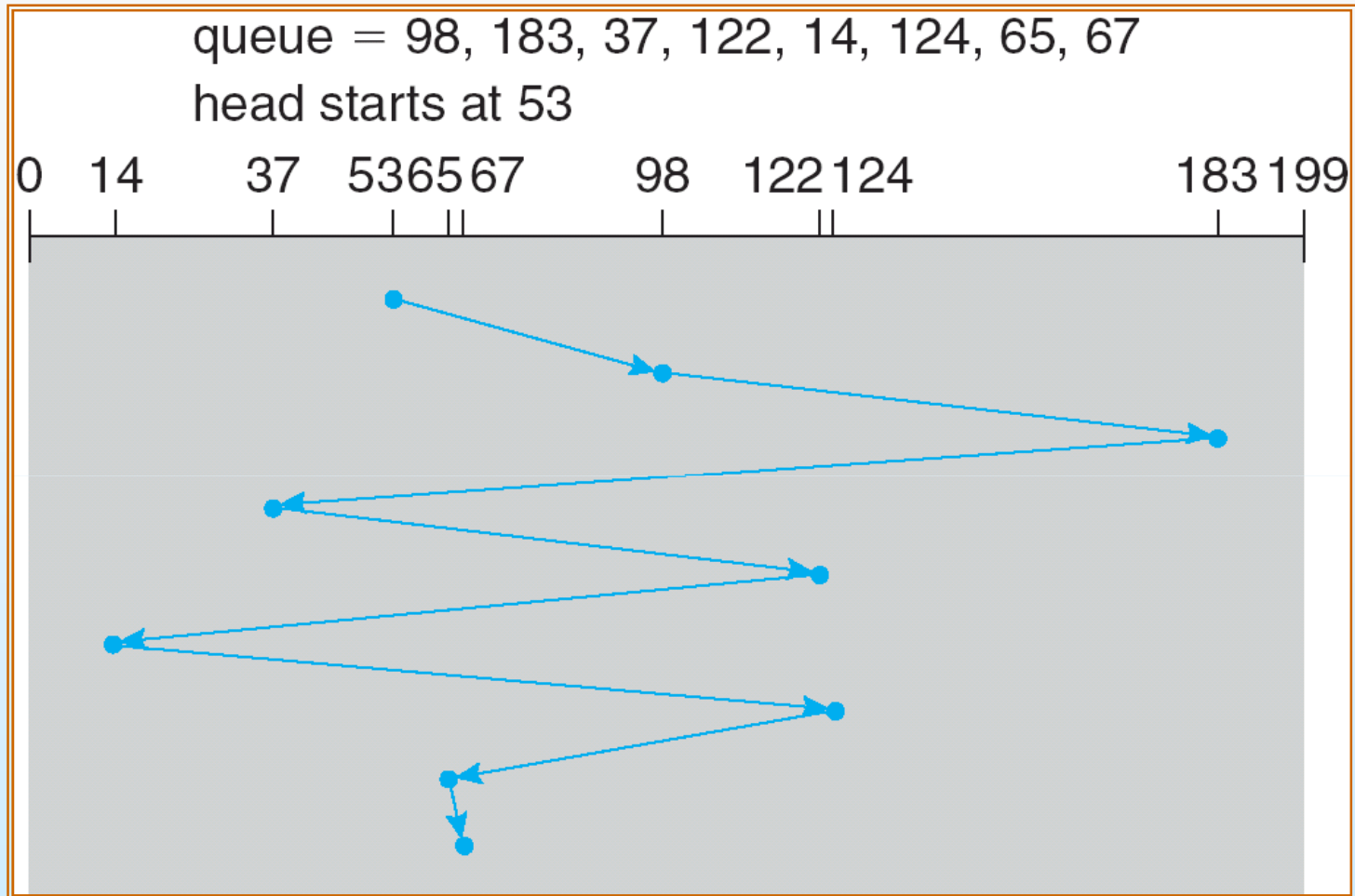
- Richiesta di I/O
 - Lettura/scrittura
 - Indirizzo del disco rispetto al quale trasferire
 - Indirizzo di memoria rispetto al quale trasferire
 - Numero di byte da trasferire
- Diversi algoritmi esistono per schedulare le richieste di I/O per il disco.
- Li presentiamo usando la seguente coda di richieste (0-199)

98, 183, 37, 122, 14, 124, 65, 67

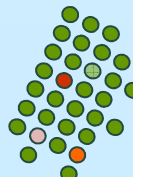
- La testina è posizionata sul cilindro 53



FCFS (First Come First Served)

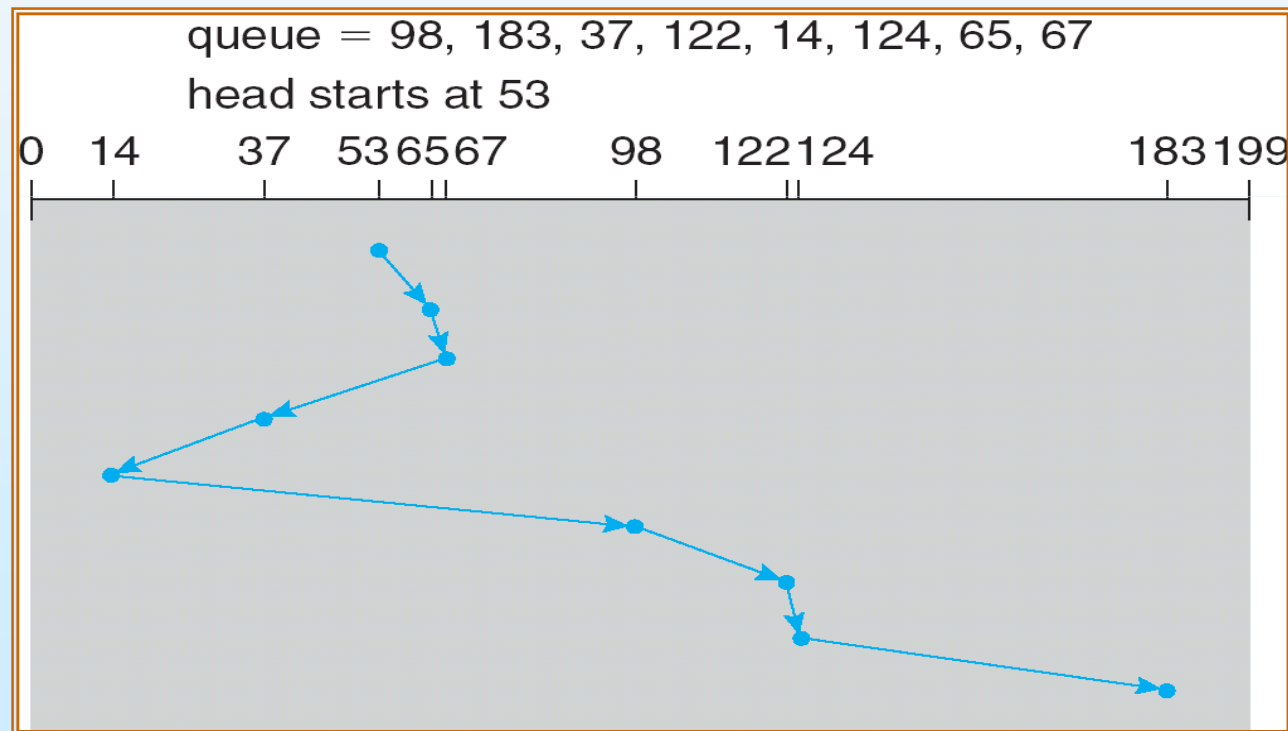


Il movimento totale della testina è di 640 cilindri



SSTF (Shortest Seek Time First)

- Seleziona la richiesta con il **minimo tempo di ricerca** dalla posizione corrente della testina.
- La schedulazione SSTF può causare attesa indefinita di alcune richieste (starvation).

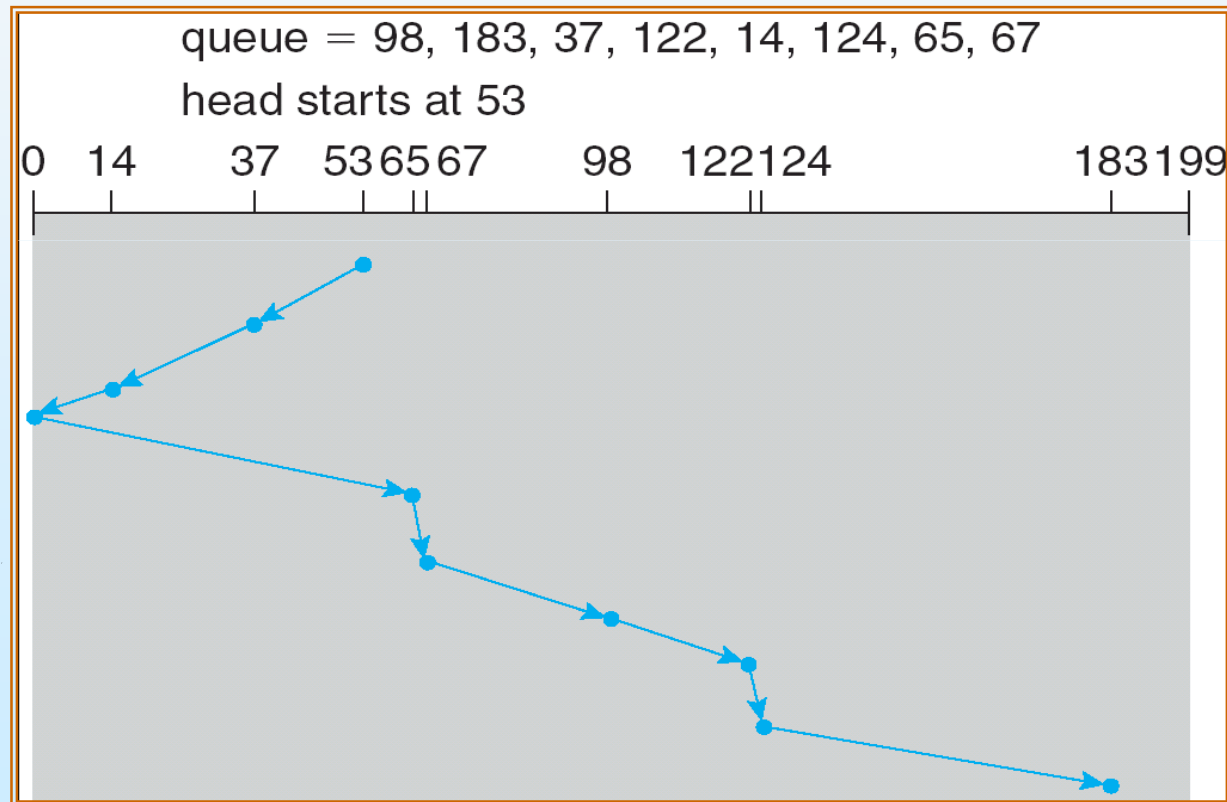


- Il movimento totale della testina è di 263 cilindri; **Non è ottimale** (servire prima le richieste per 37 e 14 riduce il movimento totale)



SCAN

- Il braccio del disco si muove da un estremo all'altro estremo, servendo le richieste che incontra. All'altro estremo il movimento viene invertito e il servizio continua.
- Talvolta chiamato algoritmo dell'ascensore.

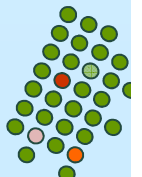


- il movimento totale della testina è 208 cilindri.

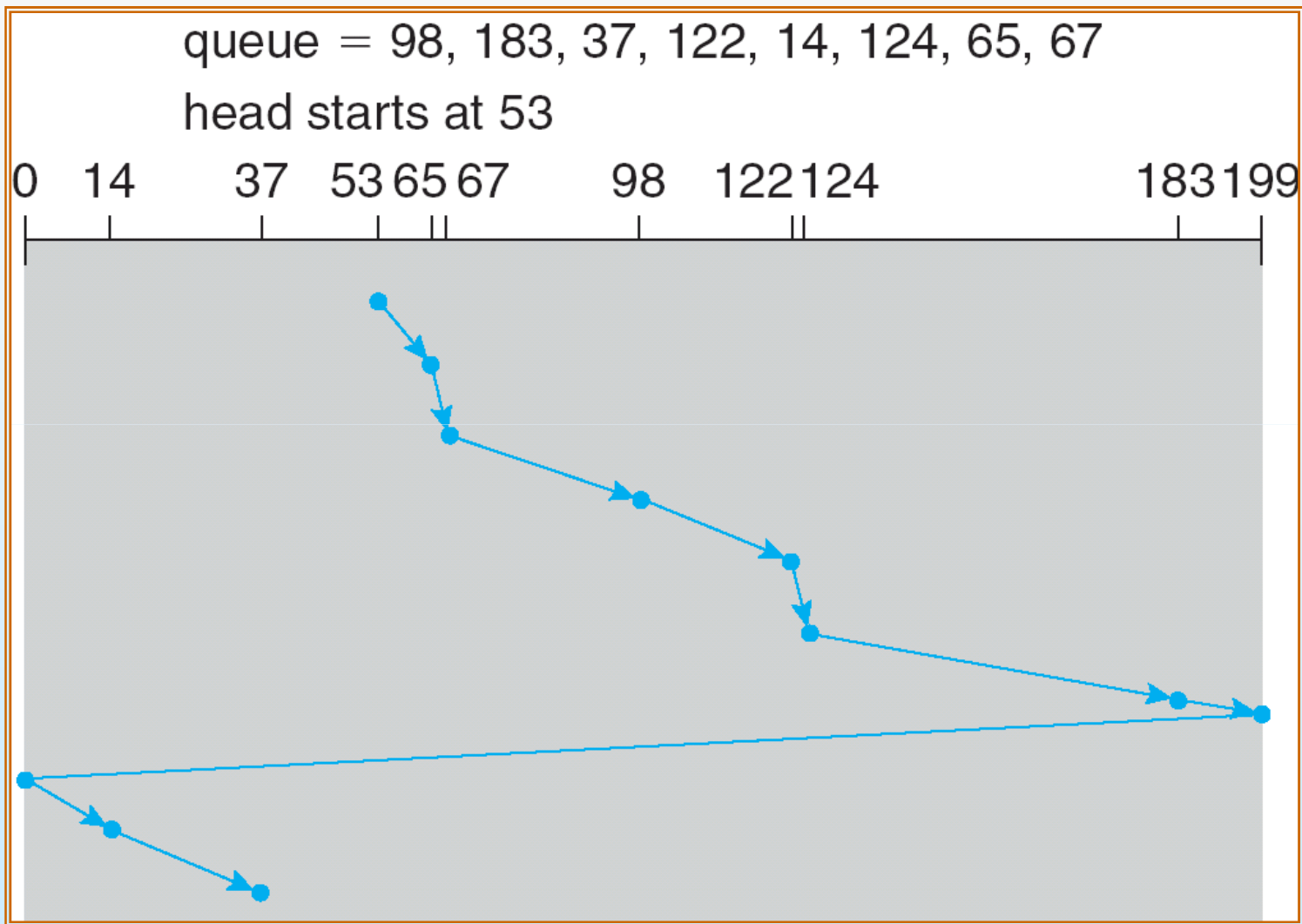


C-SCAN

- Fornisce un tempo di attesa più uniforme di SCAN.
- Il braccio si muove da un capo all'altro del disco, servendo le richieste lungo il percorso. Quando raggiunge l'altro capo ritorna direttamente all'inizio del disco, senza servire alcuna richiesta durante il ritorno.
- Tratta i cilindri come una lista circolare che si riavvolge dal cilindro finale al primo.

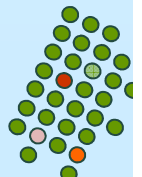
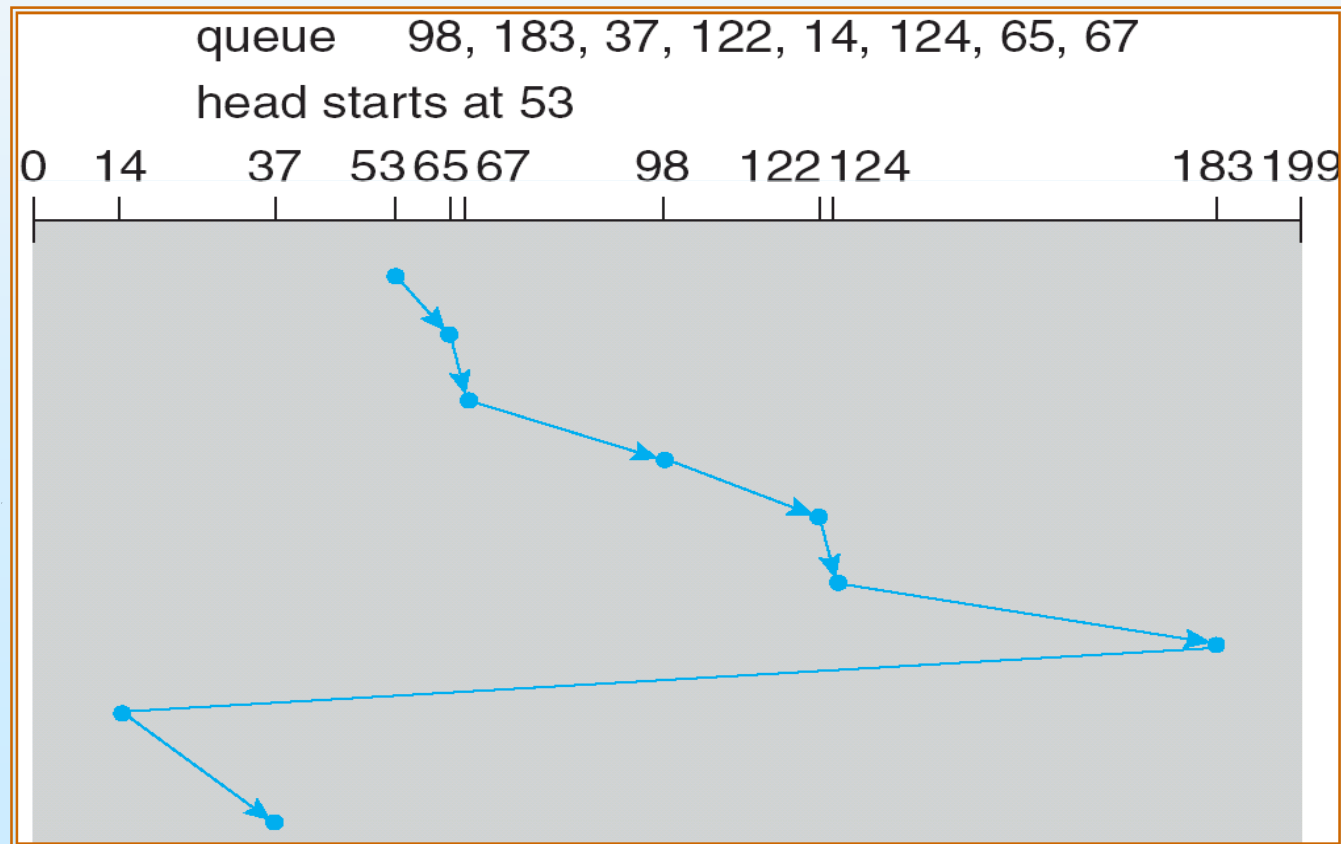


C-SCAN



C-LOOK

- Versione pratica di SCAN e C-SCAN.
- Il braccio arriva fin dove è presente la richiesta finale, per ciascuna delle due direzioni. Lì inverte immediatamente direzione, senza giungere all'estremità del disco.



Selezione dell'algoritmo di scheduling del disco

- SSTF è comune e naturale
- SCAN e C-SCAN danno migliori risultati per sistemi che pongono un carico pesante sul disco.
- Le prestazioni dipendono dal numero e dai tipi di richieste.
- Le richieste per servizio su disco possono essere influenzate dal metodo di allocazione dei file (e.g., contigua, concatenata).
- La procedura di schedulazione del disco dovrebbe essere scritta come modulo separato del sistema operativo, in modo da poter essere sostituita con una procedura differente se necessario.
- Sia SSTF che LOOK sono una scelta ragionevole come algoritmo predefinito.



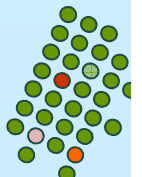
Gestione del disco

- *Formattazione a basso livello, o formattazione fisica* — Divide il disco in settori che il controller del disco può leggere e scrivere.
- Settore: header / area dati / trailer (numero settore, ECC)
 - Solitamente taglie di 256, 512, o 1024 byte
- Per usare un disco il sistema operativo crea delle partizioni contenenti uno o più cilindri. Ogni partizione è come se fosse un disco separato
- Dopo la creazione delle partizioni, occorre *formattarle logicamente* i.e., il sistema operativo ha bisogno di registrare su disco le proprie strutture dati (creazione di un file system)



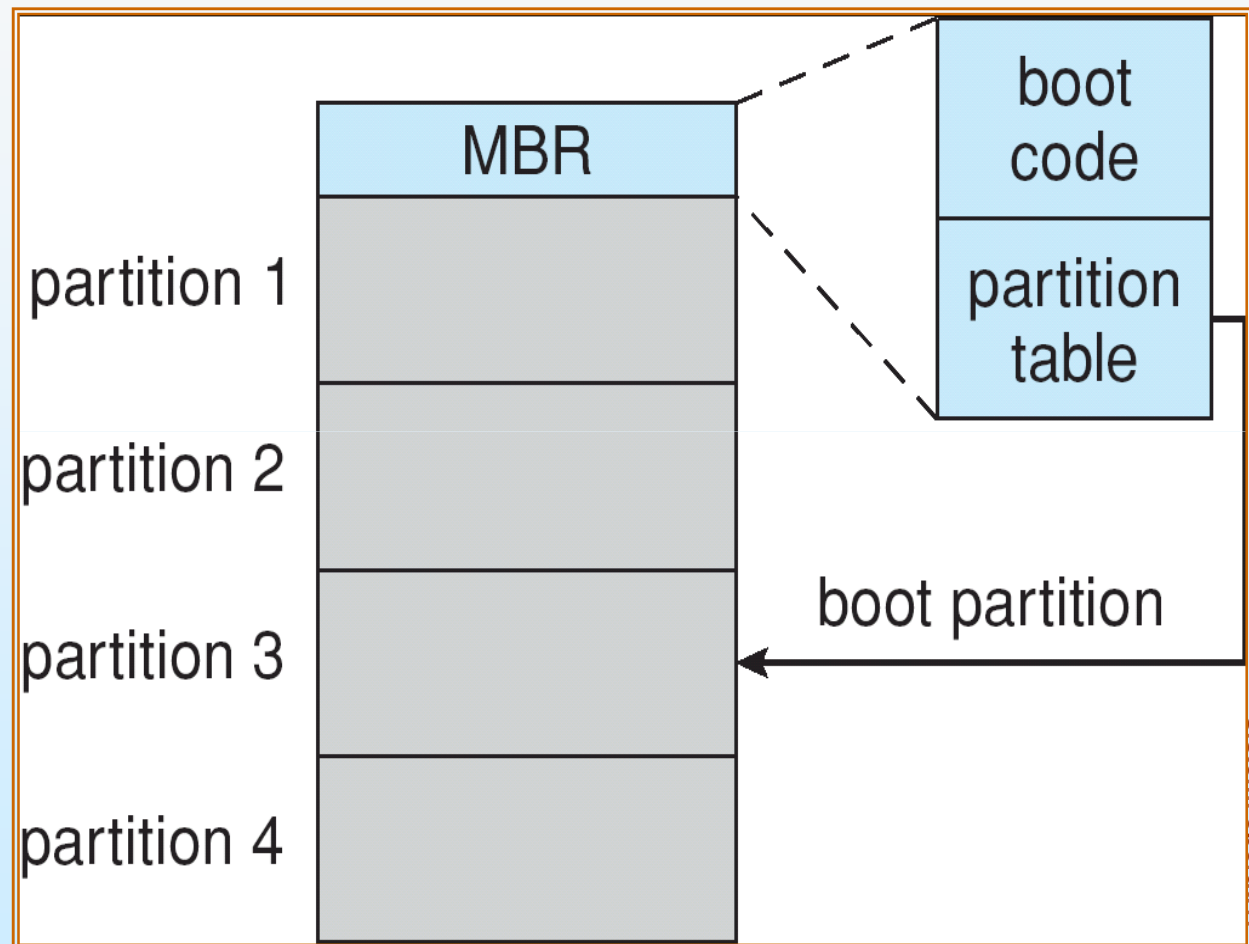
Gestione del disco

- Partizione d'avviamento contiene il sistema operativo. Il blocco di avviamento contiene codice per inizializzare il sistema.
 - Il bootstrap loader è immagazzinato in una ROM (Read Only Memory).
 - Programma d'avviamento completo (blocco d'avviamento).



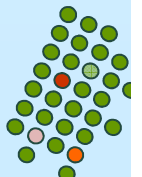
Boot da disco in Windows 2000

- ❑ Il codice di avviamento è collocato nel primo settore del disco rigido, detto MBR (Master Boot Record).
- ❑ Il sistema è guidato a leggere tale codice di avviamento.
- ❑ Facendo riferimento alla tabella delle partizioni (anch'essa presente nel MBR) si accede alla partizione di avviamento che contiene il sistema operativo ed i driver dei dispositivi



Gestione blocchi difettosi

- Metodi per gestire blocchi difettosi
 - *sector sparing (accantonamento dei settori)*
 - ▶ Il controllore è istruito così da accantonare i settori difettosi e sostituirli con settori di riserva inutilizzati
 - ▶ ogni volta che si fa riferimento al blocco logico relativo al settore difettoso, il controllore traduce la richiesta nell'indirizzo di rimpiazzo
 - ▶ Durante la formattazione viene mantenuto qualche settore di riserva per ogni cilindro



Gestione blocchi difettosi

- Metodi per gestire blocchi difettosi
 - *sector slipping (traslazione dei settori)*
 - ▶ Tutti i settori compresi tra il settore danneggiato e quello di riserva immediatamente successivo vengono spostati avanti di un settore
 - ▶ in questo modo si libera il settore successivo a quello danneggiato e quindi il settore danneggiato può essere fatto corrispondere al successivo

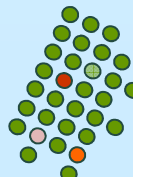
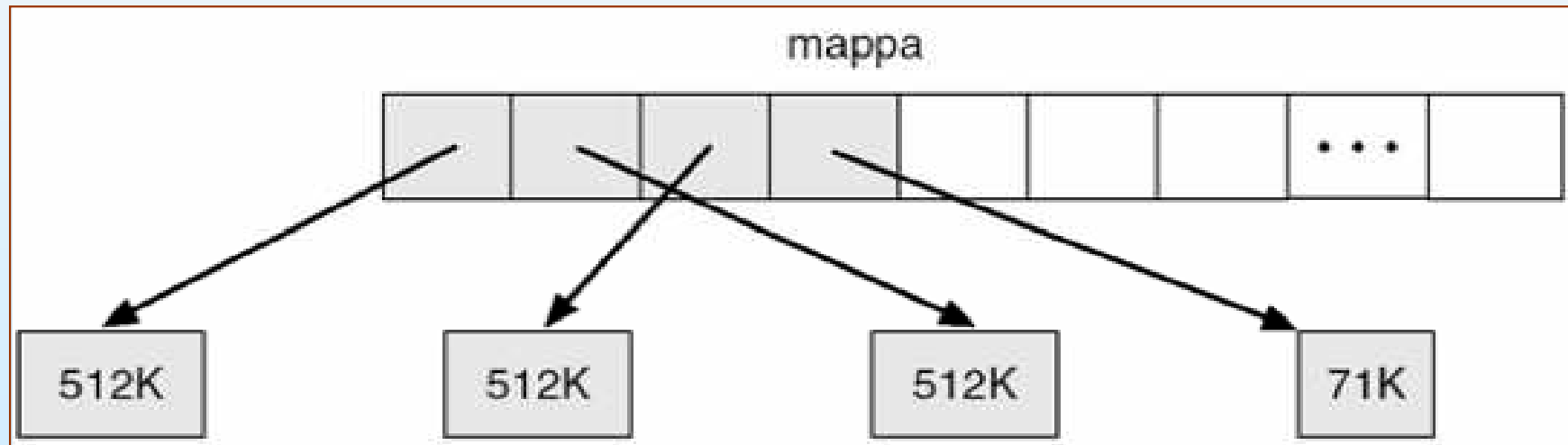


Gestione dello spazio di swap

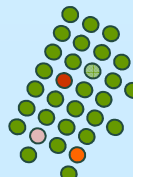
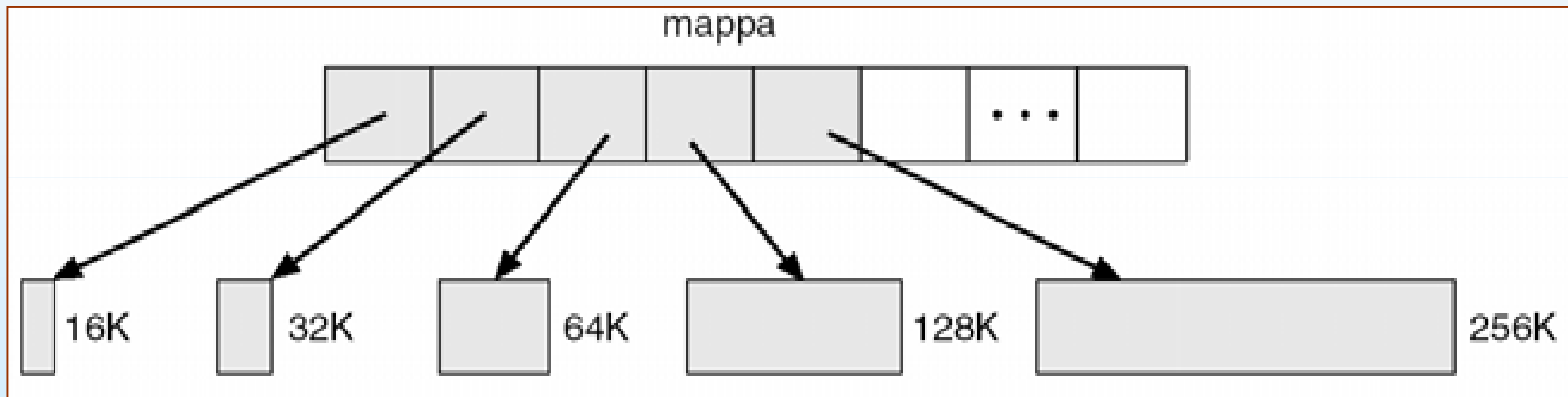
- La memoria virtuale usa lo spazio su disco come estensione della memoria centrale → per mascherare la diversa velocità di accesso si usa lo **spazio di swap** in cui le info di un processo vengono depositate per favorire un accesso più veloce
- Lo spazio di swap può essere ricavato dal normale file system, o può essere in una partizione separata del disco.
- Gestione dello spazio di swap può variare a seconda del SO.
 - BSD 4.3 assegna lo spazio di swap a un processo quando questo viene avviato. Memorizza **segmento testo** (programma), e un **segmento dati**.
 - ▶ Il kernel usa **mappe di swap** per tenere traccia dell'uso dello spazio di swap.
 - Solaris 2 assegna lo spazio di swap solo quando una pagina viene rimossa dalla memoria fisica, piuttosto che quando la pagina di memoria virtuale viene creata per la prima volta.



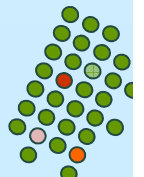
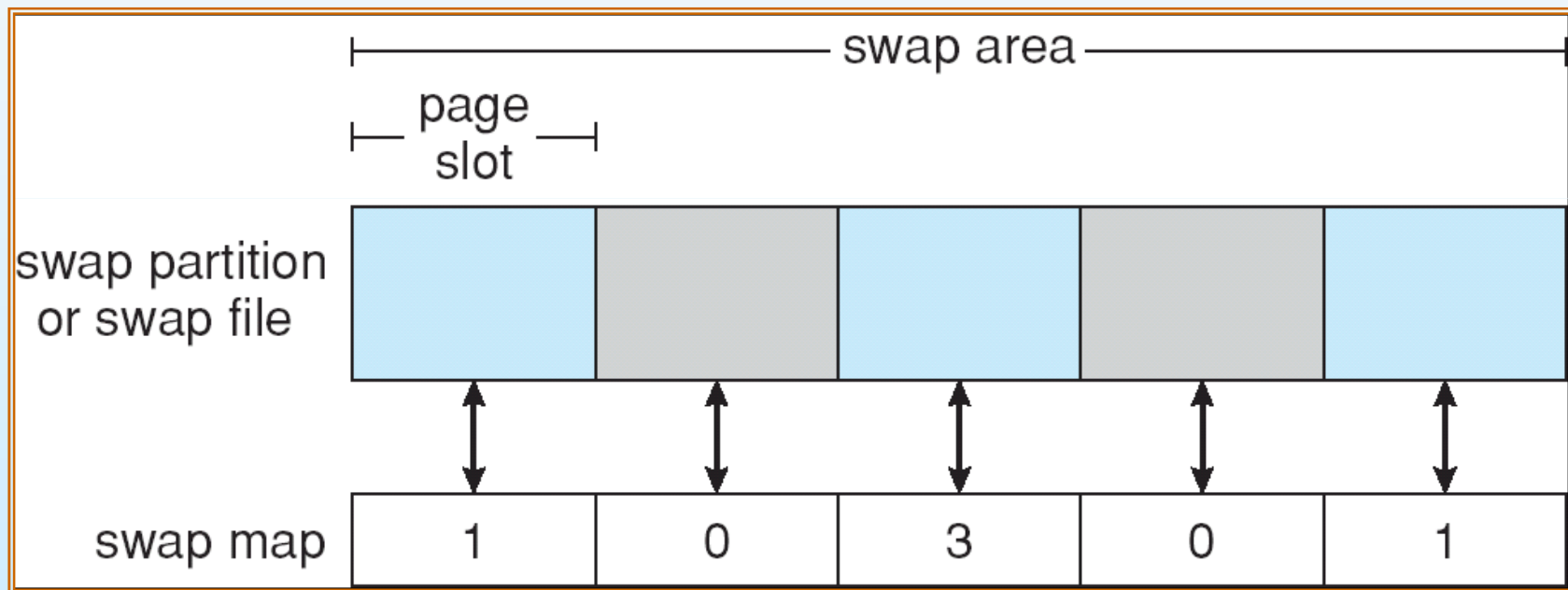
Mappa di un segmento testo in BSD 4.3



Mappa di un segmento DATI in BSD 4.3

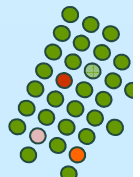


Strutture dati per lo Swapping in sistemi Linux

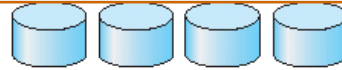


RAID

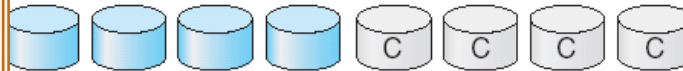
- **RAID** (redundant arrays of inexpensive disks) – dischi multipli forniscono affidabilità attraverso l'uso di ridondanza.
- Inexpensive / Independent
- Esistono sei differenti livelli di organizzazione RAID
- La copiatura speculare (*mirroring o shadowing*) mantiene una copia di ogni disco.
- Lo spezzettamento dei dati (*data striping*) utilizza un gruppo di dischi come una singola unità di memorizzazione.



Livelli RAID



(a) RAID 0: non-redundant striping.



(b) RAID 1: mirrored disks.



(c) RAID 2: memory-style error-correcting codes.



(d) RAID 3: bit-interleaved parity.



(e) RAID 4: block-interleaved parity.



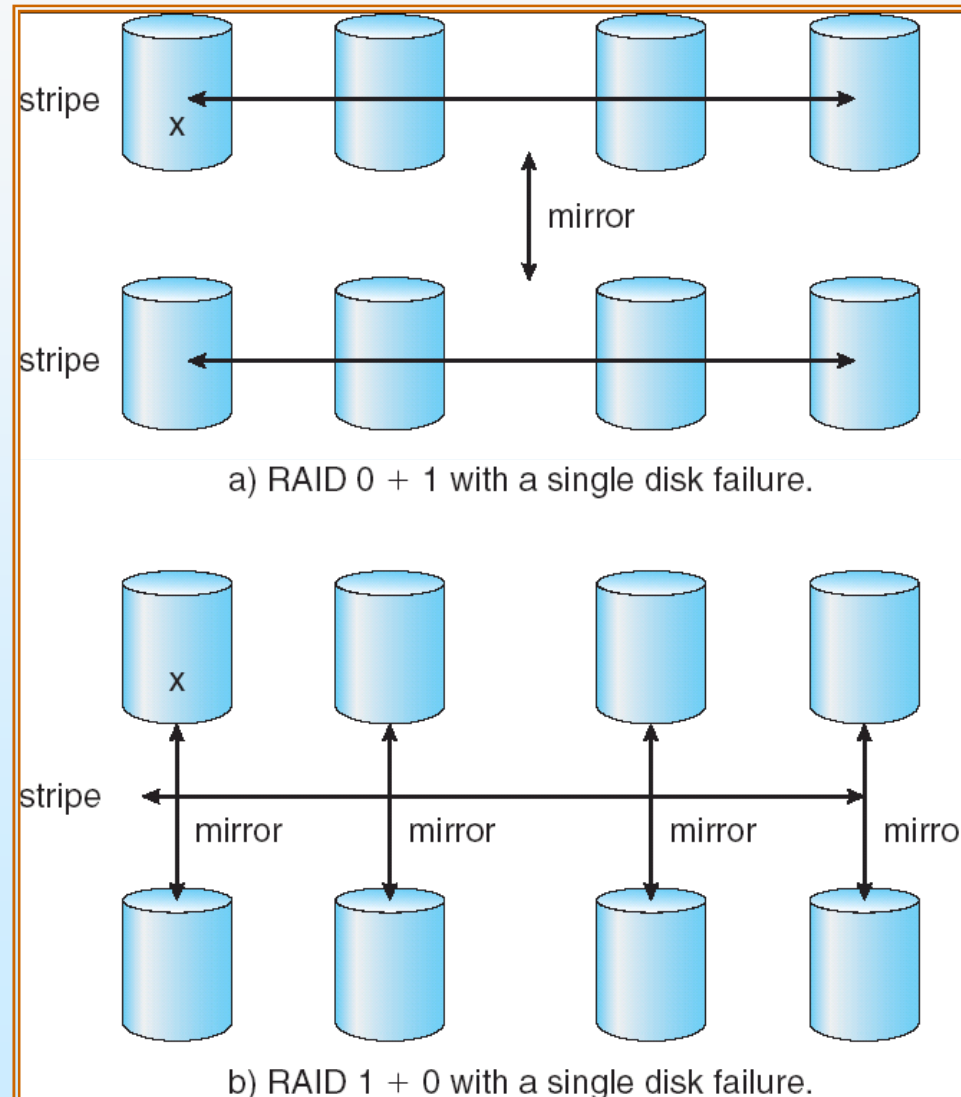
(f) RAID 5: block-interleaved distributed parity.



(g) RAID 6: P + Q redundancy.



RAID (0 + 1) and (1 + 0)



Implementazione della memoria stabile

- Lo schema di scrittura anticipata (*write-ahead*) richiede memorizzazione stabile.
- Per implementare l'archiviazione stabile bisogna:
 - replicare le informazioni necessarie su più dispositivi (solitamente dischi) con modalità di guasto indipendenti;
 - aggiornare le informazioni in modo controllato al fine di assicurare il recupero di dati stabili a seguito di un guasto durante il trasferimento o il recupero di dati



Dispositivi di memorizzazione terziaria

- Il basso costo è la caratteristica che definisce la memorizzazione terziaria.
- Generalmente, la memorizzazione terziaria viene effettuata con **supporti rimovibili**.
- Gli esempi più comuni di supporti rimovibili sono floppy disk, CD-ROM e nastri: sono disponibili anche altri tipi di dispositivi.



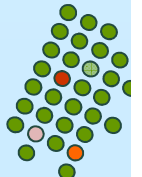
Dischi rimuovibili

- Floppy disk — sottile disco flessibile ricoperto di materiale magnetico, chiuso in una custodia di plastica protettiva.
- I dischetti possono contenere circa 1.5 MB; si usa una tecnologia simile per i dischi magnetici rimovibili che contengono fino ad 1 GB.
- I dischi magnetici rimovibili possono essere veloci quasi quanto i dischi rigidi, anche sono a più alto rischio di danneggiamento.



Dischi rimuovibili

- Un **disco magneto-ottico** registra i dati su un piatto rigido ricoperto di materiale magnetico.
 - Il calore del laser amplifica un campo magnetico grande e debole per memorizzare un bit sul disco.
 - Il raggio laser è anche usato per leggere i dati (Kerr effect).
 - La testina magneto-ottica è più lontana dalla superficie del disco rispetto alla testina di un disco magnetico, e il materiale magnetico è ricoperto da uno strato protettivo di plastica o vetro. Più resistente alla caduta delle testine.
- I **dischi ottici** non usano il magnetismo, bensì materiali speciali che possono essere alterati dal raggio laser per avere punti relativamente scuri o luminosi.



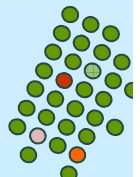
Dischi WORM

- I dati su dischi con lettura-scrittura possono essere modificati più e più volte.
- I dischi WORM (“Write Once, Read Many Times”) possono essere scritti solamente una volta.
- Sottile pellicola di alluminio interposta tra due dischi di vetro o plastica.
- Per scrivere un bit, il driver usa un raggio laser per creare un piccolo buco attraverso l’alluminio. Le informazioni possono essere distrutte ma non alterate.
- Molto durevole ed affidabile.
- I **dischi a sola lettura** (read-only), quali CD-ROM e DVD, provengono dalla fabbrica con dati preregistrati.



Nastri

- Rispetto ad un disco, un nastro è meno costoso e contiene più dati, ma l'accesso casuale è molto più lento.
- Il nastro è un mezzo economico per usi che non richiedono un accesso casuale veloce, e.g., tenere copie di riserva dei dati del disco, memorizzare grossi volumi di dati.
- Le grandi installazioni che usano nastri, utilizzano tipicamente dei robot per cambiare i nastri, che li spostano fra le unità e gli spazi di archiviazione in una libreria di nastri.
- stacker – libreria che immagazzina alcuni nastri.
- silo – libreria che immagazzina migliaia di nastri.
- Un file residente su disco, che non sarà più necessario per un po' di tempo, può venire **archiviato** su nastro; il computer potrà **ricaricarlo sul** disco per un uso attivo.



Compiti del sistema operativo

- Due importanti compiti di un sistema operativo sono la gestione dei dispositivi fisici e la presentazione alle applicazioni di un'astrazione di macchina virtuale.
- Per i dischi, il sistema operativo fornisce due astrazioni:
 - dispositivo grezzo – un array di blocchi di dati;
 - file system – il sistema operativo accoda e schedula le richieste provenienti da diverse applicazioni.

