

# Implementazione degli operatori relazionali

**proiezione** per `DISTINCT` si ordinano i risultati per eliminare i duplicati;

**restrizione** `SELECT ... WHERE` con condizione semplice:

- ricerca lineare senza indice e con dati disordinati;
- altrimenti accesso all'indice (e.g. costo logaritmico nel numero di pagine) e ai dati (dipende dal numero di match, costante per ricerche su superchiavi);

**aggregazione** con `GROUP BY` si ordina sugli attributi indicati prima di calcolare le funzioni di aggregazione;

## Giunzione

Si distinguono operatore sinistro/esterno e destro/interno. Approcci per  $R \bowtie_{\text{pred}} S$ :

**cicli annidati**

```
for r in R:
  for s in S:
    if pred(r, s):
      add(r, s)
```

- avendo a disposizione due buffer (uno per  $R$  e uno per  $S$ ), sono richieste  $\text{npag}(R) + \text{nrec}(R)\text{npag}(S)$  operazioni di I/O. Si può confrontare con il valore ottenuto con  $S \bowtie R$  per decidere il migliore; visto che

$$\text{npag}(R) + \text{nrec}(R)\text{npag}(S) \approx \text{nrec}(R)\text{npag}(S),$$

è sufficiente determinare

$$\begin{aligned} \text{nrec}(R)\text{npag}(S) &\geq \text{nrec}(S)\text{npag}(R) \\ \frac{\text{nrec}(R)}{\text{npag}(R)} &\geq \frac{\text{nrec}(S)}{\text{npag}(S)}, \end{aligned}$$

ovvero usiamo come relazione esterna quella con i record più grandi ( $\text{nrec}/\text{npag}$  minore);

- con  $\text{npag}(S)$  buffer,  $\text{npag}(R) + \text{npag}(S)$  letture;
- si preserva l'ordine dei record della relazione esterna, per cui la scelta può essere influenzata dall'ordinamento richiesto per il risultato;

**nested loop a pagine**

```
for p in pages(R):
  for q in pages(S):
    for r in p:
      for s in q:
        if pred(r, s):
          add(r, s)
```

Più efficiente,  $\text{npag}(R)(1 + \text{npag}(S))$  I/O, ma non mantiene l'ordinamento;

**nested loop con indice** se c'è un indice sugli attributi di  $S$  usati per il join, si può evitare di scorrerla per intero;

**sort-merge** applicabile quando entrambe le tabelle sono ordinate sugli attributi di join. Costo  $\text{npag}(R) + \text{npag}(S)$ .