

Database - Parte 1

Utilizzo degli Archivi

Prof. Francesco Gobbi

I.I.S.S. Galileo Galilei - Ostiglia (MN)
Materia: Informatica

16 settembre 2024

Cos'è un archivio?

In generale, un **archivio** è un insieme organizzato di informazioni caratterizzate da alcune proprietà fondamentali:

- ▶ Esiste un **nesso logico** tra di esse (cioè sono in qualche modo inerenti a un medesimo argomento);
- ▶ Sono rappresentate secondo un **formato** che ne rende possibile l'interpretazione;
- ▶ Sono registrate con un **supporto** su cui è possibile scrivere e rileggere informazioni anche a distanza di tempo;
- ▶ Sono **organizzate** in modo che siano facilmente consultabili.

Perché utilizzare gli archivi?

- ▶ **Conservazione permanente dei dati:** Gli archivi permettono di memorizzare informazioni in modo persistente, garantendo che i dati possano essere recuperati anche dopo la chiusura di un'applicazione o al riavvio di un sistema.
- ▶ **Esempi di utilizzo:**
 - ▶ **Rubrica del telefono:** Memorizza numeri di telefono e nomi in modo permanente, permettendo all'utente di richiamare queste informazioni anche dopo giorni o mesi.
 - ▶ **Anagrafe:** Conserva i dati personali dei cittadini in modo da poterli consultare per documenti ufficiali, carte d'identità, e passaporti.
 - ▶ **Biblioteca:** I sistemi di gestione delle biblioteche archiviano i dati sui libri e sui prestiti, garantendo una gestione efficiente nel tempo.
- ▶ **Vantaggi:**
 - ▶ Riduzione dell'uso di carta.
 - ▶ Accesso veloce ai dati.
 - ▶ Possibilità di aggiornare e modificare i dati esistenti.

Gestione automatizzata degli archivi

Oggi gli archivi sono gestiti in modo automatico dai computer. Questo permette una maggiore efficienza e velocità nell'accesso ai dati. Gli aspetti principali della gestione automatizzata degli archivi includono:

- ▶ **Tipologia dei supporti** usati per registrare le informazioni.
- ▶ **Attrezzature hardware** dedicate alla gestione delle unità di memorizzazione.
- ▶ **Strumenti software** per la costruzione di programmi applicativi e dell'interfaccia per l'utente.
- ▶ **Organizzazione degli archivi** per un efficiente accesso ai dati e il loro veloce ritrovamento.

Record e Campi

Record: Un record è una collezione di dati relativi a un singolo elemento. Ogni record è composto da diversi **campi**, ognuno dei quali rappresenta un attributo dell'elemento.

Campi: Un campo rappresenta una singola informazione specifica all'interno del record. Ogni campo ha un valore unico e un significato specifico.

Esempio concreto: Consideriamo una tabella che rappresenta i dati di una rubrica telefonica:

Nome	Cognome	Numero di Telefono
Mario	Rossi	123-4567890
Laura	Bianchi	098-7654321

In questo esempio, ogni riga rappresenta un **record** (ad esempio, i dati di Mario Rossi), mentre ogni colonna rappresenta un **campo** (Nome, Cognome, Numero di Telefono).

Il file: collezione di record

File: Un file è una collezione di **record** che contiene informazioni logicamente omogenee. Questi record descrivono i singoli elementi di una realtà, organizzati in modo strutturato per permettere la facile consultazione e gestione.

Esempio: Un file che rappresenta una rubrica telefonica contiene tutti i record relativi agli utenti della rubrica (come Mario Rossi e Laura Bianchi nell'esempio precedente).

I file sono fondamentali per l'archiviazione di grandi quantità di dati e per la gestione efficiente di informazioni complesse, permettendo al computer di leggere, modificare e aggiornare i dati in modo rapido.

La creazione di un archivio

La **creazione di un archivio** richiede la definizione preliminare delle seguenti specifiche:

- ▶ Il **nome dell'archivio**, che lo identifica e serve a ricordarne il contenuto; per esempio, "archivio fornitori" oppure "archivio anagrafico";
- ▶ Il **tracciato record**, in altre parole quali informazioni compongono il record;
- ▶ Il **supporto** da usare per archiviare i dati (fogli di carta, dischi o nastri magnetici, dischi ottici);
- ▶ La **dimensione massima dell'archivio**: per esempio il numero massimo di scaffali occupati in un archivio cartaceo o di abbonati in un elenco telefonico;
- ▶ Il **modo** con cui i dati sono strutturati e collegati tra loro, cioè l'organizzazione dell'archivio: ci sono diverse possibilità di organizzazione e la scelta del supporto di archiviazione è spesso legata al tipo di organizzazione e alle modalità di consultazione previste.

Realizzazione di un archivio

La decisione su queste specifiche fa parte dell'analisi del problema e precede la fase di **realizzazione fisica** dell'archivio, che può consistere nella sistemazione delle informazioni già esistenti in un archivio ben organizzato, oppure nella predisposizione del supporto per la registrazione delle nuove informazioni che verranno successivamente inserite.

Dopo aver creato l'archivio, su di esso si possono effettuare operazioni di:

- ▶ **Manipolazione**, cioè inserimento di nuovi dati o variazione dei dati registrati;
- ▶ **Interrogazione**, cioè reperimento all'interno dell'archivio delle informazioni necessarie.

Archivi digitali e memorie di massa

Gli archivi digitali richiedono l'uso di **memorie di massa**, ovvero dispositivi di memorizzazione capaci di conservare grandi quantità di dati in modo permanente.

La scelta dell'archivio corretto deve tenere conto di vari parametri, tra cui:

- ▶ **Tipo di accesso:** sequenziale o casuale.
- ▶ **Capacità della memoria:** quantità di dati che possono essere memorizzati.
- ▶ **Tempo di accesso:** il tempo necessario per accedere ai dati.
- ▶ **Velocità di trasferimento:** la velocità con cui i dati possono essere letti o scritti.

La corretta scelta di questi parametri è fondamentale per garantire l'efficienza nella gestione degli archivi digitali.

Il blocco e il record fisico

Blocco: Il blocco è l'unità fisica (o **record fisico**) di memorizzazione dei dati sulla memoria di massa.

Nei sistemi digitali, la dimensione del blocco può variare e assume valori dell'ordine di:

- ▶ 2KB, 4KB, 8KB corrispondenti rispettivamente a 2048, 4096, 8192 byte.
- ▶ 1KB = 1024 byte (in quanto 2^{10}).

Durante le operazioni di lettura e scrittura, i dati vengono copiati dal disco in una zona della memoria centrale, chiamata **buffer di I/O**, per successive elaborazioni. Successivamente, i dati vengono ricopiati su disco con operazioni di scrittura.

Differenza tra blocco e record logico

Il **blocco** non deve essere confuso con il **record logico**. Può accadere che un blocco contenga più record logici.

Questo significa che le operazioni di lettura e scrittura non riguardano singoli record logici, ma gruppi di record. Ciò riduce notevolmente il numero di accessi alla memoria di massa.

Il **fattore di blocco** di un file è il numero di record logici contenuti in un blocco. Un file con fattore di blocco uguale a 1 è detto a **record sbloccati**.

Esempio di blocco e record logico

Se la lunghezza di un **record logico** è di 500 byte e la dimensione del blocco è 2048 byte, il blocco può contenere 4 record, lasciando 48 byte inutilizzati.

In questo caso, il fattore di blocco è uguale a 4. Se un file ha un fattore di blocco uguale a 1, viene detto a **record sbloccati**.

Il file system

Il **file system** è il modulo del sistema operativo che svolge le funzioni di gestione dei file. Esso permette all'utente o al programmatore di usufruire degli archivi sulle memorie di massa, senza preoccuparsi dei dettagli delle operazioni di input/output (I/O) e facendo riferimento ai file tramite nomi simbolici.

Funzioni del file system

Il **file system** svolge le seguenti funzioni principali:

- ▶ Tiene traccia dei file, della loro posizione e del loro stato, usando strutture dati chiamate **file directory**.
- ▶ Decide, in base alle richieste, quali protezioni e diritti di accesso assegnare ai file (lettura, scrittura, solo lettura).
- ▶ Assegna i file ai programmi che li richiedono e li rende disponibili.
- ▶ Rimuove l'uso dei file dai programmi quando questi non sono più necessari.

Le directory e i descrittori dei file

Il **file system** utilizza **directory** e **tabelle dei descrittori** per memorizzare le seguenti informazioni su ogni file:

- ▶ **Nome del file:** Identifica il file nel sistema.
- ▶ **Tipo di file:** Indica il tipo di contenuto del file (testo, eseguibile, sistema, ecc.).
- ▶ **Protezioni e accessi consentiti:** Lettura, scrittura, esecuzione, ecc.
- ▶ **Nome dell'utente:** L'utente che ha creato o modificato il file.
- ▶ **Dimensione del file:** La dimensione del file in byte.
- ▶ **Indirizzo di memorizzazione:** Dove il file è memorizzato sul disco (indirizzo del primo blocco).
- ▶ **Data di creazione/modifica:** Informazioni sulla creazione, ultima modifica o ultimo accesso.
- ▶ **Password:** Eventuali password necessarie per l'accesso al file.

Organizzazione sequenziale degli archivi

Organizzazione sequenziale: In questa modalità, i record sono organizzati in sequenza, uno dopo l'altro, secondo un ordine predefinito (solitamente basato su un campo chiave).

Esempio: Consideriamo un archivio di studenti ordinato per matricola. Per trovare uno specifico record, è necessario leggere tutti i record dall'inizio fino a quando non si trova quello desiderato.

Vantaggi:

- ▶ Ideale per elaborazioni che richiedono l'accesso a tutti i record (ad esempio, calcolare la media dei voti).

Svantaggi:

- ▶ L'accesso a un singolo record può essere lento, specialmente se si trova verso la fine dell'archivio.

Organizzazione ad accesso diretto

Organizzazione ad accesso diretto: In questa modalità, i record vengono memorizzati in modo che possano essere recuperati direttamente, senza dover scorrere tutti i record precedenti. L'accesso avviene utilizzando una funzione che mappa il campo chiave al suo indirizzo in memoria (hashing).

Esempio: Un archivio che memorizza i dipendenti di un'azienda, dove ogni record può essere recuperato direttamente utilizzando il numero di identificazione del dipendente.

Vantaggi:

- ▶ Accesso rapido e diretto ai record, senza dover scorrere tutta la sequenza.

Svantaggi:

- ▶ Complesso da implementare e mantenere, specialmente quando i dati cambiano frequentemente.

Organizzazione a indici

Organizzazione a indici: In questa modalità, viene creato un indice che associa il campo chiave di ogni record all'indirizzo in cui si trova il record stesso. L'indice consente un accesso veloce ai record, pur mantenendo un'organizzazione sequenziale dei dati.

Esempio: Un archivio che contiene i libri di una biblioteca, dove l'indice memorizza il titolo del libro e l'indirizzo del record corrispondente nel file principale.

Organizzazione a chiave: In alcuni sistemi, l'indice viene organizzato sulla base di una chiave primaria, che è un campo univoco per ogni record. La chiave primaria garantisce che ogni record possa essere identificato univocamente.

Vantaggi:

- ▶ Accesso rapido ai record tramite l'indice.
- ▶ Mantiene una struttura sequenziale nei file, che è utile per la gestione complessiva dei dati.

Esempio di organizzazione a chiave

Organizzazione a chiave: In questo esempio, consideriamo una tabella anagrafica dove la chiave primaria è il codice fiscale delle persone. Ogni record è associato in modo univoco al codice fiscale.

Codice Fiscale	Nome	Cognome	Data di Nascita
RSSMRA80A01H501Z	Mario	Rossi	01/01/1980
BNCLRA85B05H501T	Laura	Bianchi	05/02/1985
VRDLGI90C09H501R	Luigi	Verdi	09/03/1990

Vantaggio: Ogni persona è identificata in modo univoco dal proprio codice fiscale, rendendo l'accesso ai record molto veloce e preciso.

Costi dell'indicizzazione

Spazio: L'uso degli indici richiede più spazio di archiviazione, poiché oltre ai dati dei record stessi, è necessario anche memorizzare la struttura dell'indice. Questa struttura occupa spazio aggiuntivo su disco per permettere il collegamento tra i record e l'indice stesso.

Tempo: Quando un nuovo record viene inserito nell'archivio, oltre a scrivere i dati del record, il sistema deve anche aggiornare l'indice per includere il nuovo record. Questo comporta un tempo maggiore durante l'inserimento rispetto a un sistema senza indicizzazione.

Beneficio: Anche se l'indicizzazione introduce un costo maggiore in termini di spazio e tempo durante l'inserimento dei record, offre significativi vantaggi nelle operazioni di lettura e accesso ai dati, poiché consente di recuperare rapidamente i record senza dover scorrere l'intero archivio.