



Informatica di base

Memoria



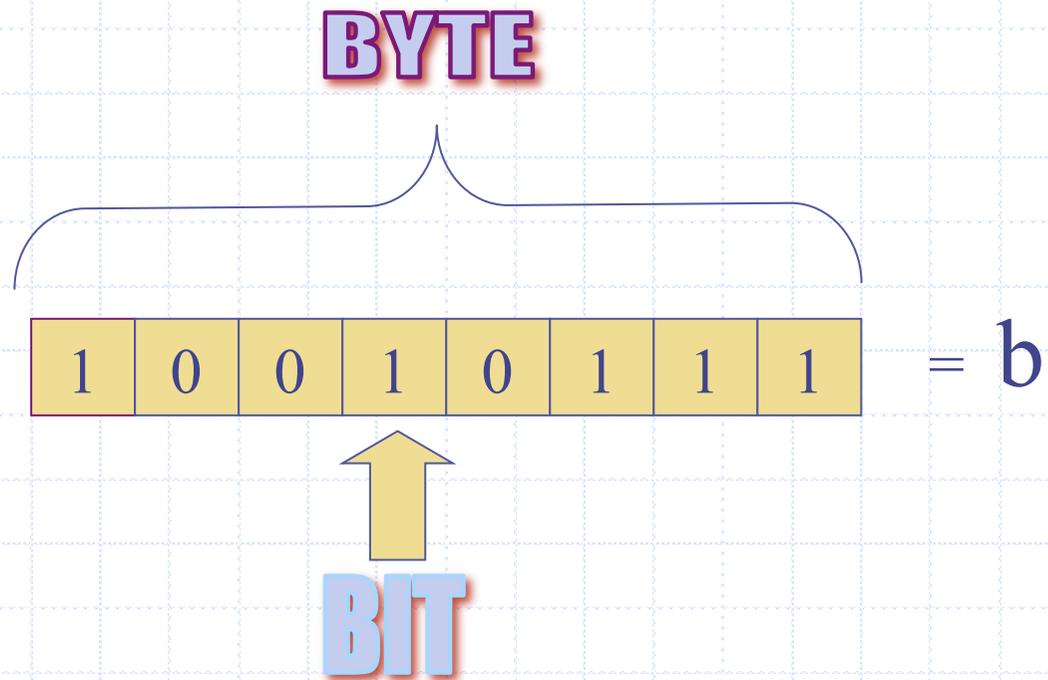
La Memoria

Massimo Marchi
<http://marchi.usr.dsi.unimi.it/>

La memoria

- ◆ Svolge la funzione di magazzino per i programmi e i dati su cui deve operare il microprocessore
- ◆ Dati e programmi sono memorizzati in unità base chiamate **byte**
- ◆ Un byte è a sua volta costituito da **8 bit**
- ◆ Un byte può memorizzare uno di 256 valori (2^8 valori)

La memoria



Il binary digit (bit)

- ◆ L'unità di informazione minima è il bit (bi-nary digi-t). Consideriamola come una casella nella quale possiamo scrivere uno di due valori.
- ◆ Il valore di un bit può essere rappresentato come:
 - ◆ 0 – 1
 - ◆ Vero – Falso
 - ◆ Accesso – Spento
 - ◆ Su – Giù
 - ◆ 0Volt - +5Volt
 - ◆ -10Volt - +10Volt

Sequenze di bit

- ◆ Avendo a disposizione un solo bit si possono rappresentare due elementi diversi:
 - Es.: si assegna al primo elemento la codifica 0 e al secondo la codifica 1
- ◆ Avendo a disposizione due bit si possono rappresentare quattro elementi diversi, assegnando a ciascuno una codifica diversa:
 - Es.: 00, 01, 10, 11

Sequenze di bit

- ◆ Avendo a disposizione tre bit si possono rappresentare otto elementi diversi.
- ◆
- ◆ Avendo a disposizione n bit si possono rappresentare 2^n elementi diversi.

La storia del Byte

- ◆ Agli albori dell'informatica c'era l'esigenza di rappresentare i caratteri stampabili (le lettere dell'alfabeto, le cifre da 0 a 9, lo spazio, la virgola etc...) per un numero totale di caratteri compreso tra 90 e 120.
- ◆ Per poter rappresentare 128 elementi diversi servono almeno 7 bit ($2^7=128$).
- ◆ E' bene avere a disposizione un bit supplementare poter effettuare controlli sulla sequenza

La storia del Byte

- ◆ Quindi per rappresentare i caratteri stampabili è necessaria una sequenza di $7+1=8$ bit.
- ◆ Una sequenza di 8 bit è detta byte, ed è diventata una unità di misura della occupazione di memoria.

Le celle di memoria

- ◆ Ogni cella di memoria contiene un byte, ed è contraddistinta da un indirizzo
- ◆ Volendo *scrivere* o *leggere* un dato dalla memoria il processore deve sempre specificare l'indirizzo della cella alla quale vuole accedere

Unità di misura per la memoria

Le dimensioni di una memoria si misurano in:

- ◆ Kbyte (kilo byte) = 2^{10} byte = 1024 byte $\approx 10^3$ byte
- ◆ Mbyte (Mega byte) = 2^{10} Kbyte = 1024 Kbyte = 2^{20} byte $\approx 10^6$ byte
- ◆ Gbyte (Giga byte) = 2^{10} Mbyte = 1024 Mbyte = 2^{30} byte $\approx 10^9$ byte
- ◆ Tbyte (Tera byte) = 2^{10} Gbyte = 1024 Gbyte = 2^{40} byte $\approx 10^{12}$ byte

La gerarchia di memoria

- ◆ Per questioni di costi ed efficienza la memoria di un elaboratore è distribuita su dispositivi hardware diversi (RAM, dischi), che nel loro insieme costituiscono il sistema memoria dell'elaboratore, più comunemente chiamato

gerarchia di memoria

La gerarchia di memoria

- ◆ Ai livelli più alti di questa gerarchia sono presenti le memorie più veloci, e quindi più costose. Ai livelli bassi sono presenti le memorie più economiche ma anche più lente
- ◆ La velocità di una memoria è misurata in base al tempo speso dal microprocessore per accedervi

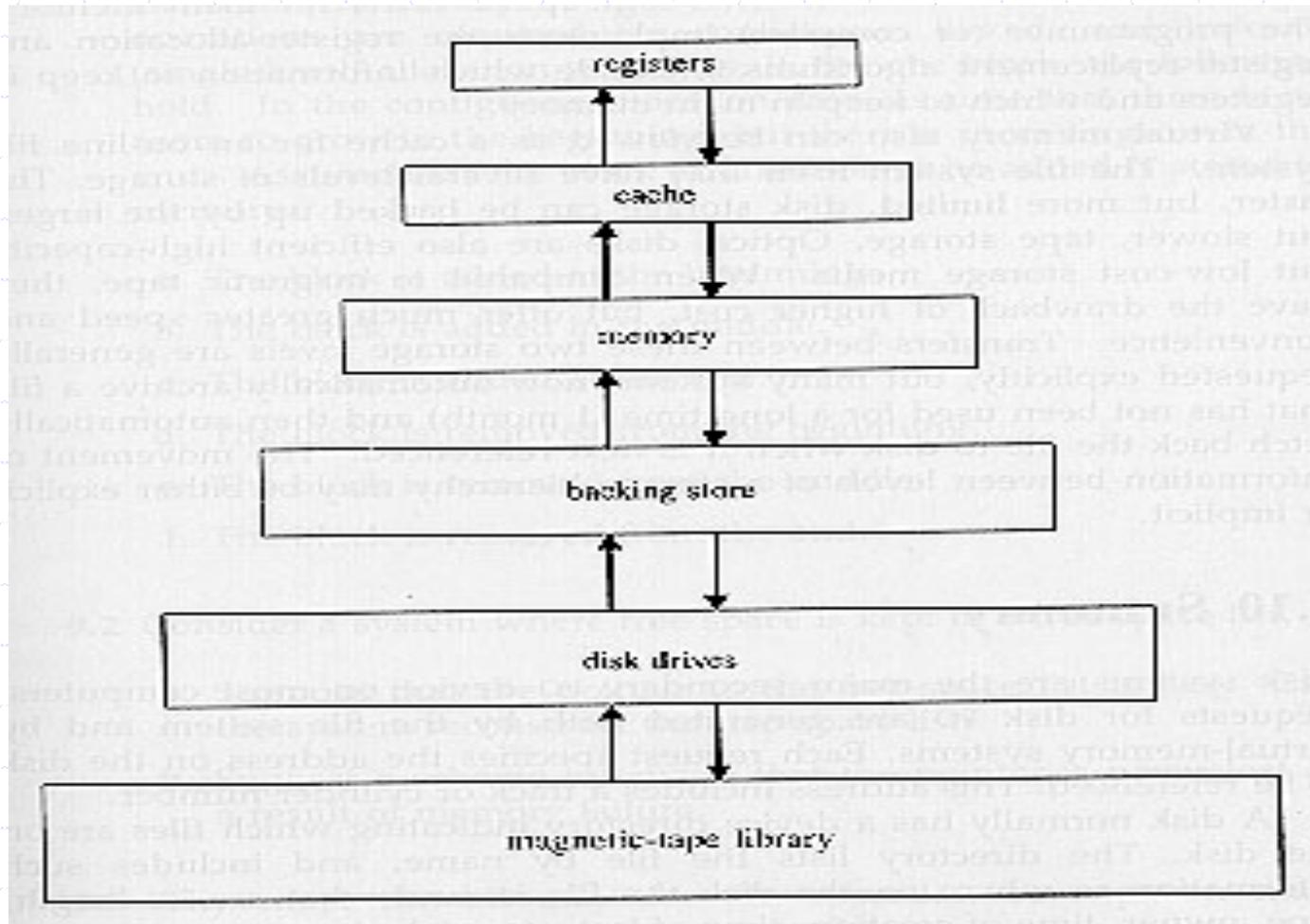
La gerarchia di memoria

- ◆ Se voglio prepararmi un piatto di spaghetti guardo se ne ho nella dispensa; se non ne trovo lì allora devo andare dal panettiere sotto casa perdendo un po' di tempo in più; se non ne trovo neanche lì devo andare al supermercato, dove perderò ancora più tempo.

Analogamente:

- ◆ Quando il processore ha bisogno di un dato, prima guarda se è presente nella memoria veloce cache; se non lo trova lì lo va cercare nella memoria centrale un po' più lenta, se non c'è neanche lì, lo va a cercare sul disco fisso che ha tempi di accesso e trasferimento superiori.

La gerarchia di memoria



La gerarchia di memoria

- ◆ I livelli di memoria generalmente presenti in un elaboratore sono:
 - i registri del microprocessore (es.: 128 registri),
 - le cache (es.: 2Mbyte di cache),
 - la memoria centrale (es.: 2GB di RAM)
 - i dischi. (es.: un disco di 160GB)

Funzionamento della memoria

- ◆ quando un elaboratore è spento tutte le informazioni risiedono su disco
- ◆ all'atto dell'accensione del calcolatore, le informazioni necessarie al funzionamento dell'elaboratore vengono trasferite da disco a memoria centrale, sino ai livelli più alti della gerarchia di memoria in funzione del loro utilizzo.

I parametri della memoria

Un dato dispositivo di memoria è caratterizzato da :

◆ velocità

- di accesso, misurata in base al tempo impiegato dal processore per accedere ad uno specificato indirizzo
- di trasferimento, misurata in base alla quantità di dati trasferita in una unità di tempo. È anche chiamata larghezza di banda, e si misura in bit/sec. oppure byte/sec.

I parametri della memoria

- ◆ volatilità, cioè la possibilità di mantenere il proprio contenuto anche in mancanza di alimentazione elettrica:
 - Le memorie volatili perdono il contenuto in assenza di alimentazione elettrica
 - Le memorie permanenti lo mantengono (non in eterno, ma per un tempo che dipende dal tipo di dispositivo: decenni o più).

I parametri della memoria

- ◆ capienza, cioè la quantità di byte che può contenere
- ◆ costo, misurato in migliaia di lire per byte
- ◆ modalità di accesso, cioè la possibilità di:
 - reperire le informazioni memorizzate (*accesso in lettura*)
 - modificare le informazioni memorizzate (*accesso in scrittura*)

Memorie: confronti

Memory technology	Typical access time	\$ per MByte in 1997
SRAM	5-25 ns	\$100-\$250
DRAM	80-120 ns	\$5-\$10
Magnetic disk	10-20 million ns	\$0.10-\$0.20

Memorie: confronti

Livello	Dimensioni tipiche	Tempo di accesso	Velocità di trasferimento
Registri	< 1 Kbyte	< 10 ns	800 MByte/s
Cache	64 — 1024 Kbyte	< 20 ns	300 MByte/s
Memoria centrale	8 — 256 MByte	60 ns	150 MByte/s
Dischi	800 — 4096 MByte	8 — 15 ms	2 — 8 Mbyte/s
Nastri	> 2048 MByte	~ 1 s	1 MByte/s

Nota: ns=nanosecondi= 10^{-9} sec.

Memorie volatili vs. non volatili

◆ Esistono due grosse categorie di dispositivi di memoria:

- le memorie volatili
- le memorie non volatili

Memorie volatili vs. non volatili

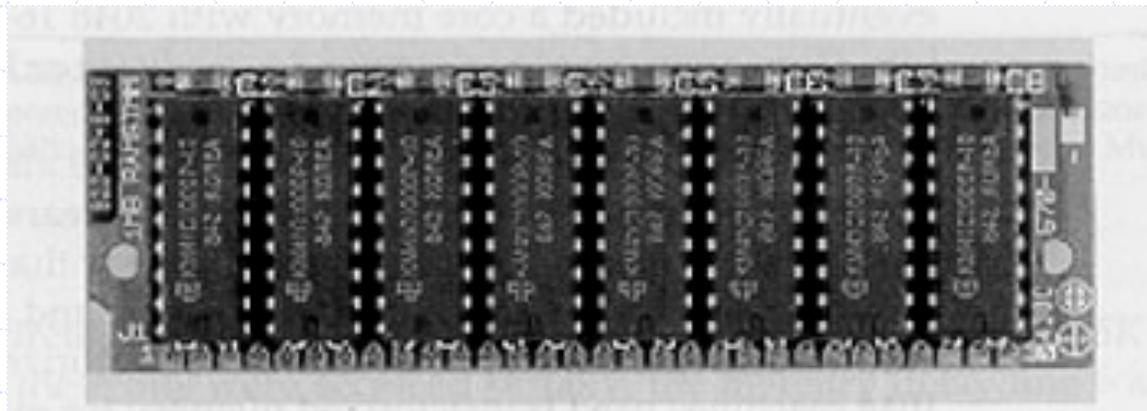
- ◆ Le memorie volatili sono tutti quei dispositivi di memoria che perdono il loro contenuto quando viene loro a mancare l'alimentazione elettrica
- ◆ Le memorie non volatili invece sono in grado di mantenere l'informazione registrata anche in assenza di alimentazione elettrica (memorie permanenti)

Le memorie volatili

- ◆ Le memorie volatili sono memorie elettroniche, realizzate con la tecnologia dei circuiti integrati (VLSI, *very large scale integration*)
- ◆ Sono indicate con il termine RAM (*Random Access Memory*) perché l'accesso ad una cella di memoria richiede un tempo indipendente dalla posizione in cui la cella risiede.
- ◆ La memoria centrale, la cache e i registri del processore sono memorie volatili

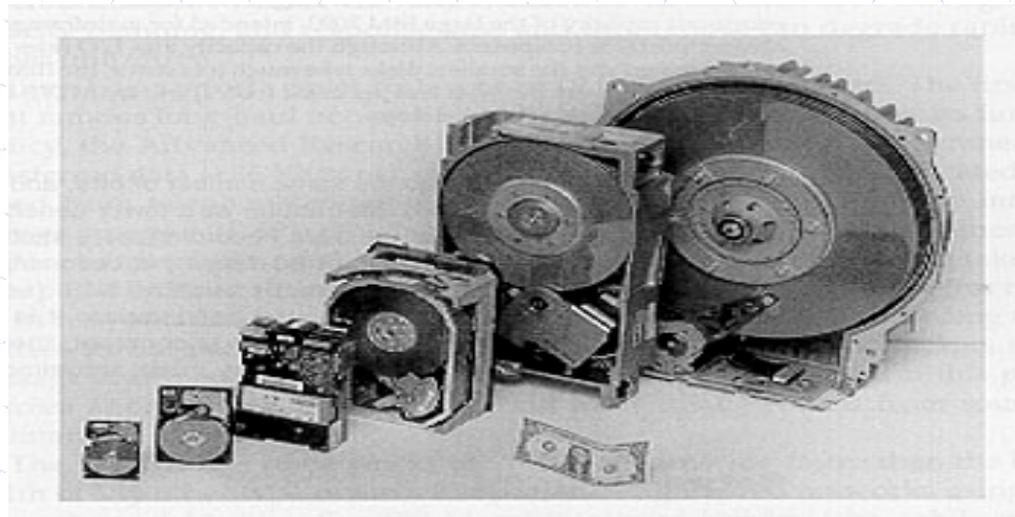
Memorie volatili

- ◆ sono genericamente indicati con il termine RAM (Random Access Memory);
- ◆ i registri di CPU, le cache e la memoria centrale.



Memorie non volatili

- ◆ I dispositivi più diffusi come memoria non volatile, sono i dischi magnetici



Le memorie permanenti

- ◆ Le memorie su supporto magnetico (dischetti, dischi, nastri)
 - Scrittura e lettura
 - Accesso sequenziale e lento
 - Basso costo

Le memorie permanenti

Le memorie permanenti si dividono in due classi:

- ◆ Le memorie di sola lettura (ROM: *read only memory*), cioè dispositivi il cui contenuto, una volta immagazzinato, è inalterabile.
- ◆ Le memorie di lettura e scrittura, i cui contenuti possono essere letti e anche cambiati nel tempo

Le memorie permanenti

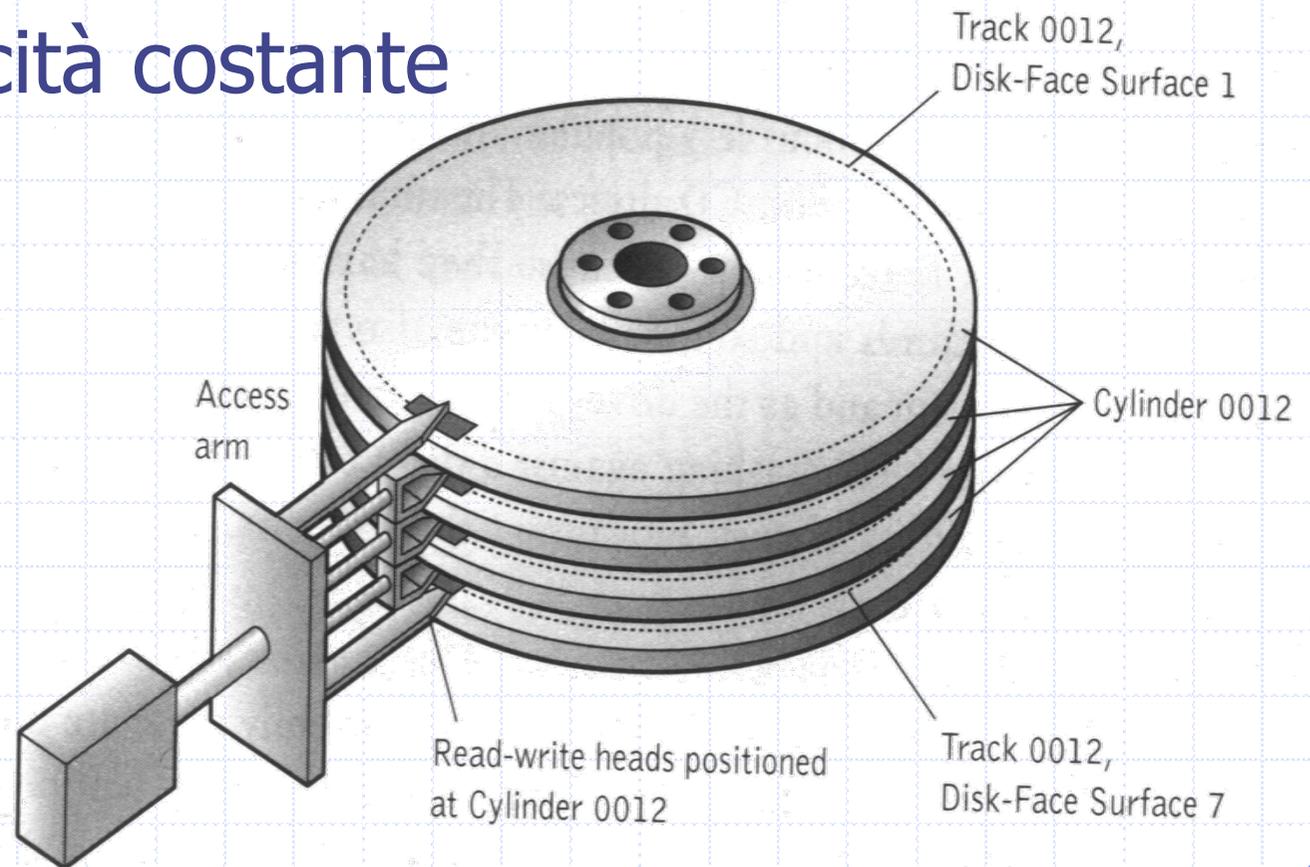
- ◆ Le memorie ottiche (CD-ROM, DVD)
- ◆ Accesso sequenziale ma veloce
- ◆ Di sola lettura (per il momento)
- ◆ Basso costo

Il disco fisso

- ◆ Il disco fisso (disco rigido, hard disk)
- ◆ Supporto di tipo magnetico con capienza (oggi) di decine di Gigabyte.
- ◆ Posizionato all'interno dello chassis.
- ◆ Consiste di un insieme di dischi magnetici (di tipo Winchester) fissati su un perno rotante.
- ◆ Le informazioni vengono memorizzate sui due lati di ciascun disco

Il disco fisso

- ◆ Suddiviso in tracce e settori.
- ◆ Ruota a velocità costante



Il disco fisso

- ◆ Un insieme di testine (una per ogni disco) leggono e scrivono le informazioni. Si possono muovere in avanti e indietro.
- ◆ Per permettere una alta densità di memorizzazione, la testina è molto vicina alla superficie del disco. Poiché il disco ruota ad alta velocità, la presenza di particelle anche molto piccole può causare la rottura della testina e/o la rigatura dei dischi. Per questo il disco è protetto da un contenitore a tenuta di polvere.

Il disco fisso

- ◆ Per eseguire una operazione di lettura o scrittura la testina:
- ◆ riceve l'indirizzo al quale leggere (numero di traccia e di settore) e la quantità i byte da trasferire.
- ◆ si muove posizionandosi sulla traccia specificata nell'indirizzo.
- ◆ aspetta che arrivi il settore specificato (il disco è sempre in rotazione)
- ◆ esegue l'operazione di lettura o scrittura

Il disco fisso

- ◆ L'organizzazione dei file sul disco fisso è complessa:
 - I file sono gestiti in blocchi di piccole dimensioni, e i singoli blocchi vengono memorizzati dove c'è spazio libero, anche non contiguo.
 - Il sistema operativo mantiene in un'area del disco fisso le informazioni relative allo spazio libero e alla "ricostruzione" di ciascun file, sparpagliato in zone diverse del disco.
 - Sistemi operativi diversi hanno criteri diversi di gestione del disco fisso.

Il disco fisso

- ◆ Se si fa un uso "intensivo" del disco, può succedere quindi che la memoria disponibile sia ancora molta, ma molto frammentata.
- ◆ In questi casi è bene utilizzare un programma di compattazione della memoria, che sposta tutti i file in aree contigue rendendo così disponibile un'area di blocchi tutti contigui

Il floppy disk

- ◆ Per permetterne l'estrazione, il dischetto è fermo e viene messo in rotazione solo quando c'è una richiesta di accesso.
- ◆ I dischetti sono dispositivi di memoria in via di estinzione, perché soppiantati da altri dispositivi più capienti, più affidabili e ugualmente poco costosi (es.: le chiavette USB).

I Compact Disk (CD-DVD)

- ◆ Dispositivi a lettura ottica di grandi capacità (centinaia di Megabyte)
- ◆ I primi CD erano di sola lettura, e venivano dunque chiamati CD-ROM.
- ◆ Le informazioni organizzate a spirale dal bordo esterno verso il centro

I Compact Disk (CD-DVD)

- ◆ La memorizzazione dei dati avviene incidendo dei solchi sulla superficie del disco (presenza del solco=1, assenza=0).
- ◆ La lettura avviene attraverso un raggio laser
- ◆ La velocità di trasferimento delle informazioni dipende dalla velocità di rotazione del disco, che a sua volta dipende soltanto dal dispositivo di lettura (i dispositivi in commercio contrassegnati dalla sigla 2x, 4x, ..., 32x per indicare la velocità del dispositivo rispetto ai primi messi in commercio).

I Compact Disk (CD-DVD)

- ◆ La tecnologia dei compact disk è in continua evoluzione, sia sul versante della capacità di memoria, sia sul versante dell'accesso in scrittura.
- ◆ I CD-R sono scrivibili una sola volta (utili per archiviare dati con la garanzia che quelle informazioni non possono essere cancellate)
- ◆ I CD-RW sono memorie di lettura e scrittura

I Compact Disk (CD-DVD)

◆ DVD (Digital Versatile Disk):

- Capacità di memorizzazione molto aumentata (ordine dei GigaByte) grazie alla densità di scrittura più elevata
- Data la differente densità di memorizzazione, il dispositivo di lettura di DVD è diverso da quello di CD.
- Il lettore di DVD è capace di leggere anche i CD

Le Read-Only Memory (ROM)

- ◆ Esiste anche una memoria non volatile realizzata con tecnologia elettronica e di sola lettura: è la ROM, che contiene informazioni di vario tipo che servono per l'esecuzione di particolari funzioni, come l'avvio del computer o la diagnosi del funzionamento delle periferiche.



Funzionamento della memoria

Massimo Marchi
<http://marchi.usr.dsi.unimi.it/>

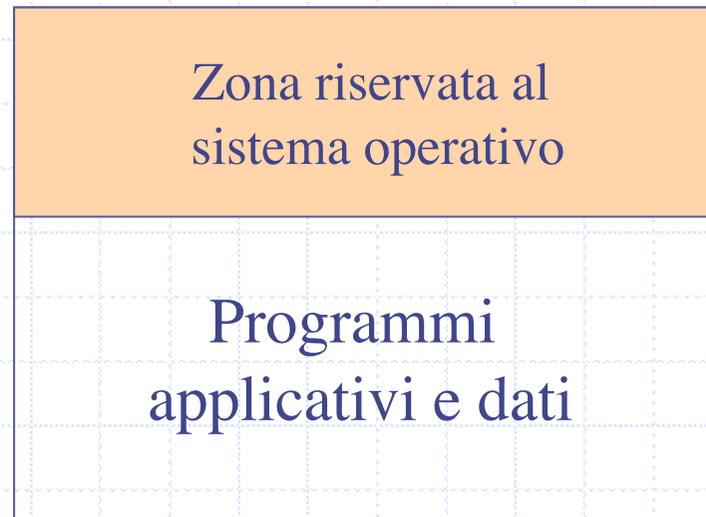
La memoria centrale

- Quando il computer è spento tutte le informazioni risiedono sul disco fisso.
- Appena lo si accende tutte le informazioni necessarie al suo funzionamento vengono trasferite dalla ROM e dal disco fisso alla memoria centrale.

La memoria centrale

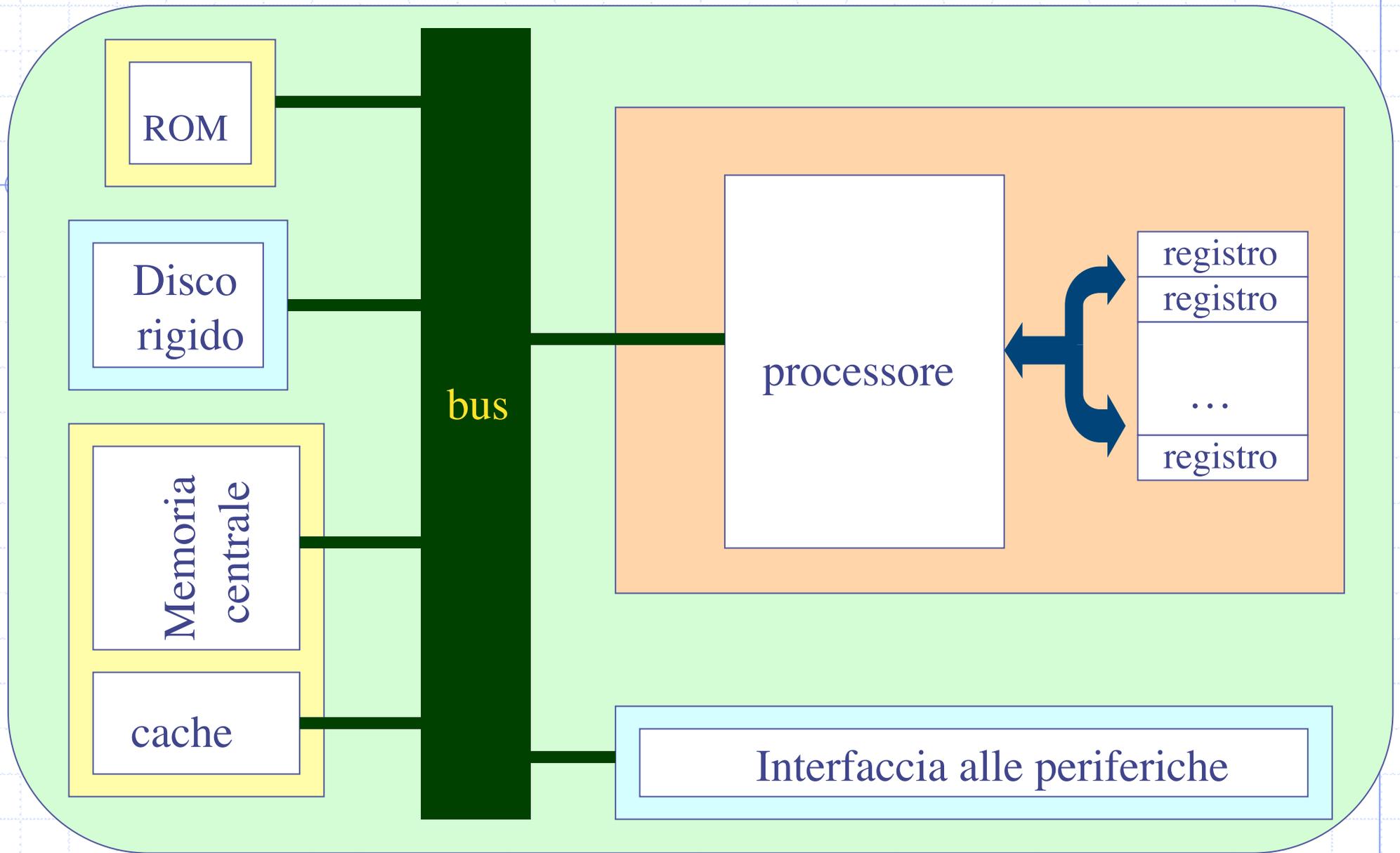
- ◆ In particolare, il sistema operativo, che a computer spento risiede sul disco fisso, viene copiato nella memoria centrale, e lì ci rimarrà fino a quando verrà spento il computer.

Memoria Centrale



La memoria cache

- ◆ La memoria cache è utilizzata come passaggio intermedio tra la memoria centrale e il processore e contiene informazioni che il sistema operativo di volta in volta ritiene più urgenti e utili.



La memoria virtuale

- ◆ Le dimensioni della memoria centrale sono relativamente piccole, e spesso succede che i programmi applicativi o i file di dati abbiano dimensioni superiori allo spazio disponibile nella RAM.
- ◆ Situazioni di questo tipo si verificano soprattutto quando un utente richiede di lavorare con più di una applicazione “contemporaneamente”.

La memoria virtuale



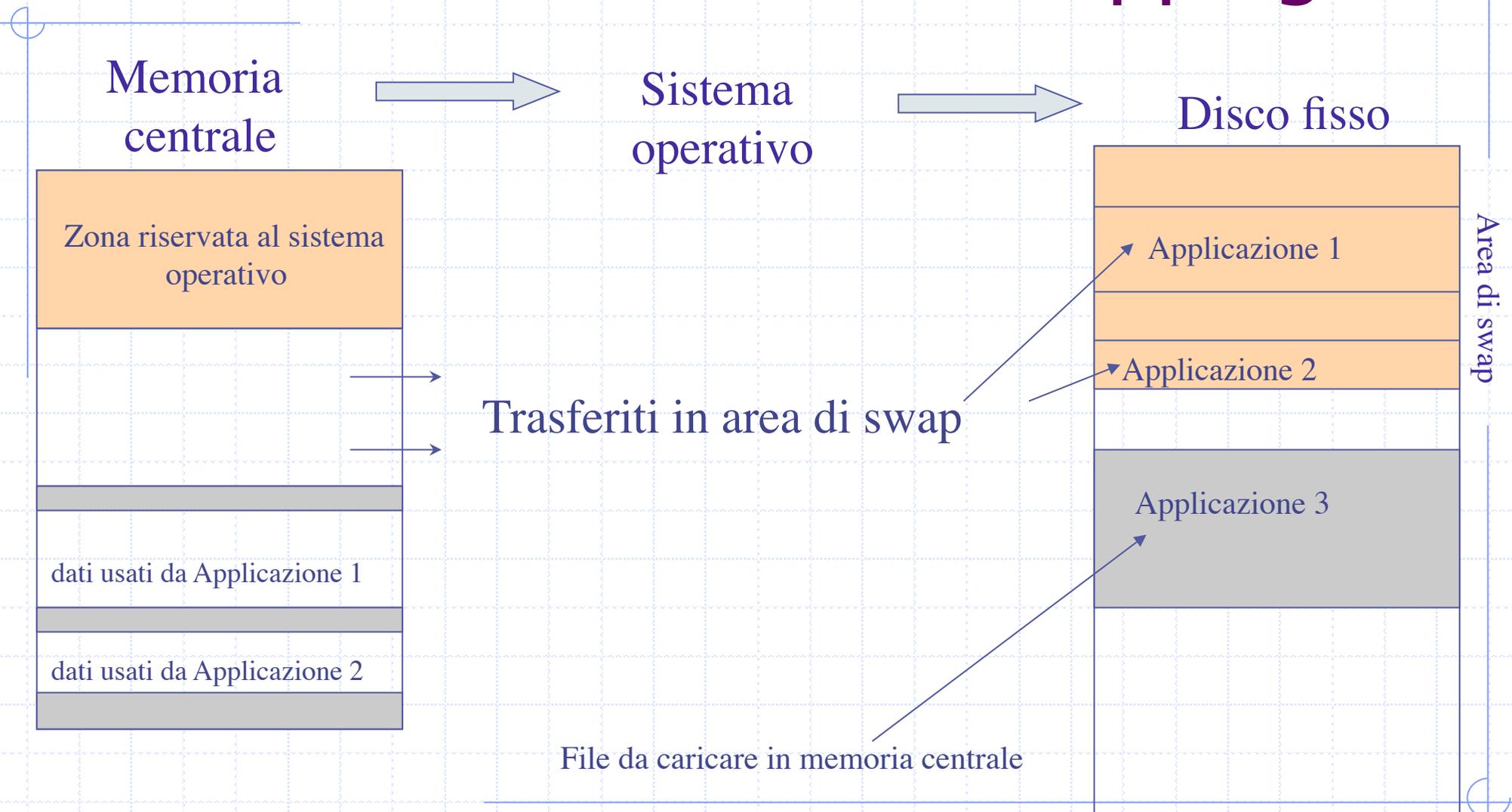
La memoria virtuale

- ◆ In questi casi il sistema operativo dovrebbe impedire l'apertura di nuovi programmi per mancanza di memoria.
- ◆ Una soluzione meno drastica consiste nel permettere al sistema operativo di utilizzare una zona del disco fisso come "espansione" della memoria centrale, e di usarla come deposito temporaneo di alcune informazioni contenute nella memoria centrale.

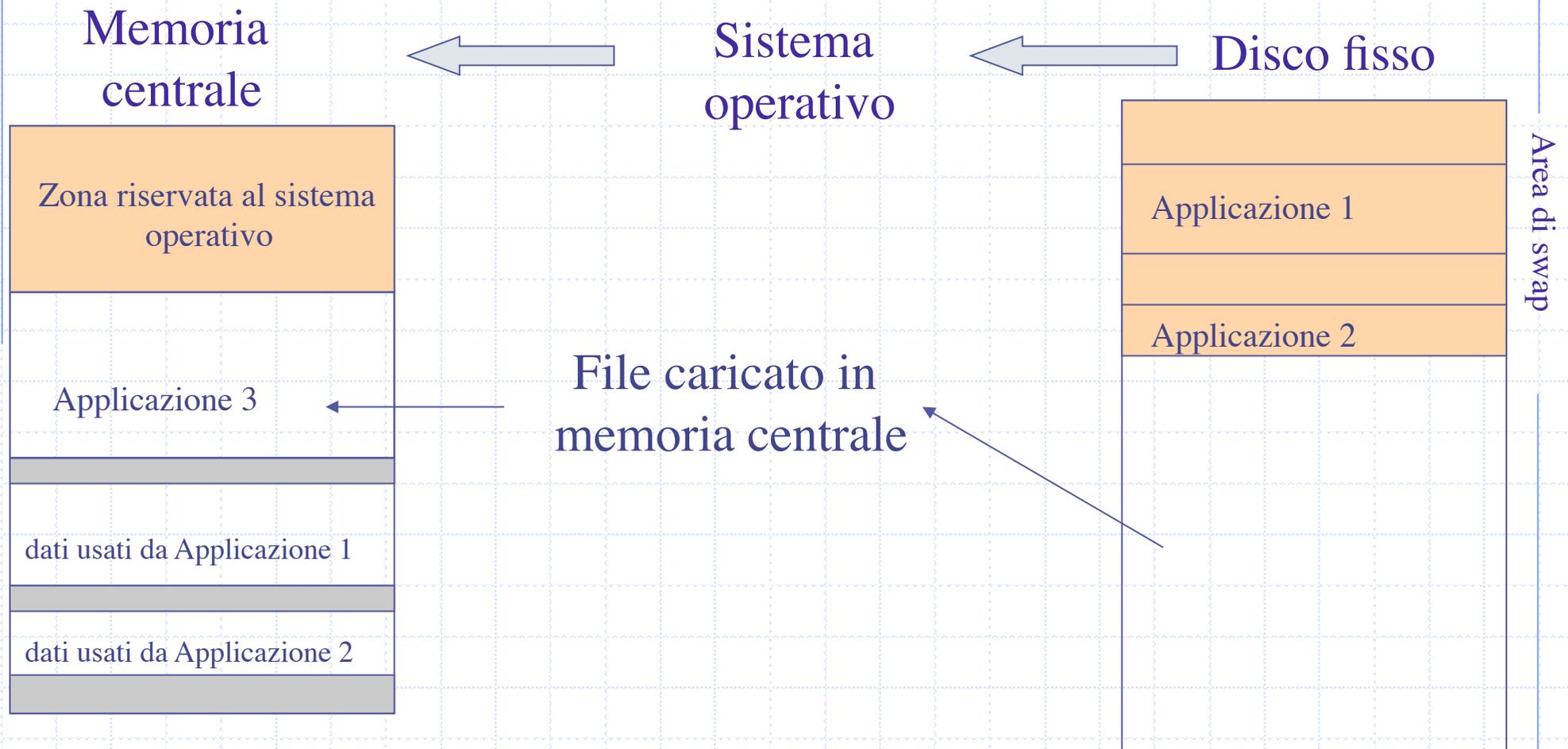
La memoria virtuale - swapping

- ◆ La parte di disco fisso dedicata a questa funzione di memoria virtuale è detta area di swap, e non è accessibile da parte dell'utente.
- ◆ Il sistema operativo sceglie quali parti della memoria centrale scaricare temporaneamente nell'area di swap e quali parti dell'area di swap riportare in memoria centrale

La memoria virtuale - swapping



La memoria virtuale - swapping



La memoria virtuale- paging

- ◆ I problemi non sono completamente risolti se si cerca spazio contiguo nella memoria centrale.
- ◆ Ecco un caso tipico non risolto applicando semplicemente lo swapping:

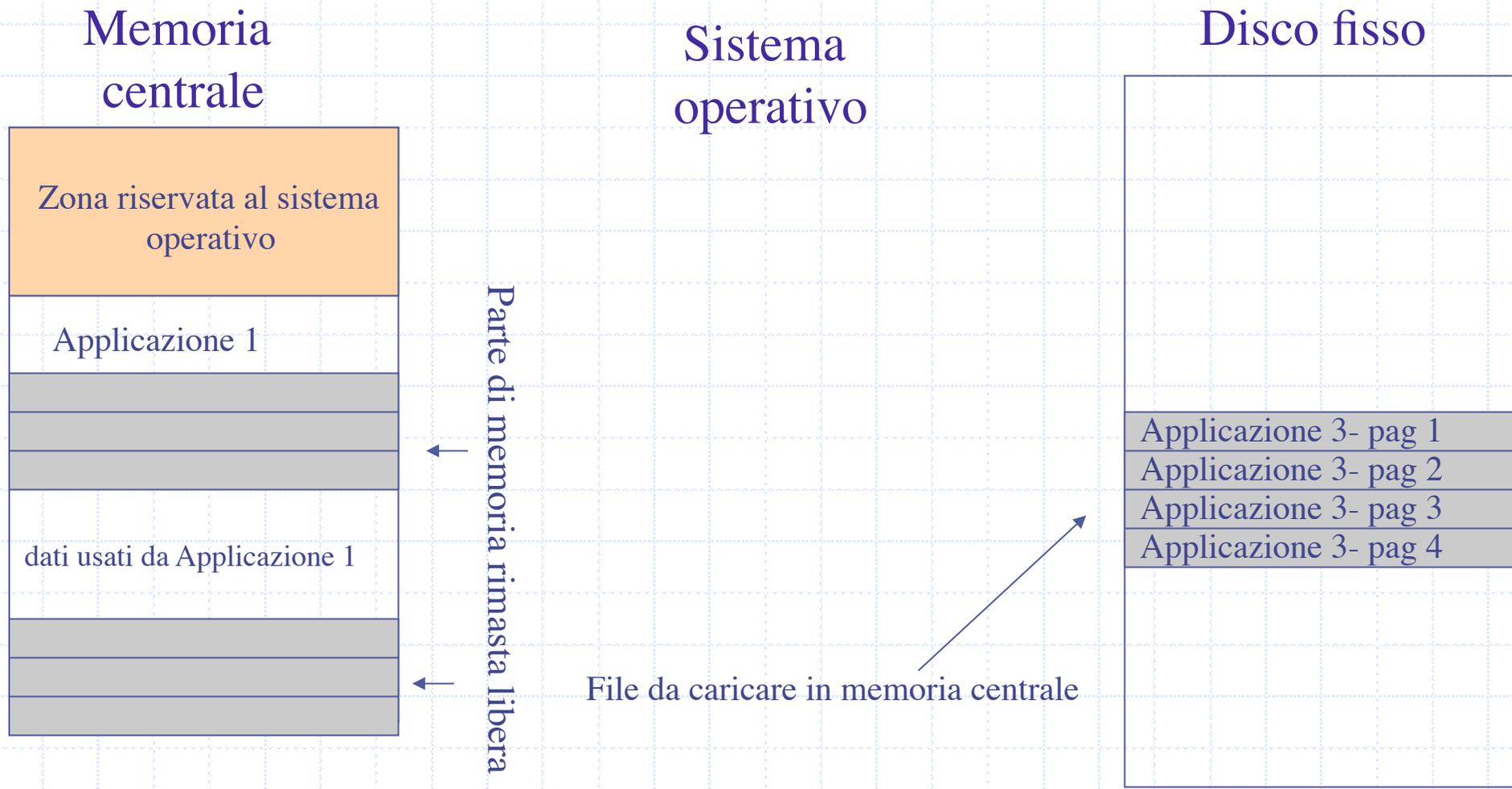
La memoria virtuale - paging



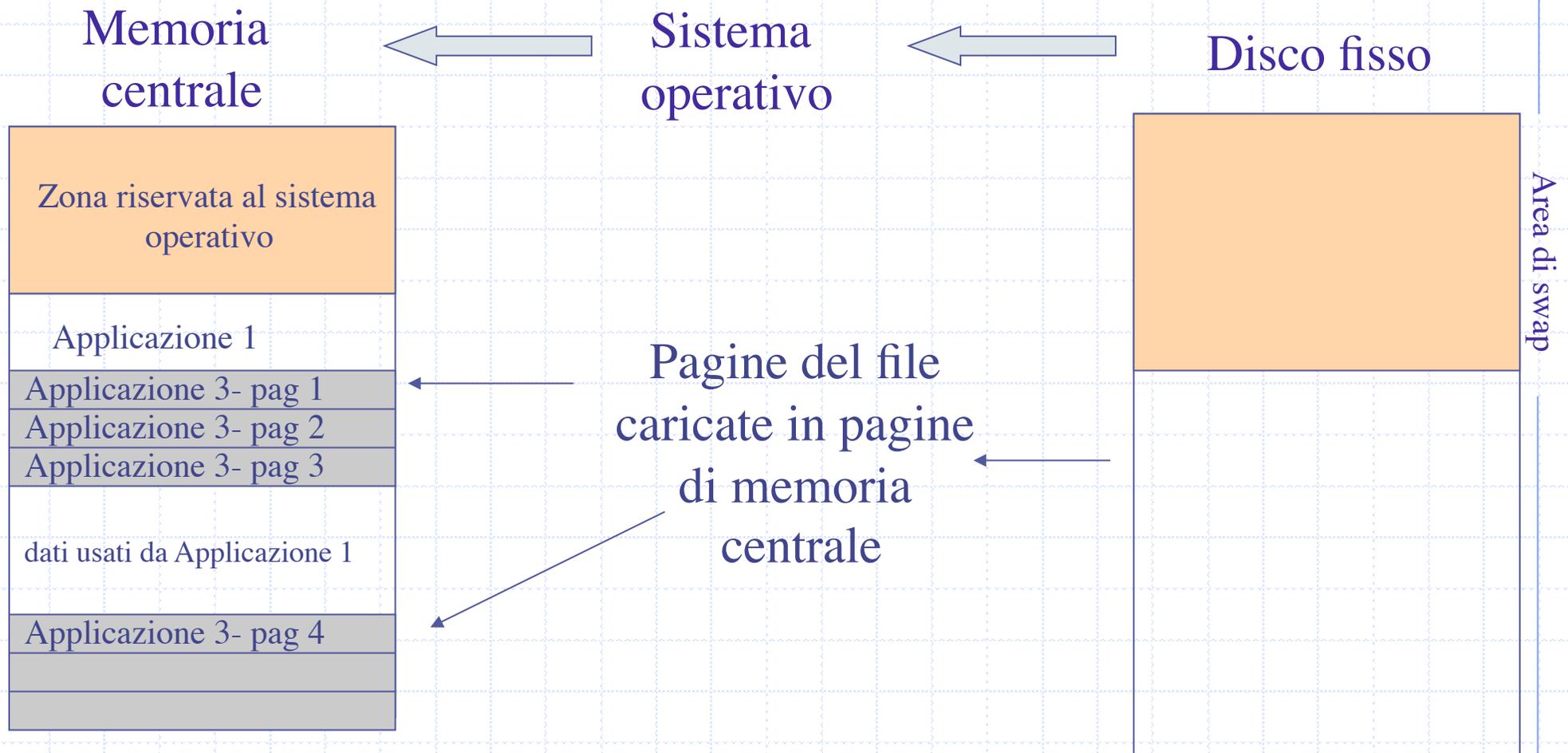
La memoria virtuale - paging

- ◆ La soluzione consiste nel gestire i file a pagine, permettendo il trasferimento di alcune pagine dalla e nella memoria centrale al posto dei file interi.

La memoria virtuale- paging



La memoria virtuale- paging



La memoria virtuale

- La combinazione di paginazione e swapping permette un utilizzo ottimizzato della memoria centrale.
- Il sistema operativo di volta in volta decide quali pagine scaricare dalla memoria centrale all'area di swap e quali portare dall'area di swap in memoria centrale.

La memoria virtuale

- L'utente può accorgersi di una *page fault* (cioè del fatto che la pagina richiesta non è presente nella memoria centrale) durante l'utilizzo di una applicazione, perché percepisce un ritardo tra un comando inviato al computer e la risposta
- ◆ Ciò è dovuto al fatto che l'algoritmo di gestione della memoria è abbastanza complesso.

La memoria virtuale

Infatti il sistema operativo deve:

- ◆ scegliere quali pagine trasferire in base all'uso che l'utente ne sta facendo
- ◆ mantenere aggiornate le tabelle che tengono traccia della corrispondenza tra indirizzi logici (cioè la sequenza di pagine di un dato file) e indirizzi fisici (cioè la loro posizione nella memoria fisica)