

Esercitazione 1

1

- Considerare la seguente base di dati:
- **Aeroporto** (Città, Nazione, Continente)
- **Volo** (CodVolo, TipoAereo, GiornoSettimana, CittàPartenza, OraPartenza, CittàArrivo, OraArrivo, CodCompagnia, NumScali)
- **Aereo** (TipoAereo, NumPasseggeri, QuantMerci)
- **Scali** (CodVolo, GiornoSettimanaArrivo, Città, Ora Arrivo, Ora Partenza)

2

- Scrivere una espressione in algebra relazionale che elenchi i codici dei voli da Roma a Bangkok che fanno scalo a Doha solo il giovedì.
- $\Pi_{CV}(\sigma_{CP='Roma' \wedge CA='Bangkok'}(Volo)) \supset \Delta$
- $(\Pi_{CV}(\sigma_{GS=giovedi \wedge Città='Doha'}(\Pi_{Città, GS, CV}(Scali)))) - \Pi_{CV}(\sigma_{GS \neq giovedì \wedge Città='Doha'}(\Pi_{Città, GS, CV}(Scali))))$

3

- Scrivere l'interrogazione precedente nel calcolo dei domini
- $\{ CV:cv \mid Volo(CV:cv, TA:ta, CP:cp, CA:ca..) \wedge cp='Roma' \wedge ca='Bangkok' \wedge \neg \exists oa, op. Scali(CV:cv, Città:c, GiornoSettimana:gs, Ora Arrivo:oa, Ora Partenza:op) \wedge c='Doha' \wedge gs \neq giovedì \}$

4

Esercizio 1

Considerare la seguente base di dati relazionale:

CLIENTI (Codice, Nome, Indirizzo, Città)

NOLEGGI (Ciente, Auto, DataPrelievo, DataRestituzione)

AUTOVETTURE (Targa, Modello, Colore, AnnoImmatricolazione, CostoGiornaliero)

con vincolo di integrità referenziale fra l'attributo Auto e la chiave della relazione AUTOVETTURE e fra l'attributo Ciente e la chiave della relazione CLIENTI formulare in algebra relazionale:

1. l'interrogazione che restituisce i clienti che hanno noleggiato più di un'autovettura
2. l'interrogazione che restituisce i clienti che hanno noleggiato autovetture di un solo modello

Soluzione:

$$1. \pi_{\text{Ciente}} (\sigma_{\text{Auto} \neq \text{Auto}'} (\pi_{\text{Ciente}, \text{Auto}} (\text{NOLEGGI}) \bowtie_{\text{Ciente} = \text{Ciente}'} \rho_{X' \leftarrow X} (\pi_{\text{Ciente}, \text{Auto}} (\text{NOLEGGI}))))$$

$$2. \quad V = \text{NOLEGGI} \bowtie_{\text{Auto} = \text{Targa}} \text{AUTOVETTURE}$$

$$V1 = V \bowtie_{\text{Ciente} = \text{Ciente}'} (\rho_{X' \leftarrow X} V)$$

$$\pi_{\text{Ciente}} \text{NOLEGGI} - \pi_{\text{Ciente}} (\sigma_{\text{Modello} \neq \text{Modello}'} (V1))$$

Esercizio 2

Considerare una relazione $R(A, \underline{B}, C, D, E)$. Indicare quali delle seguenti proiezioni hanno certamente lo stesso numero di tuple di R :

1. $\pi_{ABCD}(R)$
2. $\pi_{AC}(R)$
3. $\pi_{BC}(R)$
4. $\pi_C(R)$
5. $\pi_{CD}(R)$.

Soluzione

1. $\pi_{ABCD}(R)$ **SI**
2. $\pi_{AC}(R)$ **NO**
3. $\pi_{BC}(R)$ **SI**
4. $\pi_C(R)$ **NO**
5. $\pi_{CD}(R)$ **NO**.

3

- Si consideri il seguente schema di base di dati
 - Aeroporto (Città, Nazione)
 - Volo (IdVolo, Giorno, OraArrivo, CittàArrivo, CittàPartenza, OraPartenza)
 - Aereo (Tipo, NumPasseggeri, QuantMerci)
- Scrivere una espressione dell'algebra relazionale che elenchi gli identificatori dei voli internazionali in partenza da Pisa con durata inferiore alle 2 ore.

9

Soluzione

$\pi_{\text{IdVolo}} (\sigma_{\text{Nazione} \neq \text{Italia}}(\text{Aeroporto}))$
 $\bowtie_{\text{CittàArrivo}=\text{Città}} \Pi_{\text{IdVolo}, \text{CittàArrivo}}$
 $(\sigma_{\text{CittàPartenza} = \text{'Pisa'} \wedge (\text{OraArrivo} - \text{OraPartenza}) < 2}(\text{Volo}))$

- Rappresentare lo stesso risultato nel calcolo dei domini.

10

- $\{ \text{IdVolo:iv} \mid \text{Volo}(\text{IdVolo: iv}, \text{Giorno: g}, \text{OraArrivo: oa}, \text{CittàArrivo: ca}, \text{CittàPartenza: cp}, \text{OraPartenza: op}) \wedge \text{Aeroporto}(\text{Città: c}, \text{Nazione:n}) \wedge c=ca \wedge cp= \text{"Pisa"} \wedge n \neq \text{"Italia"} \wedge (oa-op) < 2 \}$
- Rappresentare lo stesso risultato nel calcolo delle tuple.

11

- Soluzione:
- $\{ i.(\text{IdVolo}) \mid i(\text{Volo}), a(\text{Aeroporto}) \mid i.\text{CittàArrivo}=a.\text{Città} \wedge i.\text{CittàPartenza} = \text{"Pisa"} \wedge a.\text{Nazione} \neq \text{"Italia"} \wedge (i.\text{OraArrivo}-i.\text{OraPartenza}) < 2 \}$

12

4

- Si consideri l'espressione algebrica

$$\pi_{ADH}(\sigma_{(B=C) \wedge (E=F) \wedge (A>20) \wedge (G=10)}((R1 \bowtie R3) \bowtie R2))$$

- dove R1, R2 , R3 hanno gli schemi R1(AB), R2(CDE), R3(FGH).
Trasformarla in modo da ridurre la dimensione dei risultati intermedi.

13

Soluzione

$$\pi_{ADH}(\sigma_{A>20}(R1) \bowtie_{B=C} \pi_{CDH}(R2 \bowtie_{E=F} \pi_{FH}(\sigma_{G=10}(R3))))$$

14

5

- Si consideri uno schema relazionale contenente le relazioni $R1(ABC)$, $R2(DG)$, $R3(EF)$
- Formulare in calcolo relazionale su tuple e su domini l'interrogazione realizzata in algebra relazionale dalla seguente espressione:

$$(R_3 \bowtie_{G=E} R_2) \cup (\rho_{DG \leftarrow AC}(\pi_{ACEF}(R_1 \bowtie_{B=F} R_3)))$$

15

Soluzione

- Questa espressione non è esprimibile in calcolo delle tuple a causa dell'unione tra due diverse tabelle.
- In calcolo dei domini l'espressione diventa:

$$\{ D: d, G: g, E: e, F: f \mid R_3(E:e, F:f) \wedge ((R_2(D: d, G: g) \wedge (g=e)) \vee (R_1(A: d, B: b, C: g) \wedge (b=f))) \}$$

h

16

6

- Si consideri uno schema relazionale contenente le relazioni $R1(ABC)$, $R3(EF)$. Formulare in calcolo relazionale su tuple l'interrogazione realizzata in algebra relazionale dalla seguente espressione:

$$(\rho_{DG \leftarrow AC}(\pi_{ACEF}(R1 \triangleright \triangleleft_{B=F} R3)))$$

17

Soluzione

- $\{i.(D,G), s.(E,F) \mid i(R1), g(R1), s(R3) \mid g.B=s.F \wedge i.G=g.C \wedge i.D=g.A\}$

18

Esercizio 7

-
- Se l'operatore ρ non esistesse nell'algebra relazionale, ci sarebbero query che non potrebbero essere espresse?
-
-

19

Soluzione

- Sì quelle in cui la stessa relazione compare più volte.

20

Esercizio 8

- Si consideri la seguente basi di dati
 - Skipper(codiceV, nome, cognome, codiceB)
 - Barca (codiceB, TipoB, porto)
- Scrivere una espressione in algebra relazionale per elencare i codici dei velisti che hanno fatto da skipper su ogni barca a vela ormeggiata a Lerici.

21

Soluzione

Velista = $\Pi_{\text{codiceV}, \text{codiceB}}$
 $(\sigma_{\text{TipoB}='Vela'}(\text{Barca}) \bowtie \text{Skipper})$

BarcaLerici = $\Pi_{\text{codiceB}}(\sigma_{\text{porto}='Lerici' \wedge \text{TipoB}='Vela'}(\text{Barca}))$

▪ $\Pi_{\text{CodiceV}}(\text{Velista}) -$

$\Pi_{\text{CodiceV}}((\Pi_{\text{CodiceV}}(\text{Velista}) \bowtie \text{BarcaLerici}) - \text{Velista})$

22

Divisione

- La divisione è un'operazione binaria che si applica a due relazioni r ed s , rispettivamente con schemi $R = (A_1, \dots, A_m)$ ed $S = (A_j, \dots, A_k)$, dove A_j, \dots, A_k è un sottoinsieme proprio di A_1, \dots, A_m
- La relazione risultante, $r \div s$, è detta il quoziente della divisione di r per s e ha come schema l'insieme degli attributi di R non compresi in S . In essa saranno presenti tutte (e solo) le tuple che possano essere combinate con una qualsiasi tupla di s in modo tale che la tupla risultante appartenga ad r .

23

- Velista \div BarcaLerici

24

9

- Si consideri il seguente schema di base di dati
Aeroporto (Città, Nazione)
Volo (IdVolo, TipoAereo, GiornoSettimana,
CittàPartenza, OraPartenza, CittàArrivo,
OraArrivo)
Aereo (TipoAereo, NumPasseggeri, QuantMerci)
- Scrivere un'espressione in algebra relazionale
che elenchi per volo e giorno della settimana i
collegamenti diretti tra Roma e Bucarest.

25

Soluzione

$$\Pi_{\substack{\text{IdVolo}, \\ \text{GiornoSettimana}}} \left(\sigma_{((\text{CittàPartenza}=\text{"Roma"} \wedge \text{CittàArrivo}=\text{"Bucarest"}) \vee (\text{CittàPartenza}=\text{"Bucarest"} \wedge \text{CittàArrivo}=\text{"Roma"}))} (\text{Volo}) \right)$$

Lo studente definisca la query precedente nel calcolo relazionale dei domini

26

Soluzione

$\{ \text{IdVolo:iv, GiornoSettimana: gs} \mid$
 $\text{Volo}(\text{IdVolo:iv, TipoAereo: ta, GiornoSettimana: gs, CittàPartenza: cp, OraPartenza: op,}$
 $\text{CittàArrivo: ca, OraArrivo: oa}) \wedge ((\text{ca}=\text{"Roma"} \wedge \text{cp}=\text{"Bucarest"}) \vee (\text{cp}=\text{"Roma"} \wedge \text{ca}=\text{"Bucarest"})) \}$

Lo studente scriva un'espressione in algebra relazionale che elenchi tutte le città con cui è collegata direttamente Pisa sia come città di arrivo che come città di partenza.

27

Soluzione

$\Pi_{\text{Città}} \left(\sigma_{\text{CittàArrivo}=\text{"Pisa"}} \left(\rho_{\text{Città} \leftarrow \text{CittàPartenza}} (\text{Volo}) \right) \right) \cup$
 $\Pi_{\text{Città}} \left(\sigma_{\text{CittàPartenza}=\text{"Pisa"}} \left(\rho_{\text{Città} \leftarrow \text{CittàArrivo}} (\text{Volo}) \right) \right)$

Lo studente definisca la query precedente anche nel calcolo relazionale sulle tuple.

28

Soluzione

- Non si può fare

29

10

- Dato il seguente schema:
 - AEROPORTO(Città, Nazione, NumPiste)
 - VOLO(IdVolo, GiornoSett, CittàPart, OraPart, CittàArr, OraArr, TipoAereo)
 - AEREO(TipoAereo, NumPasseggeri, QtaMerci)scrivere in algebra relazionale la interrogazione che permette di determinare gli aeroporti italiani che hanno solo voli interni.

30

Soluzione

$$\Pi_{\text{CittàPart}} \sigma_{\text{Nazione}='Italia'} (\text{AEROPORTO} \bowtie_{\text{Città}=\text{CittàPart}} \text{VOLO})$$

$$-$$

$$\Pi_{\text{CittàPart}} \sigma_{\text{Nazione}='Italia'} (\text{AEROPORTO} \bowtie_{\text{Città}=\text{CittàPart}} \text{VOLO}$$

$$\bowtie_{\text{CittàArr}=\text{CittàK}} \rho_{\text{CittàK, NazioneK, nK} \leftarrow \text{Città, Nazione, NumPiste}}$$

$$(\sigma_{\text{Nazione} \neq 'Italia'} (\text{AEROPORTO})))$$

31

11

- Si consideri il seguente insieme di relazioni:
- *Film*(CodFilm, Titolo, CodRegista, Anno)
- *Produzione*(CasaProduzione, Nazionalita, CodFilm, Costo)
- *Attore*(CodAttore, Cognome, Nome, Sesso, DataNascita, Nazionalita)
- *Interpretazione*(CodFilm, CodAttore, Personaggio)
- *Regista*(CodRegista, Cognome, Nome, Sesso, DataNascita, Nazionalita)

32

11.a

Definire in algebra relazionale una query che produca la lista dei titoli dei film che "Marcello Mastroianni" ha interpretato.

33

Soluzione

$$\begin{aligned} & \pi_{\text{Titolo}} (\\ & \pi_{\text{CodFilm}} (\\ & \pi_{\text{CodAttore}} (\sigma_{(\text{Nome} = \text{"Marcello"}) \wedge (\text{Cognome} = \text{"Mastroianni"})} (\text{Attore})) \bowtie \pi_{\text{CodFilm, CodAttore}} (\text{Interpretazione})) \\ & \bowtie \pi_{\text{CodFilm, Titolo}} (\text{Film})) \end{aligned}$$

34

Esprimere la stessa query nel calcolo relazionale dei domini e delle tuple.

$\{ \text{Titolo: } t \mid \text{Film}(\text{CodFilm: } fn, \text{Titolo: } t, \text{CodRegista: } d, \text{Anno: } y) \wedge$
 $\text{Attore}(\text{CodAttore: } an, \text{Cognome: } cogn, \text{Nome: } n, \text{Sesso: } s,$
 $\text{DataNascita: } b, \text{Nazionalità: } naz) \wedge$
 $\text{Interpretazione}(\text{CodFilm: } fn, \text{CodiceAttore: } an, \text{Personaggio: } ch) \wedge \text{Regista..} \wedge \text{Produzione...} \wedge$
 $(cogn = \text{"Mastroianni"}) \wedge (n = \text{"Marcello"}) \}$

$\{ F.\text{titolo} \mid F(\text{Film}), A(\text{Attore}), I(\text{Interpretazione}) \mid$
 $F.\text{CodFilm} = I.\text{CodFilm} \wedge A.\text{CodAttore} = I.\text{CodAttore} \wedge$
 $A.\text{Cognome} = \text{"Mastroianni"} \wedge A.\text{Nome} = \text{"Marcello"} \}$

35

11.b

- Definire in algebra relazionale una query che produca la lista dei titoli dei film che "Marcello Mastroianni" ha interpretato senza "Sofia Loren".

36

Soluzione

$$\begin{aligned} & \pi_{\text{Titolo}} \left(\pi_{\text{CodFilm}} \left(\left(\pi_{\text{CodAttore}} \left(\sigma_{(\text{Nome}="Marcello") \wedge (\text{Cognome}="Mastroianni") } \right. \right. \right. \right. \\ & \left. \left. \left. \left. (\text{Attore}) \right) \right) \right) \right) \bowtie \pi_{\text{CodFilm,CodAttore}} (\text{Interpretazione}) \bowtie \pi_{\text{CodFilm,Titolo}} (\text{Film}) - \\ & \pi_{\text{Titolo}} \left(\pi_{\text{CodFilm}} \left(\left(\pi_{\text{CodAttore}} \left(\sigma_{(\text{Nome}="Sofia") \wedge (\text{Cognome}="Loren") } \right. \right. \right. \right. \\ & \left. \left. \left. \left. (\text{Attore}) \right) \right) \right) \right) \bowtie \text{Interpretazione} \bowtie \text{Film} \end{aligned}$$

37

Esprimere la stessa query nel calcolo dei domini.
Errori??

```
{Titolo: t |
  Film(CodFilm: cf, Titolo: t, CodRegista: cr, Anno: a) ^
  Attore(CodAttore:ca, Cognome:c, Nome:n, Sesso:s, DataNascita:dn, Nazionalità:nz) ^
  Interpretazione(CodFilm: cf, CodAttore:ca, Personaggio:p, SessoPersonaggio:sp) ^
  ¬ (∃ca'(∃p'(∃sp'(∃c'(∃n'(∃dn'(∃nz'(
    Film(CodFilm: cf, Titolo: t, CodRegista: cr, Anno: a) ^
    Attore(CodAttore:ca', Cognome:c', Nome:n', Sesso:s', DataNascita:dn', Nazionalità:nz') ^
    Interpretazione(CodFilm: cf, CodAttore:ca', Personaggio:p', SessoPersonaggio:sp') ^ c'
    ="Loren" ^ n' ="Sofia"))))))))}}
```

38

Esercizio 13

- Si consideri una base di dati sulle relazioni:
 $R1(A,B,C)$ e $R2(D,E, F)$.
- Dire se le seguenti interrogazioni sono equivalenti

- $\pi_{AB}(R1) - \pi_{AB}(R1 \bowtie_{C=D} R2)$
- $\pi_{AB}(\sigma_{D \text{ IS NULL}}(R1 \bowtie_{\text{left}(C=D)} R2))$

- R1

- a1 b1 c1

- a1 b2 c3

- a1 b1 c2

R2

c1 e1 f1

c3 e2 f2

- Se A è chiave?

Esercizio 14

- Indicare nome, cognome e specializzazione dei medici che hanno effettuato visite eccetto che il giorno 1° Marzo 2013

```
SELECT M.Nome, M.Cognome, M.Specializzazione
FROM Medico M
WHERE M.Matricola IN
    ( SELECT V.Medico
      FROM Visita V
    )
AND M.Matricola NOT IN
    ( SELECT V.Medico
      FROM
        Visita V
      WHERE V.Data = '2013-03-01'
    );
```

43

- risolvere la stessa query in algebra relazionale

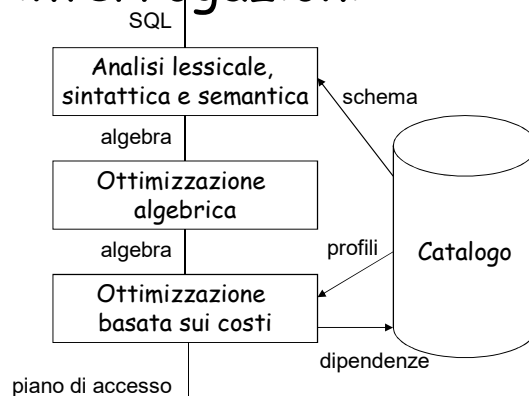
$$\pi_{N,C,S}(\text{Medico} \supset \triangleleft_{\text{Mat=Medico}} \text{IN})$$

except

$$(\pi_{\text{Medico}}(\text{Visita}) - \pi_{\text{Medico}}(\sigma_{\text{Data}='2013-03-01'}(\text{Visita})))$$

44

Il processo di esecuzione delle interrogazioni



45

Ottimizzazione algebrica

- Il termine ottimizzazione è improprio perché il processo utilizza euristiche
- Euristiche fondamentali:
 - selezioni e proiezioni il più presto possibile (per ridurre le dimensioni dei risultati intermedi):
 - "push selections down"
 - "push projections down"

46

Una procedura euristica

- Decomporre le selezioni congiuntive in successive selezioni atomiche
- Anticipare il più possibile le selezioni
- In una sequenza di selezioni, anticipare le più selettive
- Anticipare il più possibile le proiezioni (anche introducendone di nuove)
- Combinare prodotti cartesiani e selezioni per formare join

47

Esempio

R1(ABC), R2(DEF), R3(GHI)

```
SELECT  A , E
FROM    R1, R2, R3
WHERE   C=D AND B>100 AND F=G AND H=7 AND I>2
```

viene trasformata in

- prodotto cartesiano (FROM)
- selezione (WHERE)
- proiezione (SELECT)

```
PROJAE (SELB>100 AND H=7 AND I>2 AND C=D AND F=G (
  (R1 JOIN R2 JOIN R3)
```

48

Esempio, continua

- viene trasformata in

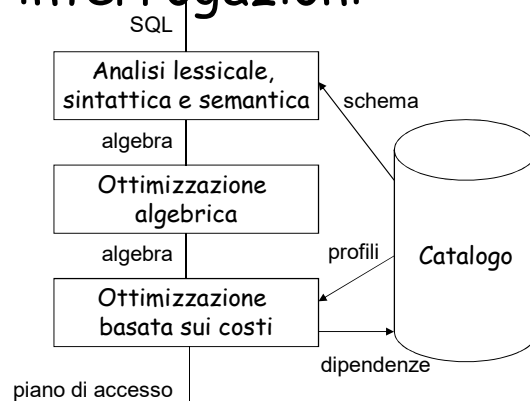
$$\text{PROJ}_{AE}(\text{PROJ}_{AEF}((\text{PROJ}_{AC}(\text{SEL}_{B>100}(R1))) \text{JOIN}_{C=D} R2) \text{JOIN}_{F=G} \text{PROJ}_G(\text{SEL}_{I>2}(\text{SEL}_{H=7}(R3))))$$

l'ordine in cui si fanno i join va valutato:

dopo l'ottimizzazione algebrica (si ottengono tutte le espressioni minimali equivalenti) va fatta quella in base ai costi

49

Il processo di esecuzione delle interrogazioni



50

"Profili" delle relazioni

- Informazioni quantitative:
 - cardinalità di ciascuna relazione
 - dimensioni delle tuple
 - dimensioni dei valori
 - numero di valori distinti degli attributi
 - valore minimo e massimo di ciascun attributo
- Utilizzate nella fase finale dell'ottimizzazione, per stimare le dimensioni dei risultati intermedi

51

Esecuzione delle operazioni

- I DBMS implementano gli operatori dell'algebra relazionale (o meglio, loro combinazioni) per mezzo di operazioni di livello abbastanza basso, che però possono implementare vari operatori "in un colpo solo"
- Operatori fondamentali:
 - accesso diretto
 - scansione
- A livello più alto:
 - ordinamento
- Ancora più alto
 - Join, l'operazione più costosa

52

Ottimizzazione basata sui costi

- Un problema articolato, con scelte relative a:
 - operazioni da eseguire (es.: scansione o accesso diretto?)
 - ordine delle operazioni (es. join di tre relazioni; ordine?)
 - i dettagli del metodo (es.: quale metodo di join)
- Architetture parallele e distribuite aprono ulteriori gradi di libertà

53

Ordine del join

- Join associativo
- La scelta dell'ordine delle coppie a cui si applica prima l'operatore è rilevante
- Attributo di join chiave

54

Metodi di join

- Nested loop
- Merge scan
- Hash-based

55

nested loop

- per ogni tupla nella tabella esterna si esaminano tutte le tuple di quella interna per verificare la condizione di join.
- date R ed S , si hanno due possibilità R esterna o R interna
- il costo dipende dal numero di accessi e dal fatto che la tabella esterna sia piccola

56

merge scan

- si ordinano le tabelle in base agli attributi di join
- si trova l'elemento della seconda tabella da cui partire rispetto al primo elemento della prima tabella e poi si continua da lì
- il costo è l'ordinamento

57

hash

- è costruita una tabella ad accesso hash sull'attributo di join per cercare le tuple corrispondenti più velocemente
- ogni riga della prima tabella è inserita in una tabella aggiuntiva, poi si cercano le tuple della seconda con la stessa codifica hash
- serve memoria ed è migliore del nested loop nel caso di equijoin

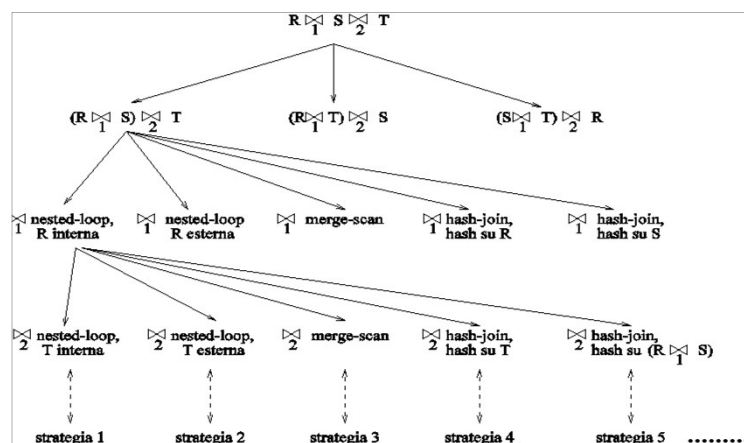
58

Il processo di ottimizzazione

- Si costruisce un albero di decisione con le varie alternative ("**piani di esecuzione**")
- Si valuta il costo di ciascun piano
- Si sceglie il piano di costo minore
- L'ottimizzatore trova di solito una "buona" soluzione, non necessariamente l'"ottimo"

59

Un albero di decisione



60