



# 1. Introduzione ai DB

Prof. Viglietti Francesco

[Www.in4matika.altervista.org](http://www.in4matika.altervista.org)

# Gli Archivi

Un archivio viene definito come un'insieme organizzato di informazioni caratterizzate da alcune proprietà fondamentali:

- Collegate tramite un **nesso logico** (inerenti un argomento)
- Rappresentate secondo un **formato** interpretabile
- Registrate su **supporto** di INPUT/OUTPUT
- **Organizzate** per essere facilmente consultate.

Es. Elenco telefonico → provincia e comune (nesso logico)

→ informazioni in righe (formato)

→ carta (supporto)

→ ordine alfabetico (organizzazione)

# *Operazioni sugli Archivi*

- Creazione
- Consultazione o Interrogazione
- Inserimento
- Modifica o Aggiornamento
- Cancellazione
- Ordinamento
- Stampa
- Fusione

# Organizzazione degli Archivi

Dal punto di vista logico i dati sono organizzati come in una tabella, in cui le righe prendono il nome di record e le colonne di campi. L'elenco dei campi che compongono i record viene detto **tracciato del record**

Dal punto di vista fisico gli archivi vengono memorizzati in supporti che vengono detti file. Un file è una collezione di record, che descrivono istanze di un'entità. Ogni record è composto da un insieme di campi che contengono i valori assunti dalle caratteristiche scelte per descrivere l'entità.

La creazione di un archivio richiede:

Il nome (lo identifica e ne ricorda il contenuto), Tracciato record (informazioni che compongono i record), Tipo di Supporto da usare, Dimensione massima, Organizzazione dell'archivio.

# *I supporti fisici*

Nella maggior parte dei casi gli archivi sono costituiti da insiemi di record omogenei (File di record). I file sono formati da sequenze di codice che rappresentano i caratteri a cui vengono sommati bit di parità. (bit → byte → ASCII → Unicode → UTF)

Ogni operazione di trasferimento dati dalla periferica alla memoria centrale e viceversa viene effettuata su un insieme di caratteri (blocco), le cui dimensioni variano a seconda del sistema (2K, 4K, 8K). Il blocco (Record Fisico) è l'unità fisica di memorizzazione dei dati su memoria di massa. **Record Fisico ≠ Record Logico**

Es. se in un file, il record contiene 400 caratteri (400 byte) ed il blocco è da 2KB, in esso possono essere contenuti 5 record logici e avanzano 48byte.

Fattore di blocco → numero di record logici contenuti in un record fisico

I file possono essere anche con record a lunghezza variabile, in questo caso ogni record deve terminare con un carattere che identifichi la fine (end of record)

# Memorie di massa

Schede perforate, HDD, Nastri Magnetici, Dischi Ottici, Pen disk,...

Le memorie di massa sono caratterizzate da:

Tipo di accesso (random o sequenziale), Capacità, access time(ms), Transfer rate (KB/s o MB/s), Compatibilità (possibilità di usare i supporti rendendo disponibili i dati su altre unità del sistema), Fault Tolerance (possibilità di non interruzione su un guasto), Sicurezza e Integrità (protezione dei dati su eventi naturali ed umani), Operatività (facilità d'uso), Durata (dei supporti), Trasferimento (possibilità d'uso su diversi sistemi), Espansione (aggiunta di nuove unità), Costo.

# *Backup*

Backup → consiste nel prevedere delle copie di sicurezza dei dati in modo da evitarne la perdita. Il backup può essere completo o incrementale.

Restore → è l'operazione inversa al backup, cioè il ripristino dei dati salvati

E' molto importante prevedere una strategia di backup.

# Fault Tolerance

La tolleranza del guasto dev'essere garantita nei sistemi la cui fermata è sconsigliata (sistemi bancari, siti web, ...)

Il Fault Tolerance è realizzato a diversi livelli:

Dischi → RAID (0,1,...,6)

Controller+Dischi → Duplexing

Sistema → Cluster

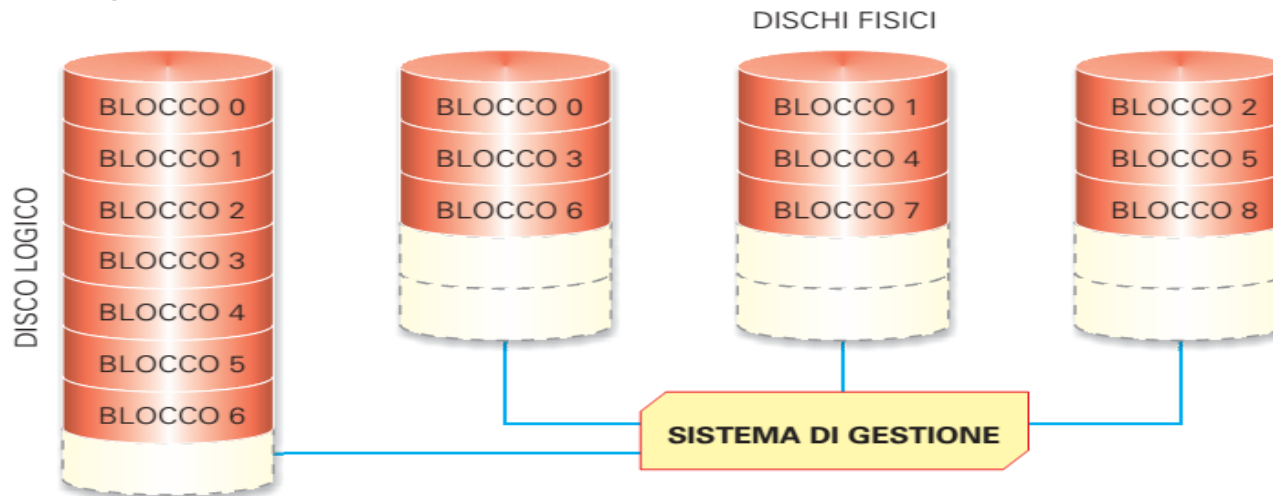
Il RAID nasce con lo scopo di velocizzare il trasferimento dati, e consiste nella creazione di un solo disco logico gestito con array di più dischi fisici (in teoria si sfrutterebbe il trasferimento in parallelo dei dati). L'evoluzione del

RAID ha consentito anche la tolleranza del guasto.



# Raid

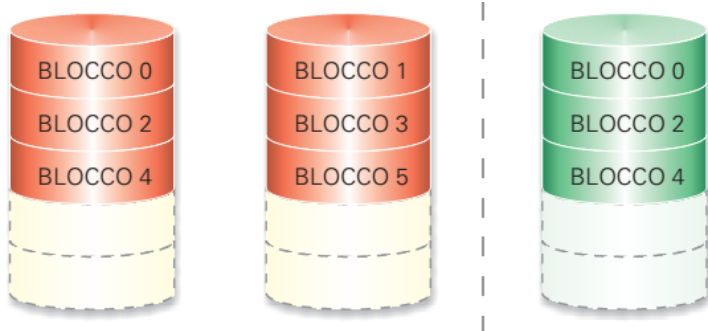
Distribuendo le informazioni su un array di più dischi fisici si può pensare di operare i trasferimenti da e verso i dischi in parallelo aumentando, in teoria, il transfer rate di tante volte quante sono i dischi. Il OS vede l'array RAID come un unico disco logico; il controller del disco RAID provvede a parallelizzare le letture e le scritture su più dischi.



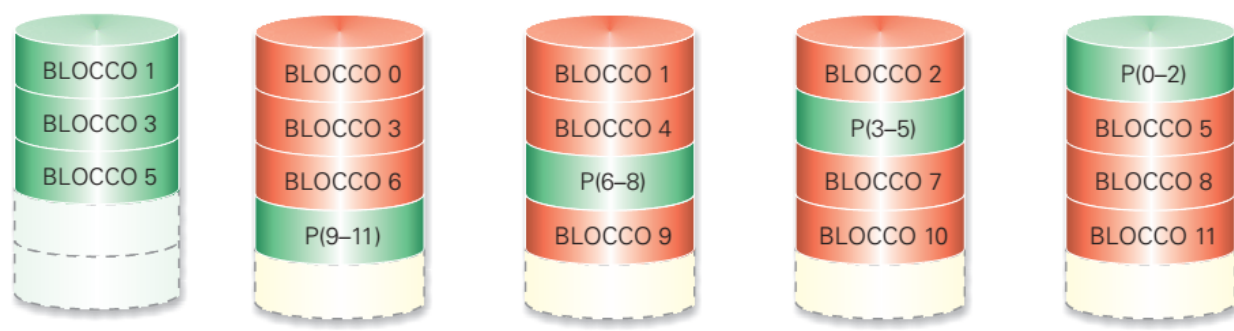
# Raid tipi

La tecnologia RAID può essere sfruttata per realizzare tecniche alternative alla duplicazione dei dischi, per garantire l'operatività del sistema anche in caso di guasti. Per es, nel caso in cui la serie RAID sia costituita da 5 dischi, 4 contengono i dati, mentre il quinto garantisce la ridondanza e viene utilizzato come disco di parità.

RAID 1



RAID 5



# Compressione Dati e Multimedialità

Necessità sempre maggiori di memorizzazione → Memorie di massa di capacità sempre maggiori.

## Occupazione di spazio su HDD

Lettera	10 KB
Foto in B/N	20 KB
Voce (1 sec)	30 KB
Video (1 sec)	3,5 MB
Libro (medio no foto)	4 MB
Quotidiano (low resol.)	15 MB
Foto color (high resol.)	12 MB

La Multimedialità aggiunge

Immagini, suoni, video

ai testi...

Esempio di spazio occupato

Da un video non compresso

Risoluzione VGA 640x480

→ VHS

$640 \times 480 = 300 \text{KB} \times 30 \text{ fps}$

= 9MB

# Collegamento delle periferiche

Tutte le periferiche sono collegate all'unità centrale attraverso le porte, e il colloquio avviene per mezzo delle interfacce e utilizzando protocolli di comunicazione.

Le porte possono essere seriali, o parallele.

STANDARD → RS232, CENTRONICS, IEEE1394, USB

Controller → dispositivo che pilota la periferica secondo le funzioni che deve eseguire.

Driver → permette al OS di controllare la periferica

BUS I/O → bus dedicato che collega la RAM al device indipendentemente dalla CPU. Il bus PCI è il + diffuso

Le interfacce più diffuse sono:

IDE (vecchio), EIDE, SCSI, SATA

# Cache e memorie

La memoria cache è una memoria temporanea che permette il trasferimento dati tra due dispositivi operanti con velocità diverse. (es. CPU ← Cache → RAM)

CPU a 3GHz → tempo di accesso 0,33 ns

Cache con t.a. 0,33 ns

RAM DDR2 a 800MHz → tempo di accesso 2,5ns

Se si deve cercare una parola, se essa è in cache il t.a. è di 0,33 ns altrimenti sarà  $0,33\text{ns} + 2,5\text{ns} = 2,83\text{ns}$  → Importanza della gestione della cache → PRINCIPI di LOCALITÀ

L. spaziale → probabile accesso a celle vicine

L. temporale → probabile accesso alla stessa cella

# Il software per la gestione dei file

Il file system è il modulo del OS che si occupa della gestione dei file. Esso consente di mettere a disposizione dell'utente-programmatore gli archivi nascondendone le proprietà fisiche di I/O.

# Principali tipi di organizzazione

Spazio e tempo sono caratteristiche fondamentali in fase di progetto dell'archivio. Ottimizzazione dello spazio in modo da rendere basso il tempo d'accesso e di ricerca.

O. Sequenziale → è la più semplice. Tutti i record sono uno di seguito all'altro. Inserimento in coda dei nuovi record, nessuna riscrittura dei record. Semplice la gestione, ma lenta la ricerca.

O. diretta → ogni record viene identificato dalla posizione occupata nel file. I nuovi record sono consecutivi. Quindi per cercare un determinato record sarà necessario conoscere l'ID. Usato negli archivi identificati attraverso un numero progressivo.

O. ad indice → occorre un'altra tabella, oltre a quella dei record, in cui vengono inserite le chiavi di ricerca

Esistono anche altri tipi di organizzazione... ad indici su più livelli, ad albero, ...