Gestione della memoria secondaria

Memoria secondaria

- Struttura del disco
- Scheduling del disco
- Gestione del disco
- Gestione dello spazio di swap
- Struttura RAID
- Collegamento dei dischi
- Memoria stabile
- Costi

Dischi magnetici

- I **dischi magnetici** forniscono grandi quantità di storage secondario nei moderni computer
  - I drive ruotano tra 60 e 200 volte al secondo
  - Il **transfer rate** è la velocità alla quale i dati vengono trasferiti dal drive al computer
  - Il **tempo di accesso** è il tempo necessario a muovere il braccio sul cilindro desiderato, più quello necessario al settore desiderato a ruotare sotto la testina
  - Un **head crash** si verifica quando la testina tocca la superficie del disco (causa una rottura del disco)
- I dischi possono essere rimovibili
- I drive si collegano al computer attraverso l'**I/O bus**
  - Esistono diversi tipi di bus, tra cui **EIDE**, **ATA**, **SATA**, **USB**, **Fibre Channel**, **SCSI**
  - L'**host controller** nel computer usa il bus per comunicare con il **disk controller** che si trova all'interno del drive

Nastri magnetici

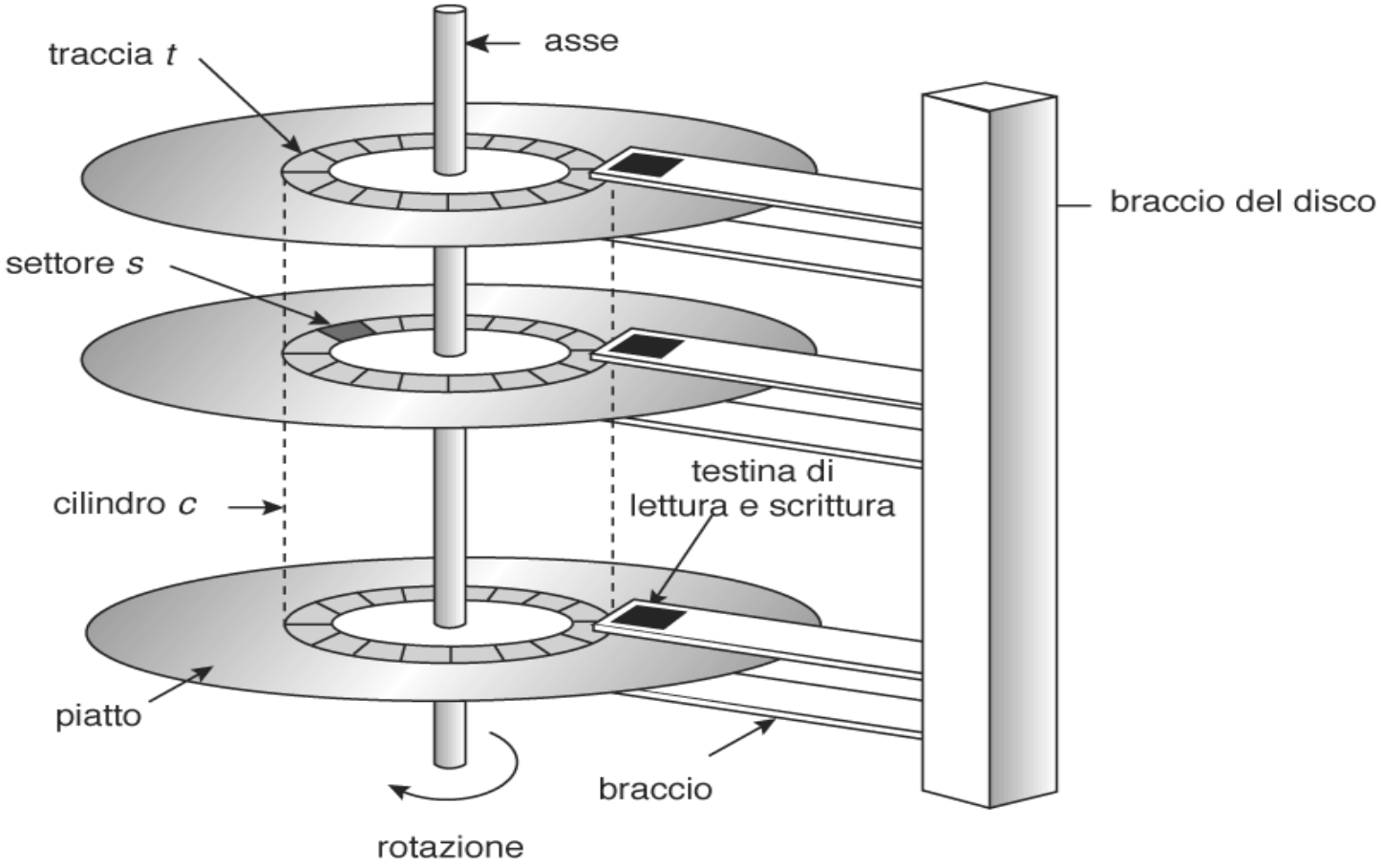
- **Nastro magnetico**
  - E’ stato uno dei primi supporti di memoria secondaria
  - Relativamente permanente e in grado di contenere grandi quantità di dati
  - Tempo di accesso elevato
  - Accesso casuale ~ 1000 volte più alto di quello del disco
  - Dopo che i dati sono sotto la testina, la velocità di trasferimento è paragonabile a quella di un disco
  - Principalmente usato per il backup, per archiviazione di dati utilizzati di rado, e per il trasferimento di dati tra sistemi

    .

Dischi ottici

- Lettura ottica basata sulla riflessione (o sulla mancata riflessione) di un raggio laser.
- Densità di registrazione più alta dei dischi magnetici.
- Creati in origine per registrare i programmi televisivi, poi usati come dispositivi di memoria nei calcolatori.
- Diversi tipi/caratteristiche
  - CD-ROM
  - CD-R
  - CD-RW
  - DVD
  - DVD-RAM
  - …

Struttura fisica di un disco magnetico



Struttura logica del disco

- I dischi sono gestiti come array mono-dimensionali di blocchi logici che costituiscono l'unità elementare di trasferimento.

- L'array mono-dimensionale di blocchi logici è realizzato sequenzialmente nei settori del disco.
  - Il settore 0 è il primo settore della prima traccia del cilindro più esterno.
  - Quindi si prosegue con le tracce del cilindro e poi nei cilindri più interni.
  - I cilindri più esterni contengono più settori dei cilindri interni.

Scheduling del disco

- Il sistema operativo è responsabile della gestione efficiente del disco: ***tempi di accesso*** bassi e ***ampiezza di banda*** (numero di bit trasferiti / tempo di trasferimento) **elevata**.
- Tempo di accesso:
  - *Seek time* - per lo spostamento delle testine sul cilindro del settore interessato.
  - *Rotational latency* - tempo di rotazione del disco per portare le testine sul settore.
- Obiettivo dello scheduling: **minimizzare il *seek time* e *massimizzare l'ampiezza di banda***.
- Seek time $\approx$ distanza di seek

Scheduling del disco

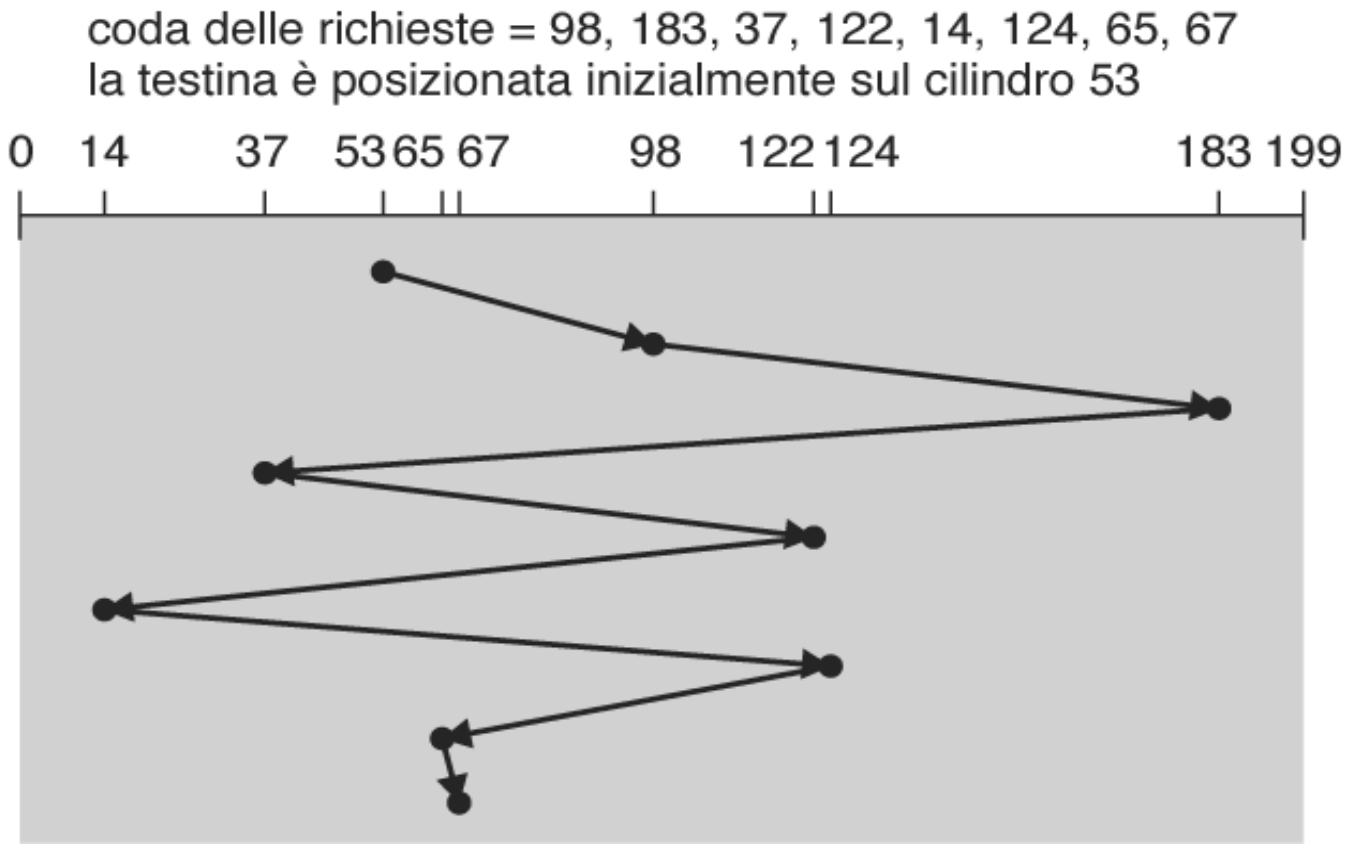
- Esistono numerosi algoritmi per schedulare le richieste di accesso al disco.

- Per valutarli usiamo la seguente sequenza di cilindri su cui si trovano i settori richiesti:

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

Posizione iniziale della testina: ***53***

Scheduling FCFS

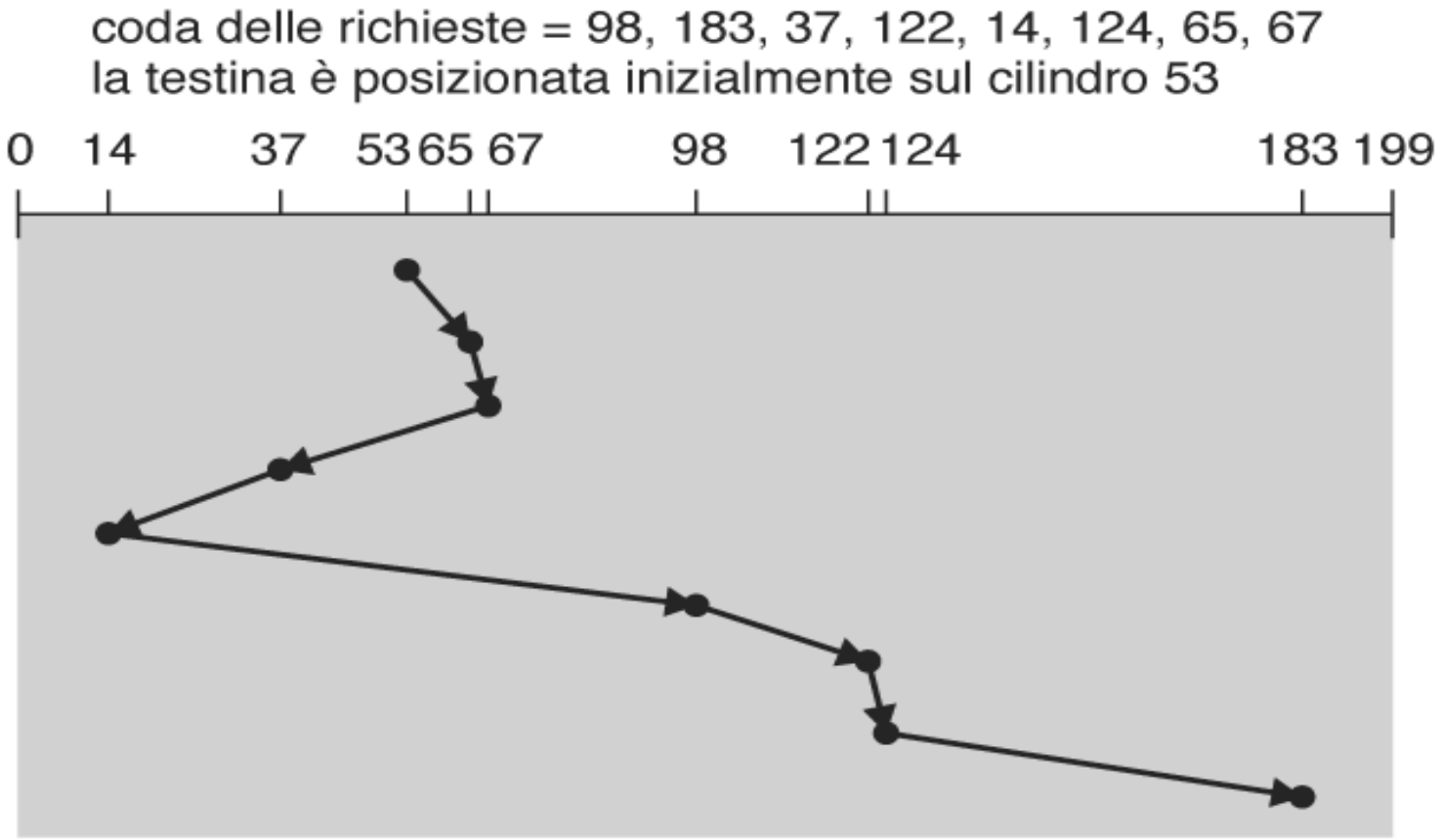


Movimento totale della testina: **640** cilindri.

Scheduling SSTF

- **SSTF: Shortest Seek Time First**.
- Seleziona la richiesta che comporta il minimo tempo di *seek* dalla posizione corrente della testina.
- Lo scheduling SSTF è una forma di scheduling SJF.
- Può causare *starvation*.

Scheduling SSTF



Movimento totale della testina: **236** cilindri.

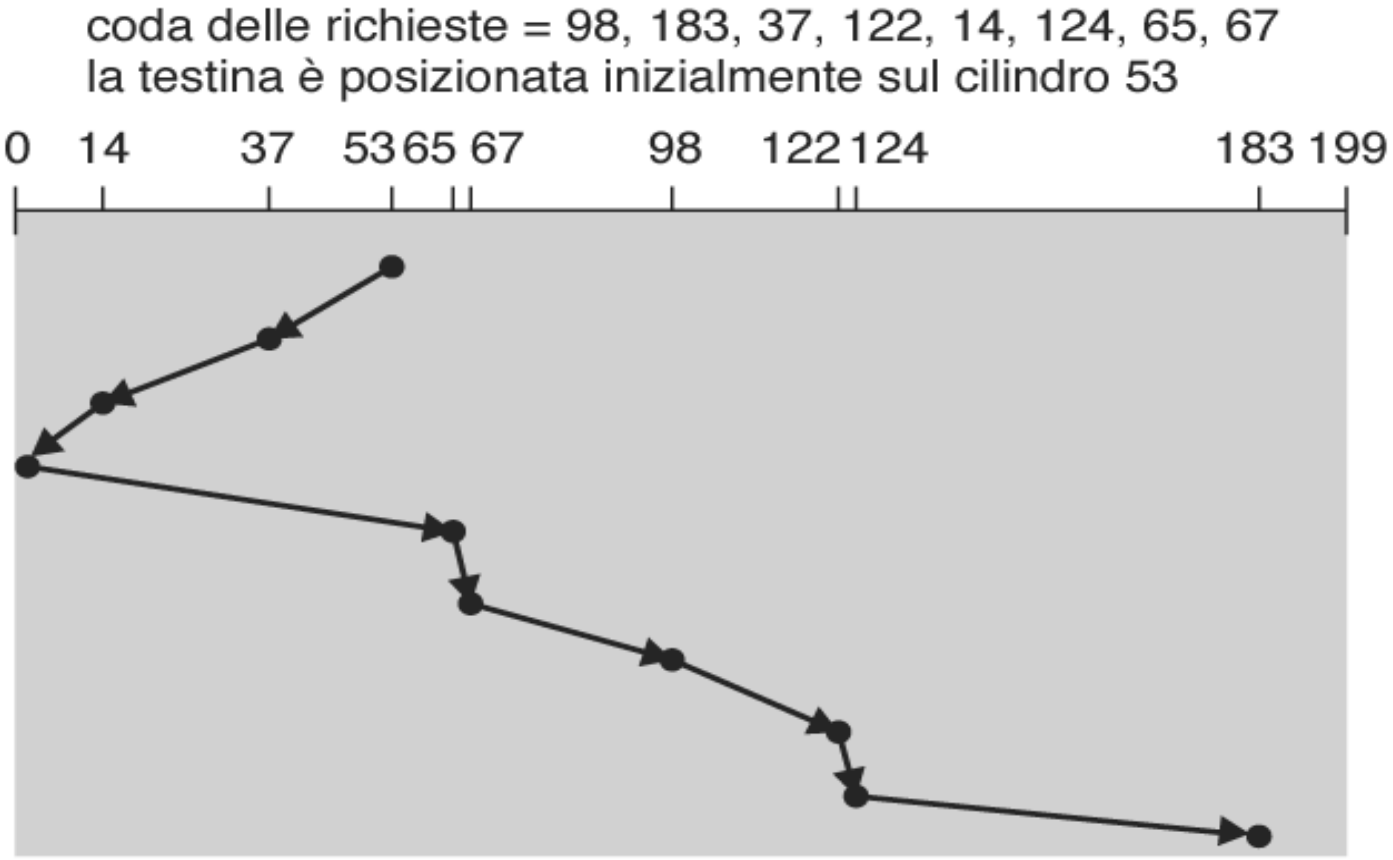
Scheduling SCAN

- La testina parte da un estremo del disco e muovendosi fino all'altro estremo serve tutte le richieste di blocchi che si trovano lungo il percorso, quindi inverte la marcia e fa lo stesso nell'altra direzione.

- Una nuova richiesta per un blocco che si trova davanti alla testina viene subito servita.

- E' anche chiamato ***algoritmo dell'ascensore***.

Scheduling SCAN

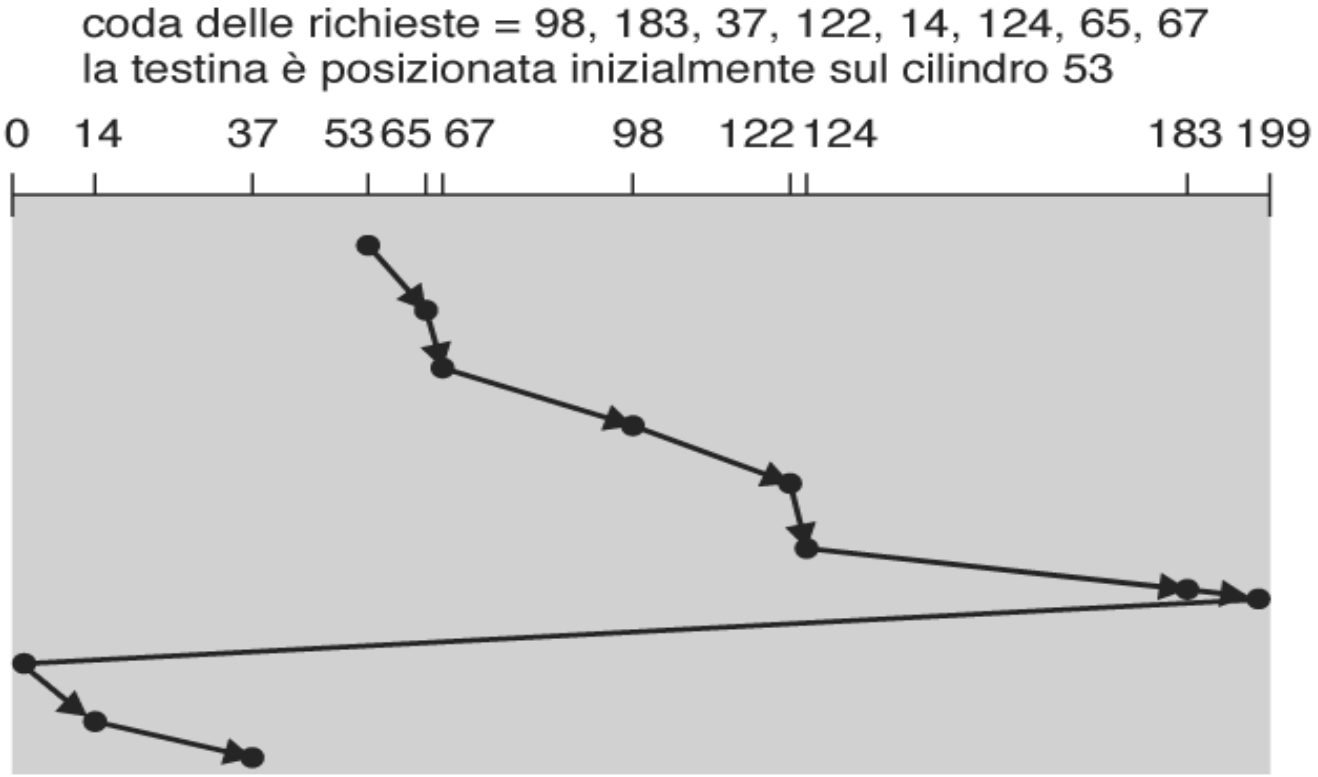


Movimento totale della testina: **236** cilindri.

Scheduling C-SCAN

- Variante di SCAN che offre un tempo di attesa più uniforme.
- La testina parte da un estremo del disco e muovendosi fino all'altro estremo serve tutte le richieste di blocchi che si trovano lungo il percorso, quindi ritorna all'altro estremo del disco (senza servire le richieste nello spostamento) per ripartire da lì.
- Tratta i cilindri come una lista circolare con l'ultimo collegato al primo.

Scheduling C-SCAN

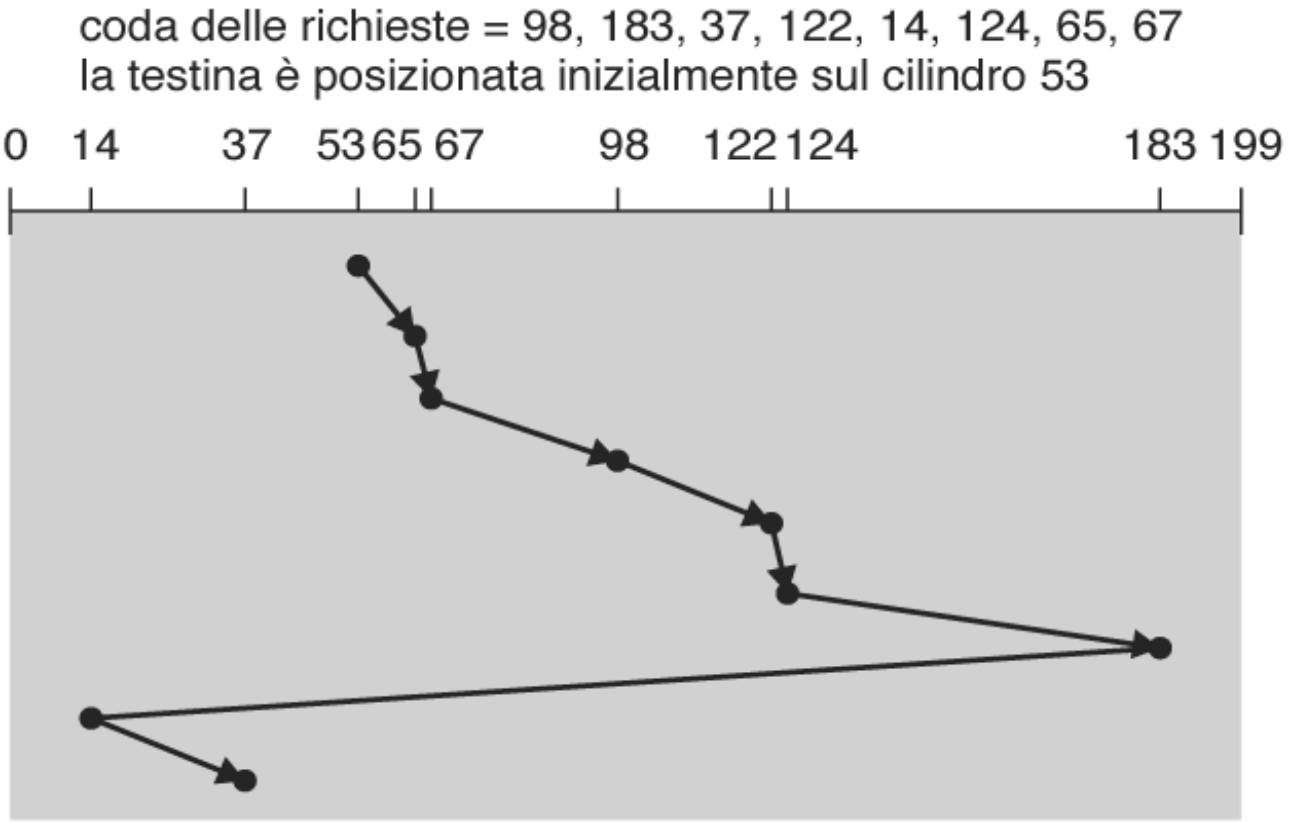


Movimento totale della testina: **183** cilindri (escluso il movimento per riportare la testina all'inizio, che non si conta, ma ha un tempo di latenza).

Scheduling C-LOOK

- Variante di C-SCAN
- La testina viene spostata non fino alla fine del disco ma solo fino a che ci sono richieste in quella direzione.

Scheduling C-LOOK



Movimento totale della testina: **153** cilindri (escluso il movimento per riportare la testina all'inizio).

Selezione dell’algoritmo di scheduling

- SSTF è usato comunemente.
- SCAN e C-SCAN sono migliori per sistemi che fanno un uso intensivo dei dischi.
- Le prestazioni dipendono dal numero e dal tipo delle richieste.
- Le richieste possono essere influenzate dal metodo di allocazione dei file.
- Gli algoritmi SSTF e LOOK sono degli algoritmi di default ragionevoli.
- L’algoritmo di scheduling del disco dovrebbe essere un modulo a se stante del sistema operativo.

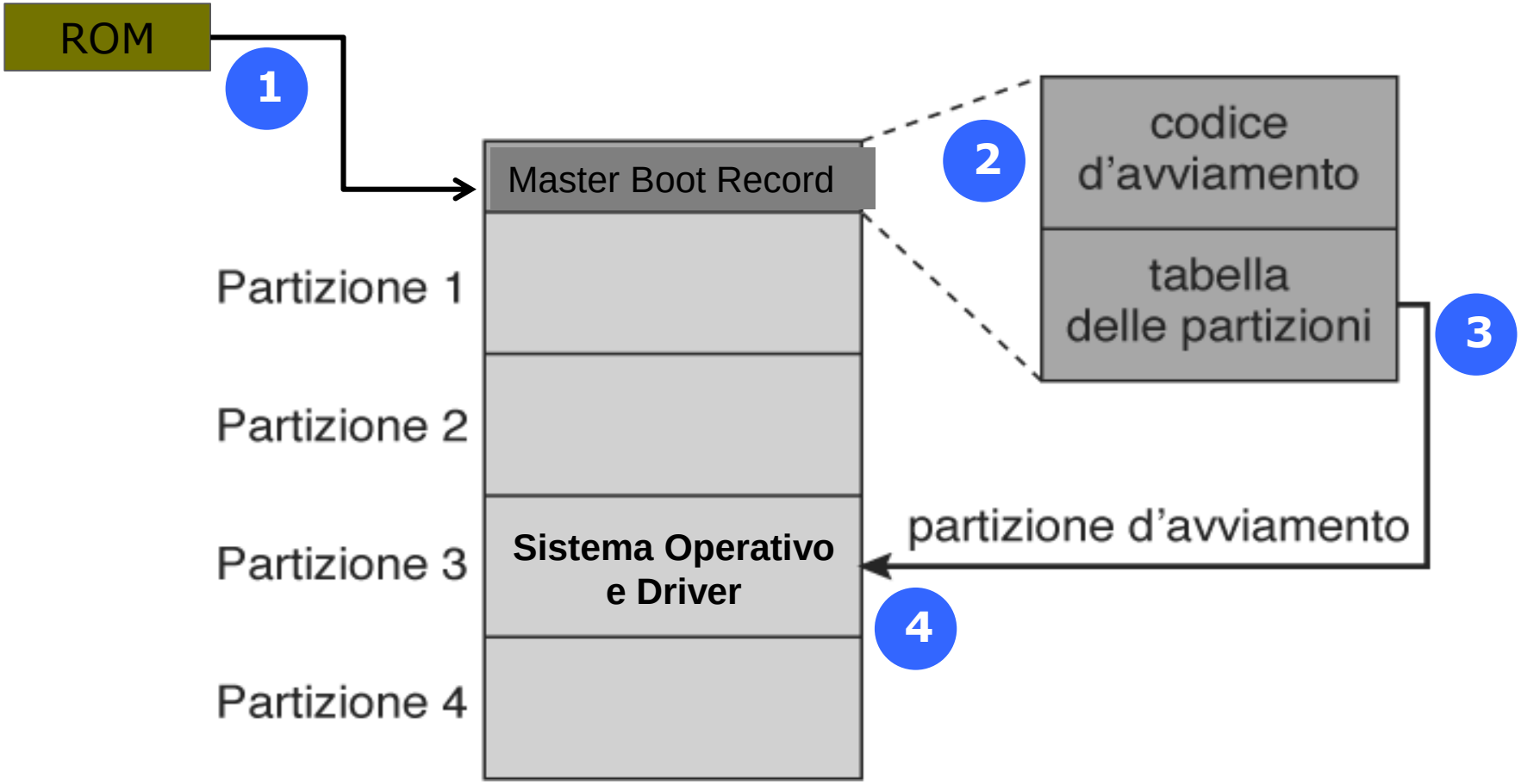
Gestione del disco

- *Formattazione di basso livello o fisica* — divisione del disco in settori che il controller può leggere o scrivere.

- Per usare un disco come contenitore di file, il S.O. ha bisogno di memorizzare le proprie strutture sul disco:
  - *Partizionamento del disco* in uno o più gruppi di cilindri (partizioni).
  - *Formattazione logica* per creare un file system.

- Il programma di boot inizializza il sistema.
  - Programma di bootstrap memorizzato nella ROM.
  - *Bootstrap loader* nella ROM per caricare il bootstrap dal disco (dai blocchi di boot).

Avviamento dal disco (esempio Windows)

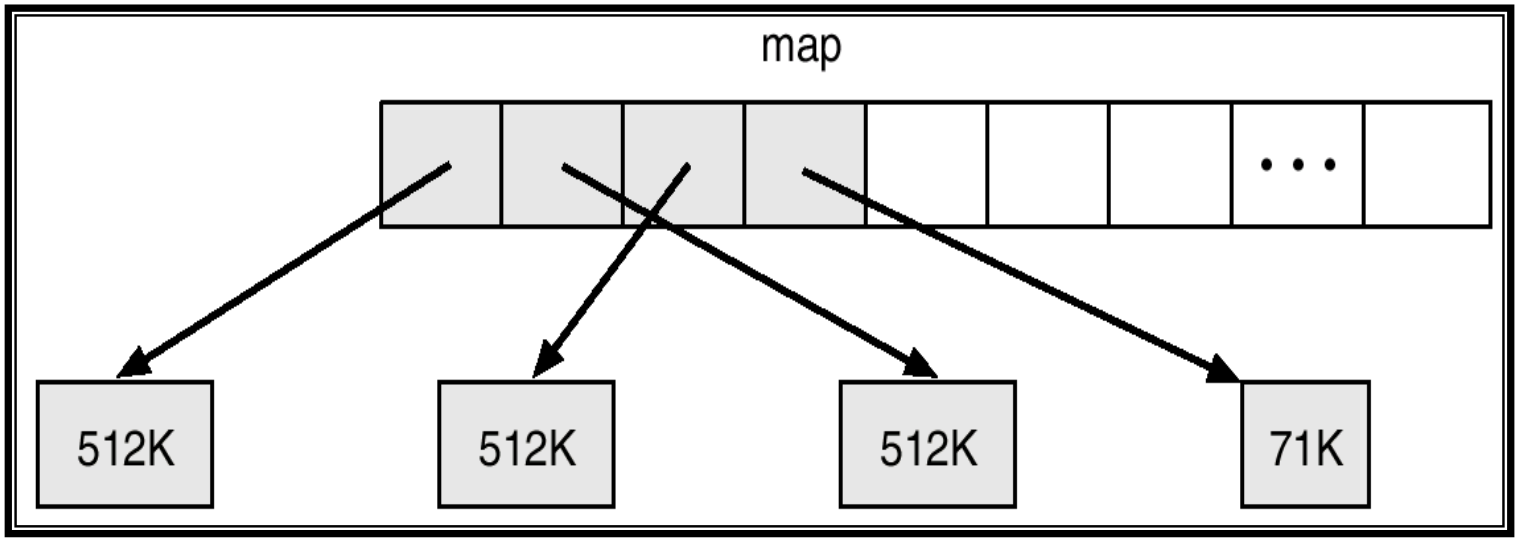


Gestione dello spazio di swap

- La memoria virtuale usa spazio sul disco come estensione della memoria centrale (**spazio di swap**).
- Lo spazio di swap può essere sullo stesso spazio usato dal file system (meno efficiente) o su una diversa partizione del disco (più efficiente). A volte si usano ambedue.
- Gestione dello spazio di swap:
  - UNIX 4.3BSD alloca lo spazio di swap quando un processo è attivato; con *text segment* (il programma) e *data segment* (per i dati).
  - Il kernel usa le *mappe di swap* per tenere traccia dell'uso dello spazio di swap.
  - Solaris 2 alloca spazio di swap su necessità.

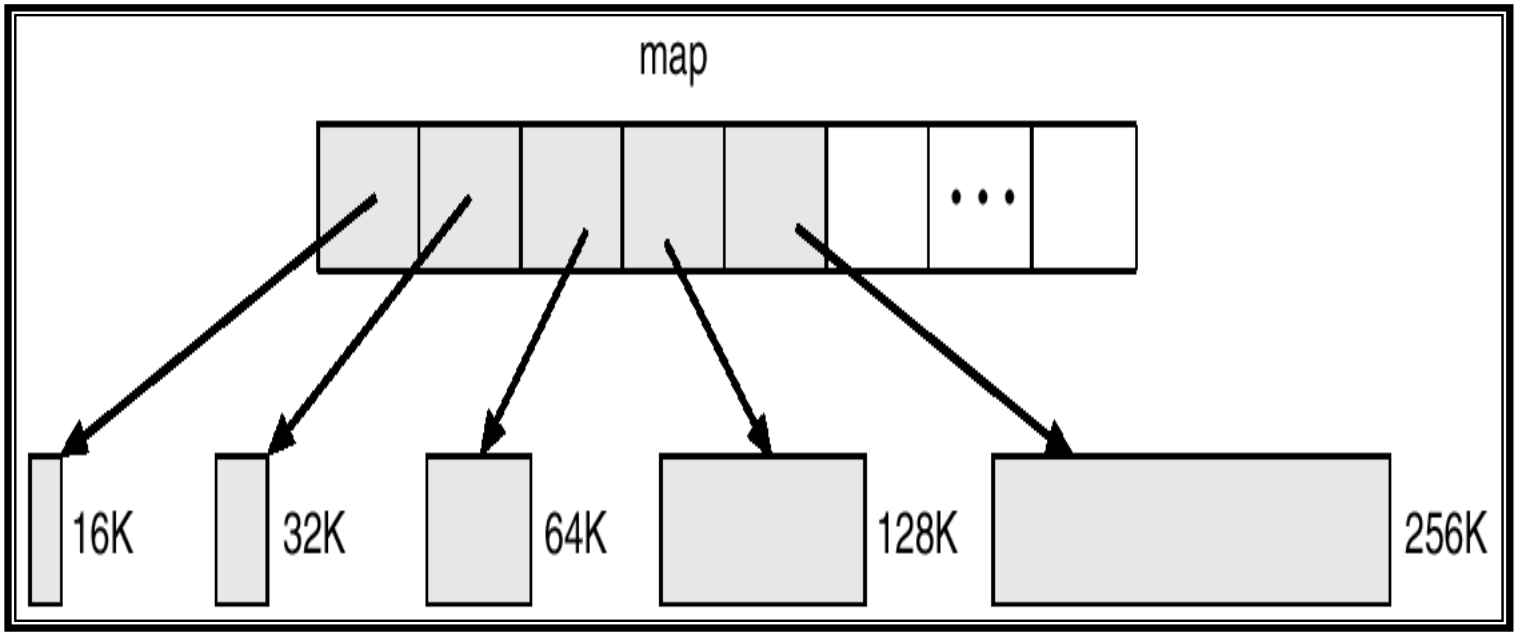
Mappa di swap in UNIX 4.3 BSD: testo

- Lo **spazio di swap** del segmento di testo è allocato in blocchi di 512 K tranne per l'ultimo blocco.



Mappa di swap in UNIX 4.3 BSD: dati

- Lo **spazio di swap** del segmento dati è allocato in blocchi di dimensione variabile (multipli di 16 K).
- Per processi piccoli si usano blocchi piccoli, per processi grandi si allocano blocchi di dimensione sempre maggiore.



Dischi RAID

- Alcune tecniche efficienti per la gestione dei dischi si basano sull'uso di un insieme di dischi su cui si opera contemporaneamente.

- Quando serve memorizzare grandi moli di dati e/o mantenere una elevata affidabilità della memoria secondaria si può usare una batteria di dischi detta **struttura RAID**.

- **RAID** (*redundant array of indipendent disks*).

- Una struttura RAID funziona come una ***singola unità*** di memoria secondaria.

Dischi RAID

- I dischi RAID migliorano le prestazioni e l'affidabilità della memoria secondaria usando dati replicati.

- Esempi:
  - *Mirroring* o *shadowing* : ogni disco viene replicato.
    (migliore affidabilità e maggiore velocità).

  - *Block interleaved parity* : blocchi di parità (contengono bit di controllo / **bit di parità**).
    (affidabilità e minore spazio di replica).

- Una struttura RAID può essere organizzata in sette livelli differenti.

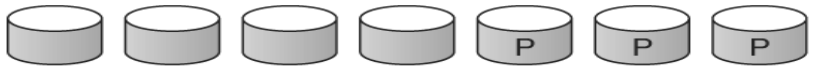
Livelli RAID

(a) RAID 0: sezionamento senza ridondanza

RAID 0 divide i dati equamente tra due o più dischi con nessuna informazione di parità o ridondanza.

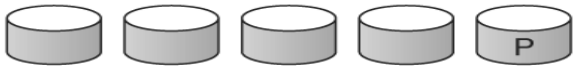
(b) RAID 1: copiatura speculare

RAID 1 crea una copia esatta (o mirror) di tutti i dati su più dischi.



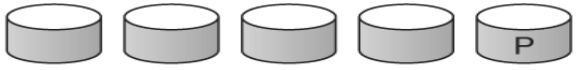
(c) RAID 2: codici per la correzione degli errori

RAID 2 divide i dati al livello di bit invece che di blocco con dischi di parità.



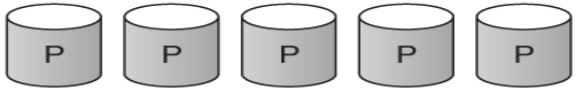
(d) RAID 3: bit di parità intercalati

RAID 3 usa una divisione al livello di byte con un disco dedicato alla parità.



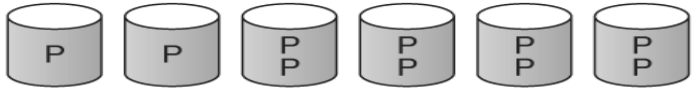
(e) RAID 4: blocchi di parità intercalati

RAID 4 usa una divisione (striping) a livello di blocchi con un disco dedicato alla parità.



(f) RAID 5: blocchi intercalati a parità distribuita

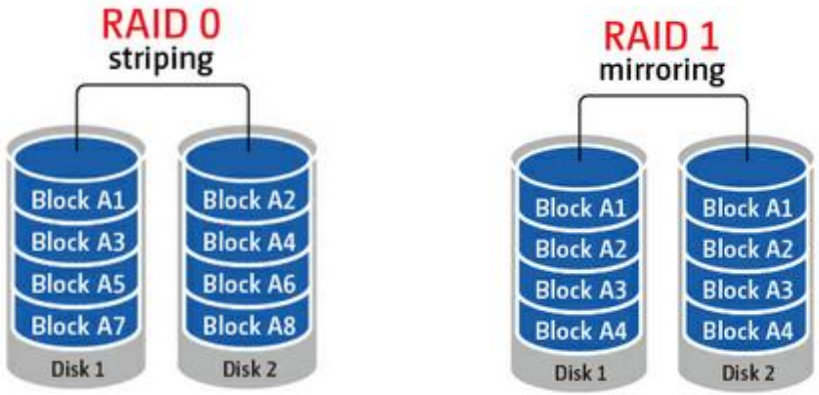
RAID 5 usa una divisione dei dati a livello di blocco con i dati di parità distribuiti tra tutti i dischi appartenenti al RAID.

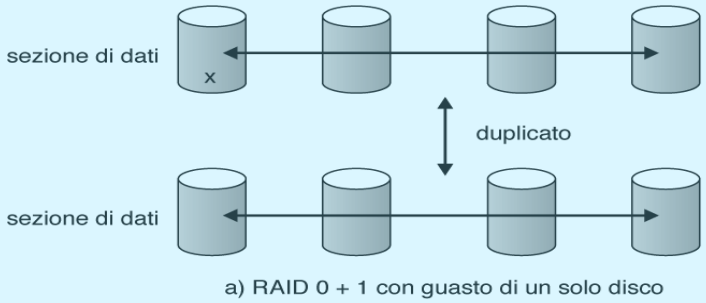


(g) RAID 6: ridondanza P + Q

RAID 6 usa una divisione a livello di blocchi con i dati di parità distribuiti due volte tra tutti i dischi.

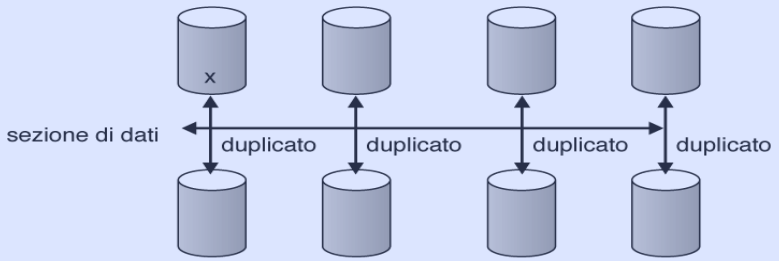
RAID (0 + 1) e (1 + 0)





a) RAID 0 + 1 con guasto di un solo disco

Un sistema **RAID 0+1** è un RAID che viene usato sia per replicare che per condividere dati tra diversi dischi.



b) RAID 1 + 0 con guasto di un solo disco

Un sistema **RAID 1+0**, è simile al RAID 0+1 ma i livelli RAID sono usati in senso invertito.

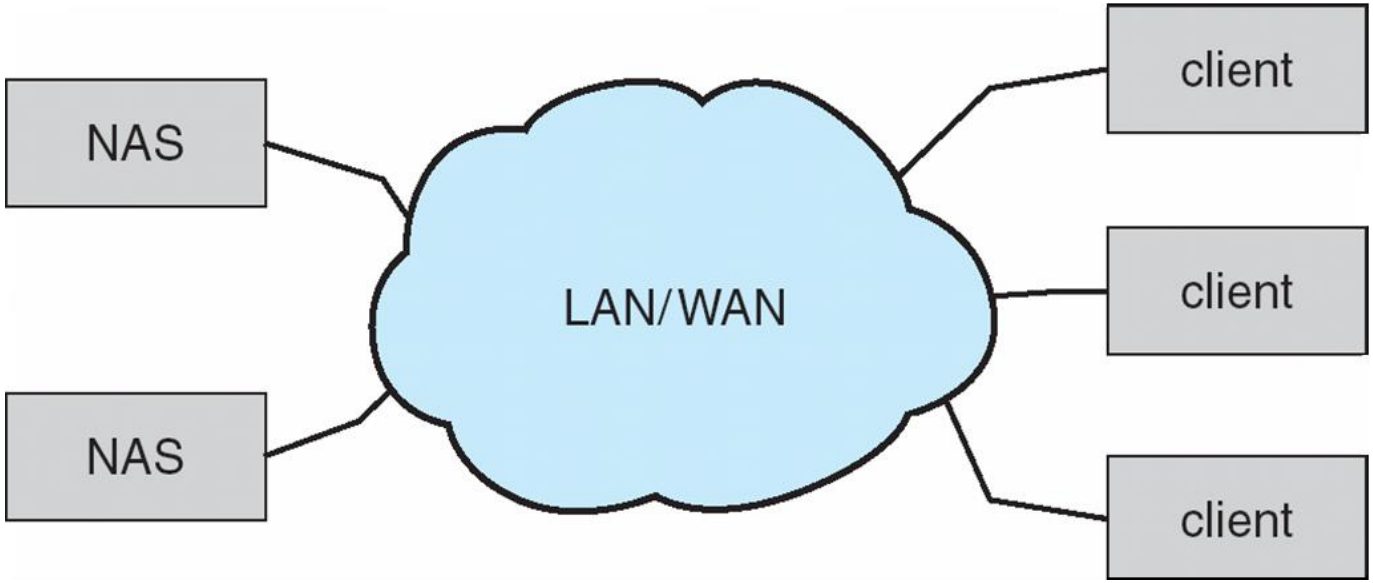
Collegamento dei dischi

- In un sistema di calcolo i dischi possono essere connessi in due modi principali:

1. **Collegati ad un host (PC/workstation)**
   - attraverso una porta di I/O.
2. **Collegati alla rete**
   - attraverso una connessione di rete.

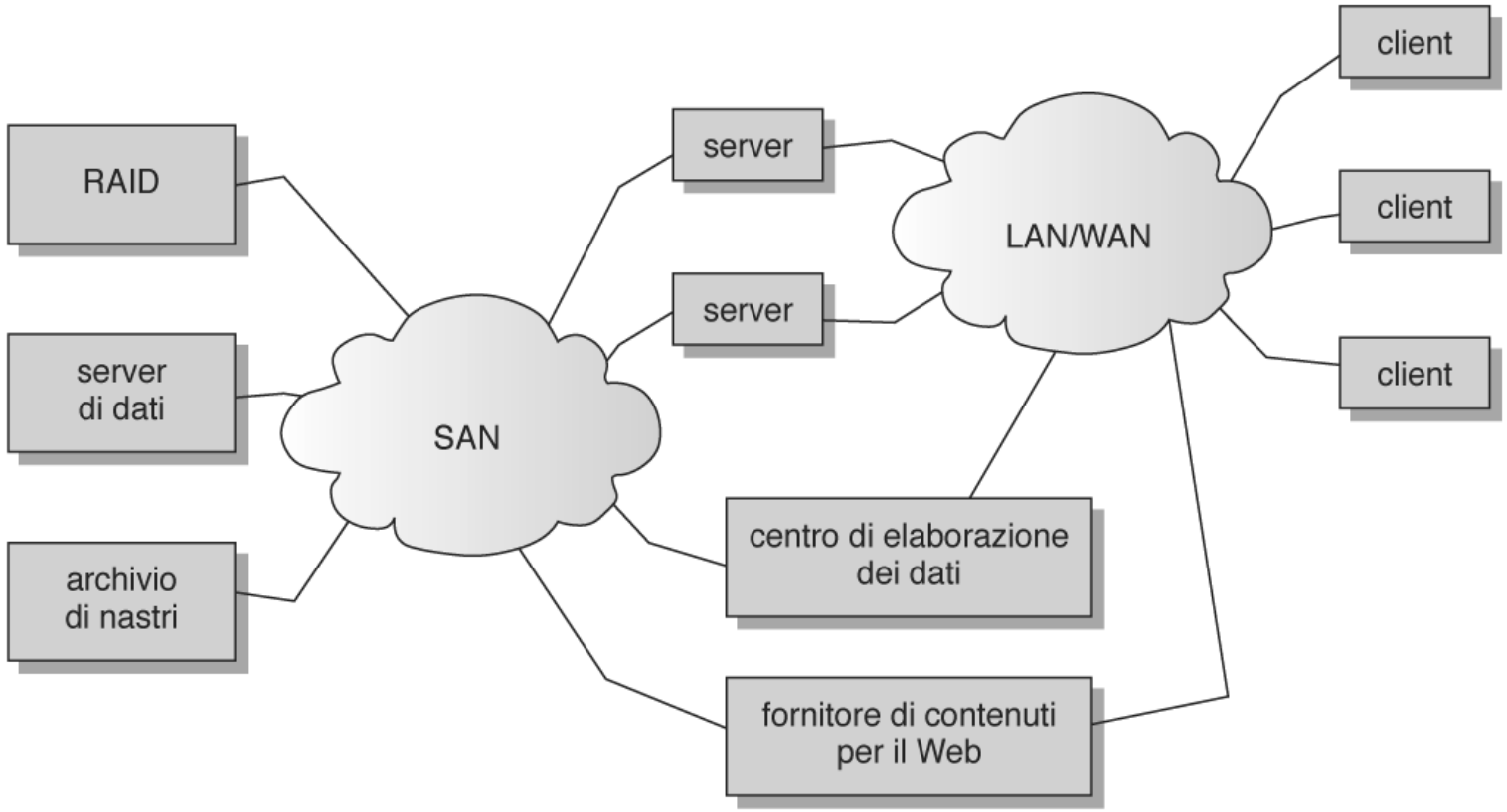
Memoria secondaria connessa alla rete

- **Network-attached storage** (**NAS**) è uno storage disponibile su rete anzichè su una connessione locale (come un bus)
- NFS è un tipico protocollo NAS
- Implementato mediante chiamate di procedura remota (RPCs) tra l'host e lo storage



Storage-Area Network (SAN)

- Comune in ambienti di storage di grandi dimensioni
- Più host sono collegati a più dispositivi di storage



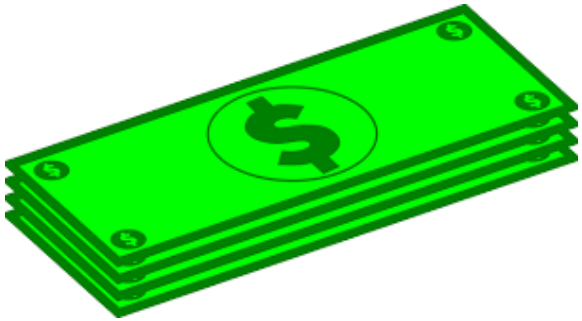
Implementazione della memoria stabile

- La memoria stabile garantisce che non ci siano mai perdite di dati.

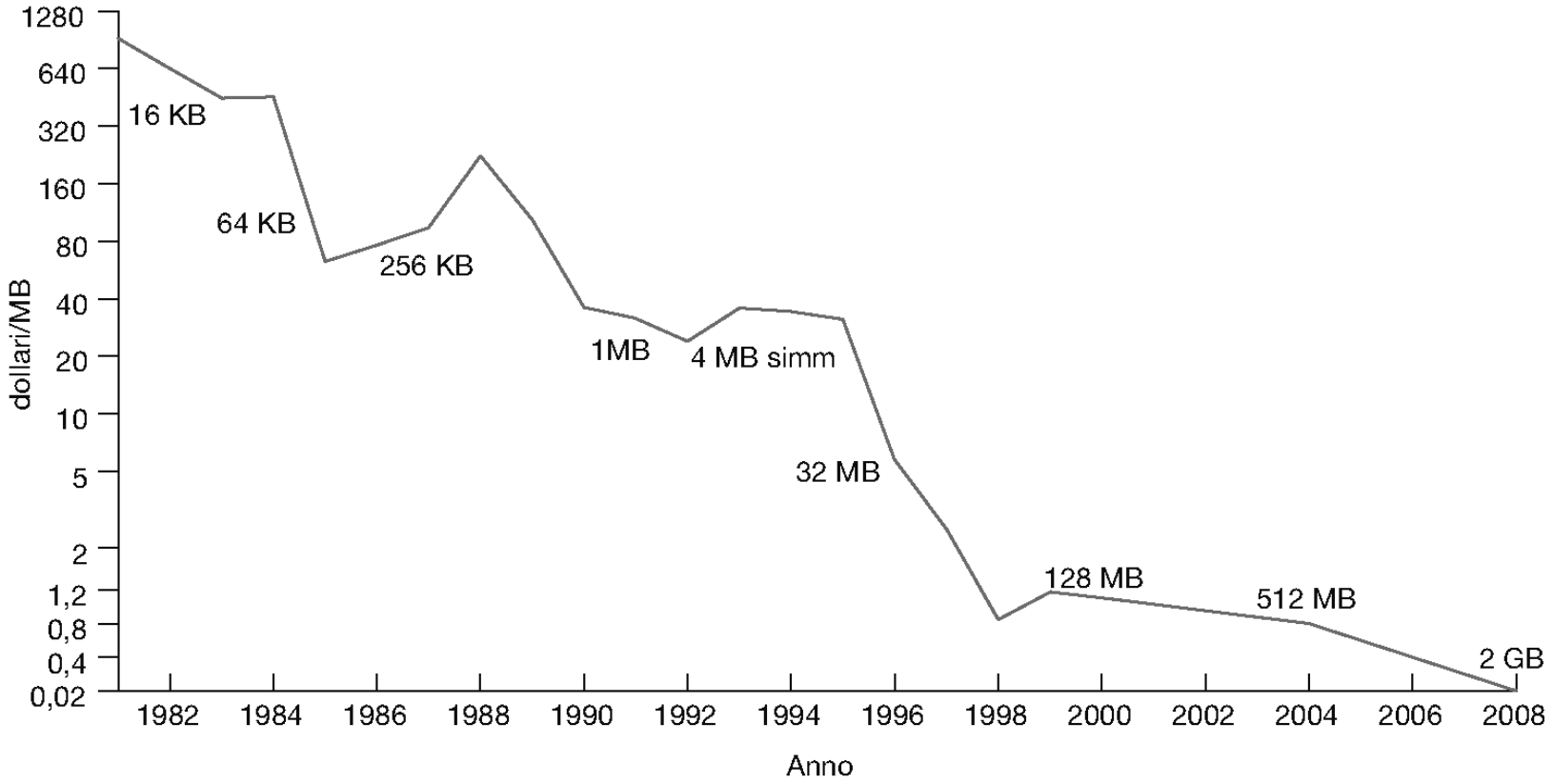
- Per implementare la memoria stabile:
  - Replicare l'informazione su più supporti di memoria non volatile che abbiano modalità di guasto differenti e indipendenti.
  - Aggiornare l'informazione in maniera controllata per assicurare che si possano ottenere i dati stabili dopo un fallimento:
    - sia durante il trasferimento dei dati
    - sia durante un ripristino.

Costi

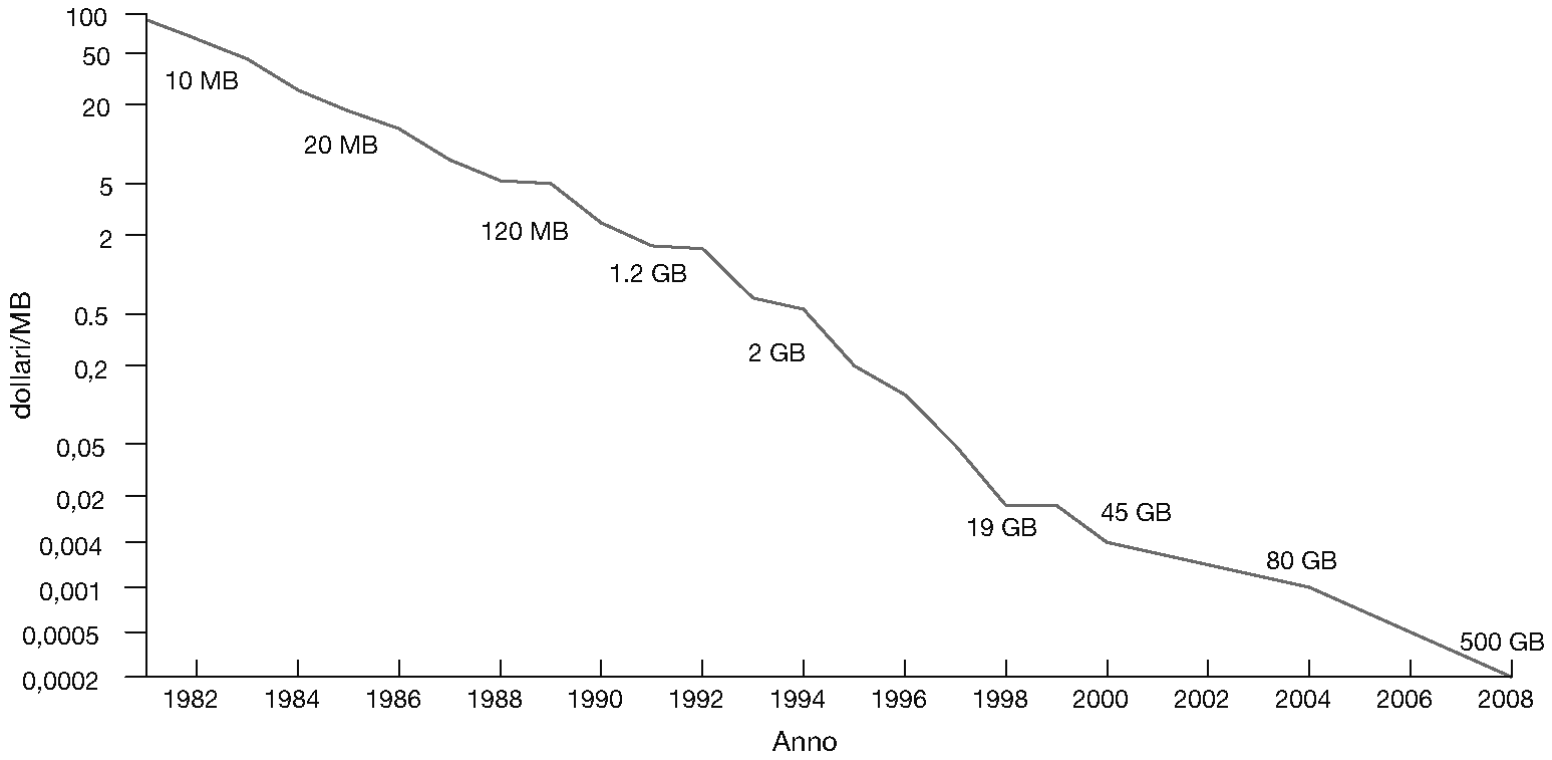
- La memoria principale è più costosa della memoria secondaria.
- Il costo per megabyte di memoria su disco è competitivo con quello della memoria su nastro.
- I nastri vengono generalmente usati per la cosiddetta **memoria terziaria**.



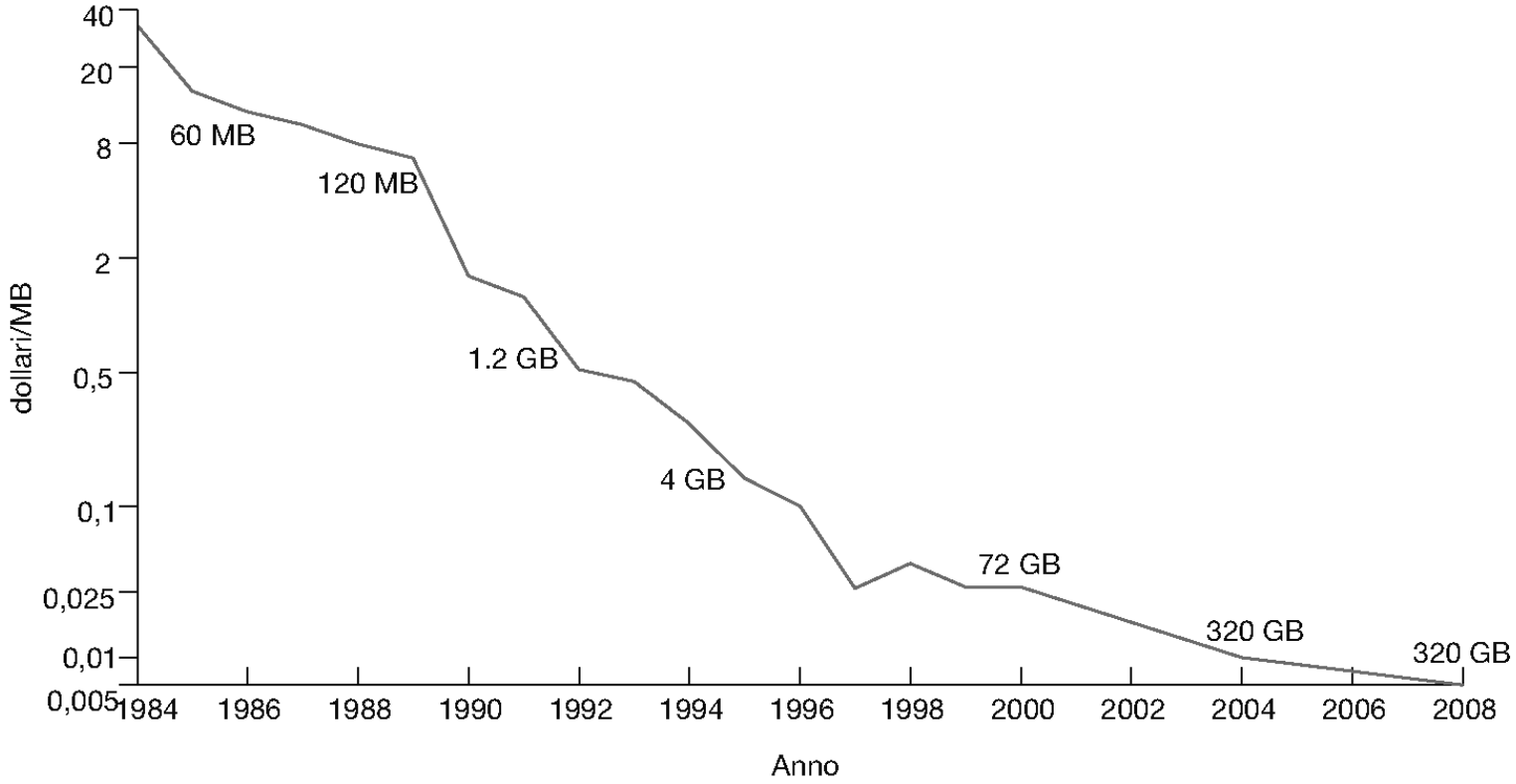
Prezzo al MB della memoria RAM, dal 1981 al 2008



Prezzo al MB dei dischi magnetici, dal 1981 al 2008



Prezzo al MB delle unità a nastro, dal 1984 al 2008



Dischi/Drive a Stato Solido (SSD)

- **Solid state drive**: I dischi a stato solido non hanno le stesse caratteristiche dei dischi rigidi che abbiamo discusso finora.
- Realizzano una memoria non volatile, ma non hanno parti in movimento e non richiedono tenpi di ricerca.
- Consumano meno energia, hanno meno capacità di memorizzazione e ne viene fatto un uso più limitato.
- Si realizzano tramite memorie flash e DRAM con batteria.