

02. File di dati

database

database

Un **archivio di dati**, o **file** (di dati), è un **insieme di dati correlati identificato da un nome, memorizzato permanentemente** su un supporto di memorizzazione di massa di un elaboratore e avente vita indipendente dal programma utilizzato per la sua creazione e/o modifica.

All'interno di un file, i dati sono normalmente raggruppati in unità logiche denominate registrazioni o **record**.

database

Nell'elenco telefonico i dati dei vari abbonati sono costituiti da un insieme di elementi (cognome, nome, numero telefonico, indirizzo, ecc.) che costituisce la struttura del record o **tracciato del record** (viene definito *aspetto intensionale* dei dati).

Ogni singolo elemento all'interno di questa struttura è noto come **campo** o *field*.

Prendendo in considerazione uno specifico abbonato, l'insieme dei dati che lo rappresentano in accordo con la struttura predefinita («Bianchi», «Giovanni», «0634162567», «Via Roma 14», ecc.) costituisce uno specifico record.

L'insieme di tutti i record di un file costituisce ciò che viene indicato come *aspetto estensionale* dei dati.

database

Ad es. la figura mostrata rappresenta un insieme di record (registrazioni) relativamente all'archiviazione dei dati anagrafici dei soci di un club:

Lo **shema**, cioè l'aspetto intensionale, è uno e coincide con il tracciato del record; le **istanze**, aspetto estensionale, invece possono essere molteplici e corrispondono a ciascun insieme di record organizzati secondo la struttura del *tracciato del record*.

	COGNOME	TIT.	NOME	DATA NASCITA	CONIUGATO	N. TELEFONO
REC. 01	ROSSI	AVV.	MARIO	12/04/1956	VERO	47367823
REC. 02	CONTE	RAG.	MARCO	10/09/1962	FALSO	45009876
REC. 03	MAN FREDI		ENZO	09/02/1960	VERO	12985558
REC. 04	PETRUZZI	DOTT.	VIOLA	23/06/1943	VERO	77823009
...						
REC. N	ALBERTI		CARLO	12/03/1971	FALSO	76904451

Il tracciato del record è:

SOCI CLUB: (Cognome, Titolo, Nome, Data nascita, Coniugato, Telefono).

file di dati - organizzazione dei file

I file di dati sono strutture informative che realizzano sistemi per la memorizzazione e l'elaborazione dei dati: queste strutture sono caratterizzate da un'organizzazione interna, dalla quale dipende sia il modo in cui esso è memorizzato sul supporto fisico della memoria di massa, sia il modo in cui viene elaborato.

L'organizzazione interna di un file è distinta in due livelli:

- **fisico**, relativo al supporto fisico di memorizzazione dei dati del file;
- **logico**, relativo alle modalità di gestione dei dati del file.

Il supporto su cui sono memorizzati i dati individua l'organizzazione fisica e condiziona quella logica: **i programmi applicativi interagiscono sempre con il livello logico.**

file di dati - organizzazione dei file

L'organizzazione logica fa riferimento alla gestione del file, cioè **al modo in cui i record sono disposti all'interno del file stesso** e, di conseguenza, al modo in cui possono essere ricercati e individuati (*modalità di accesso*).

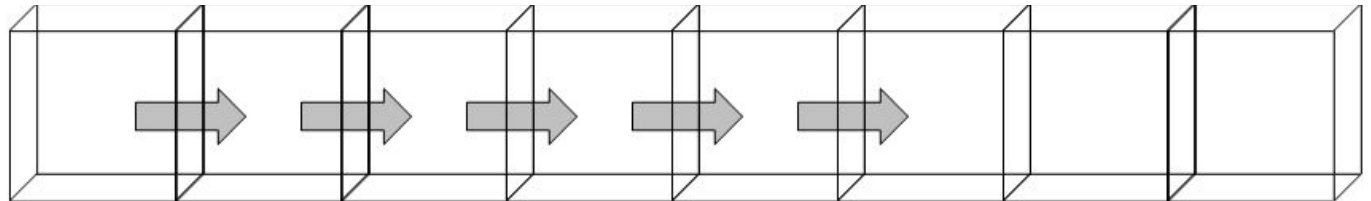
I tipi fondamentali di organizzazione logica dei file sono tre:

- *sequenziale*;
- *ad accesso diretto*;
- *indicizzata*.

file di dati - organizzazione dei file

Nei file ad **accesso sequenziale** i record sono registrati uno successivamente all'altro, nell'ordine in cui sono stati inseriti; per accedere ad uno specifico record è necessario «scorrere» tutti quelli che lo precedono.

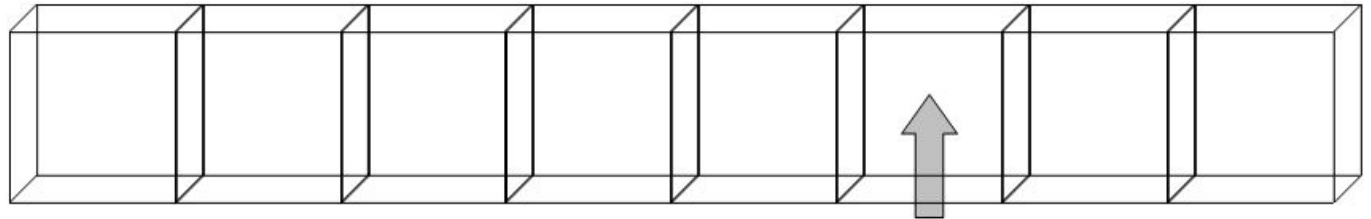
La ricerca di uno specifico record può essere facilitata se questi sono ordinati in base al campo oggetto della ricerca stessa, che può essere interrotta se, ad es. in caso di ordinamento crescente, viene elaborato un record in cui il valore del campo in oggetto è maggiore di quello ricercato (*ricerca sequenziale ottimizzata*), altrimenti è necessaria una *ricerca sequenziale completa*.



file di dati - organizzazione dei file

Accesso diretto (o *casuale, random*): **ogni singolo record è individuato da un numero che ne rappresenta la posizione all'interno del file**; quando i record sono inseriti sequenzialmente, tale numero viene assegnato incrementando quello dell'ultimo record memorizzato.

Per accedere *direttamente* a uno specifico record in modo da effettuare un'operazione di lettura o scrittura dei dati è necessario specificare il numero che ne individua la posizione.



file di dati - organizzazione dei file

In figura viene schematizzato un archivio con accesso random dove i singoli record occupano le posizioni che vanno da 1 a n .

Numero record	Nominativo	Telefono	Indirizzo	...
1	Rossi Mario	0586578182	Via del mare 12	
2	Bianchi Giovanni	0634162567	Via Roma, 14	
3	Verdi Carla	0265134257	Piazza Garibaldi, 21	
4	Rossini Giuseppe	0554567123	Viale Marconi, 1	
5	Neri Marta	0567821453	Corso Mazzini, 56	
6	Bianchi Andrea	0587223311	Viale Italia, 144	
7	Neri Daniele	0586441234	Via Leopardi, 5	
...		
n	Rossi Maria	0552432185	Scali Manzoni, 68	

file di dati - organizzazione dei file

Indicizzata (o *indexed*): nella struttura del record del file viene individuato un campo (o un insieme di campi) denominato **chiave**, i cui valori identificano univocamente i singoli record.

In questo tipo di organizzazione, oltre ad un file primario ad accesso diretto in cui sono memorizzati i record in ordine di inserimento, viene gestito anche un ulteriore file indice contenente una tabella delle chiavi.

Questa tabella viene utilizzata dal codice che effettua l'accesso al file che la gestisce mantenendola costantemente ordinata secondo un criterio *lessicografico*.

file di dati - organizzazione dei file

La ricerca di uno specifico record avviene, a partire da un valore fornito in input, consultando la tabella delle chiavi sulla quale, per la sua organizzazione, la ricerca risulta essere molto rapida, e utilizzando la posizione del record per accedere al file primario dei dati.

In questo modo è possibile accedere «direttamente» ad un record specificandone la chiave e, a partire da questa, elaborare sequenzialmente i record precedenti e successivi seguendo l'ordinamento dei valori chiave.

file di dati - organizzazione dei file

La figura mostra un file indice formato dalla coppia [Chiave, Puntatore], con i valori chiave in ordine crescente ed il corrispondente puntatore che contiene la posizione del primo byte del record corrispondente all'interno del file dati.

INTERVALLO DI RICERCA INIZIALE																																									
Nr.	CHIAVE	PUNTAT.																																							
1	1	8473																																							
2	4	0983																																							
3	7	2749																																							
4	8	0019																																							
5	12	3409																																							
6	14	7194																																							
7	15	4589																																							
8	16	2817																																							
9	19	0009																																							
10	21	9095																																							
11	23	6693																																							
12	25	7612																																							
13	26	4009																																							
14	28	2930																																							
15	29	4059																																							
16	30	6785																																							
17	33	9377																																							
18	35	3000																																							
19	37	2938																																							
20	38	4958																																							
21	41	6768																																							
22	44	5859																																							
23	45	1237																																							
24	46	0940																																							
25	49	3829																																							
26	50	0958																																							
27	52	0048																																							
28	56	5678																																							
29	58	4503																																							
30	59	1024																																							
31	60	0104																																							
32	62	7849																																							
33	65	0048																																							
34	66	3309																																							
35	67	2975																																							
36	68	0095																																							
37	71	7809																																							
38	72	1829																																							
39	74	4859																																							
40	75	0594																																							
41	78	0738																																							
42	82	9873																																							

Inizialmente l'intervallo di ricerca coincide con tutto il file indice (da **1** a **42**), mentre l'elemento di confronto è l'elemento centrale calcolato come: $(1+42)/2 = 21$.

Vogliamo cercare il record con valore **chiave = 58**.

file di dati - organizzazione dei file

Dal momento che il valore della chiave corrispondente alla posizione nr.21 = 41 < 58, questa, se presente, non può trovarsi nella prima metà del file indice; ridefiniamo di conseguenza l'intervallo di ricerca.

Nr.	CHIAVE	PUNTAT.
1	1	8473
2	4	0983
3	7	2749
4	8	0019
5	12	3409
6	14	7194
7	15	4589
8	16	2817
9	19	0009
10	21	9095
11	23	6693
12	25	7612
13	26	4009
14	28	2930
15	29	4059
16	30	6785
17	33	9377
18	35	3000
19	37	2938
20	38	4958
21	41	6768
22	44	5859
23	45	1237
24	46	0940
25	49	3829
26	50	0958
27	52	0048
28	56	5678
29	58	4503
30	59	1024
31	60	0104
32	62	7849
33	65	0048
34	66	3309
35	67	2975
36	68	0095
37	71	7809
38	72	1829
39	74	4859
40	75	0594
41	78	0738
42	82	9873

Adesso l'intervallo di ricerca va da **22** a **42**, mentre l'elemento di confronto può essere calcolato come: $(22+42)/2 = 32$.

file di dati - organizzazione dei file

Il valore della chiave corrispondente alla posizione nr.32 = 62 > 58, dunque la chiave cercata si trova nella prima metà dell'intervallo di ricerca.

		INTERVALLO DI RICERCA	
Nr.	CHIAVE	PUNTAT.	
1	1	8473	
2	4	0983	
3	7	2749	
4	8	0019	
5	12	3409	
6	14	7194	
7	15	4589	
8	16	2817	
9	19	0009	
10	21	9095	
11	23	6693	
12	25	7612	
13	26	4009	
14	28	2930	
15	29	4059	
16	30	6785	
17	33	9377	
18	35	3000	
19	37	2938	
20	38	4958	
21	41	6768	
22	44	5859	
23	45	1237	
24	46	0940	
25	49	3829	
26	50	0958	
27	52	0048	
28	56	5678	
29	58	4503	
30	59	1024	
31	60	0104	
32	62	7849	
33	65	0048	
34	66	3309	
35	67	2975	
36	68	0095	
37	71	7809	
38	72	1829	
39	74	4859	
40	75	0594	
41	78	0738	
42	82	9873	

Il nuovo intervallo di ricerca va da **22** a **31**, e l'elemento di confronto viene calcolato come: $(22+31)/2 = \mathbf{26}$.

file di dati - organizzazione dei file

Il valore della chiave corrispondente alla posizione nr.26 = $50 < 58$; la chiave cercata si trova nella seconda metà dell'intervallo di ricerca.

		INTERVALLO DI RICERCA			
Nr.	CHIAVE	PUNTAT.			
1	1	8473	0983	2749	0019
2	4	3409	7194	4589	2817
3	7	0009	9095	6693	7612
4	8	4009	2930	4059	6785
5	12	9377	3000	2938	4958
6	14	6768	5859	1237	0940
7	15	3829	0958	0048	5678
8	16	0048	5678	4503	1024
9	19	9095	6693	7612	4009
10	21	2817	0009	9095	6693
11	23	7612	4009	2930	4059
12	25	6785	9377	3000	2938
13	26	4958	6768	5859	1237
14	28	0940	3829	0958	0048
15	29	5048	5678	4503	1024
16	30	0104	7849	0048	3309
17	33	3309	2975	0095	7809
18	35	7809	1829	4859	0594
19	37	0738	9873		
20	38				
21	39				
22	40				
23	41				
24	42				

Il nuovo intervallo di ricerca va da **27** a **31**, e l'elemento di confronto viene calcolato come: $(27+31)/2 = \mathbf{29}$, al cui interno è presente la chiave cercata.

file di dati - organizzazione dei file

L'organizzazione *indexed* offre la possibilità di definire più indici in funzione di chiavi diverse (relativamente ad un elenco telefonico potrebbe essere necessario definire un indice per il numero di telefono e uno, distinto, per il cognome e il nome dell'abbonato) mantenendo unico il file primario dei dati.

In questo modo è possibile vedere il file ordinato in modi diversi a seconda delle necessità di elaborazione: questa caratteristica determina quella che generalmente viene chiamata *multidimensionalità di un archivio*.

file di dati - organizzazione dei file

La figura mostra un archivio organizzato in maniera *indexed* su due distinti indici definiti rispettivamente sul nominativo e sul numero di telefono.

L'archivio centrale (primario) è gestito in maniera random, mentre i due indici hanno le rispettive chiavi mantenute costantemente ordinate e accanto a ognuna di esse viene gestito il riferimento al relativo record nell'archivio primario.

Accedendo all'archivio attraverso un indice o l'altro è possibile «vedere» i record ordinati secondo i due diversi criteri.

	Nominativo	Telefono	Indirizzo
1	Rossi Mario	0586578182	Via del mare 12
2	Bianchi Giovanni	0634162567	Via Roma, 14
3	Verdi Carla	0265134257	Piazza Garibaldi, 21
4	Rossini Giuseppe	0554567123	Viale Marconi, 1
5	Neri Marta	0567821453	Corso Mazzini, 56
6	Bianchi Andrea	0587223311	Viale Italia, 144
7	Neri Daniele	0586441234	Via Leopardi, 5
...	
n	Rossi Maria	0552432185	Scali Manzoni, 68

Indice nominativo

Bianchi Andrea	6
Bianchi Giovanni	2
Neri Daniele	7
Neri Marta	5
Rossi Maria	n
Rossi Mario	1
Rossini Giuseppe	4
...	
Verdi Carla	3

Indice telefono

0265134257	3
0552432185	n
0554567123	4
0567821453	5
0586441234	7
0586578182	1
0587223311	6
...	
0634162567	2