

Organizzazione dei dati sul disco (DBMS)

I file sono insieme di pagine, che tipicamente contengono:

header #entry, identificatore pagina (PID)...

directory puntatori ai record della pagina

record disposti negli *slot* della pagina

I record sono identificati tramite RID (*row identifier*), coppie di PID e slot. Lo slot è la posizione nel vettore directory, quindi il record può essere spostato liberalmente all'interno della pagina aggiornando solo un puntatore.

Le pagine di un file possono essere fisicamente disposte in vari modi:

seriale (heap file) dati disordinati inseriti uno dopo l'altro — semplice e senza overhead di spazio ma poco efficiente (ricerca lineare), solitamente strategia di default;

sequenziale ordinamento dei dati per consentire ricerca binaria. Efficiente per ordinamento, raggruppamento e ricerca (accessi logaritmici nel numero di pagine del file), ma complica la gestione di inserimento, cancellazione e aggiornamento. Non è utile se le operazioni più frequenti richiedono di analizzare tutta la tabella;

procedurale (hash file) la pagina in cui si colloca un record è quella con indirizzo uguale a un hash della chiave del record, overflow gestito con liste di trabocco (e solo se si riempie una pagina), efficienti su ricerche per uguaglianza della chiave (in particolare non per intervalli) e la dimensione del file non varia molto nel tempo (costo di rehashing) — se si può prevedere il numero di ennuple, si crea un file di dimensione tale da avere un determinato load factor;

tabellare indice (associazioni chiave → riferimento record) memorizzato con un B+ tree. Efficiente per ricerche di uguaglianza e intervallo sulla chiave.