Optimizacija poizvedb

Hevristični postopek in primeri

Osnovne tehnike optimizacije poizvedb

- **Hevristična pravila** za pravilno razvrščanje operacij poizvedbe.
- Primerjava relativnih stroškov različnih strategij, kjer izberemo najcenejšo, t.j. predstavlja najmanjšo porabo sredstev.

Procesiranje poizvedbe (1)

- Procesiranje poizvedbe sestavlja:
 - Razčlenjevanje oz. dekompozicija,
 - Cilj je pretvoriti poizvedbo iz visoko nivojskega jezika (SQL) v relacijsko algebro.
 - 2. Optimizacija,
 - o Izbira najučinkovitejše izvedbene strategije.,
 - 3. Generiranje kode,
 - 4. Izvedba.

Procesiranje poizvedbe (2)

Dekompozicija

- o Procesiranje poizvedbe sestavlja:
 - Razčlenjevanje oz. dekompozicija,
 - Analiza,
 - o Normalizacija,
 - Semantična analiza,
 - Poenostavitev,
 - Reorganizacija poizvedbe.

Procesiranje poizvedbe (3)

Dekompozicija » Analiza (1)

- Analiza zajema leksikalno in sintaktično analizo.
- Odatno je potrebno preveriti:
 - Ali so relacije in atributi zapisani v sistemskem katalogu?
 - Ali so operacije poizvedbe primerne glede na tip objekta, nad katerim se izvajajo?

Procesiranje poizvedbe (4)

Dekompozicija » Analiza (2)

- Navadno po opravljeni analizi poizvedbo pretvorimo v drevo poizvedbe za lažjo nadaljnjo obravnavo:
 - Osnovna relacija postane list drevesa.
 - Vmesna relacija, ki je rezultat operacij algebre, postane vozlišče drevesa.
 - Rezultat poizvedbe je v korenu drevesa.
 - Zaporedje operacij poteka od listov drevesa do korena.

Procesiranje poizvedbe (5)

Dekompozicija » Normalizacija

- Normalizacija zajema pretvorbo poizvedbe v normalizirano obliko, ki jo je lažje obdelovati.
 - Pretvori se predikatni (WHERE) del poizvedbe.
 - Obstajata dva načina:
 - Konjunktivna NO: Zaporedje konjunkcij (pogoji povezani z OR operatorjem) je povezanih z operatorjem AND.
 - Disjunktivna NO: Zaporedje disjunkcij (pogoji povezani z AND operatorjem) je povezanih z operatorjem OR.

Procesiranje poizvedbe (6)

Dekompozicija » Semantična analiza (1)

- Pri semantični analizi želimo preprečiti napačno formulirane (komponente ne prispevajo k ciljnemu rezultatu) ali kontradiktorne poizvedbe (predikatu ne zadošča nobena n-terica).
- Algoritmi za preverjanje semantične pravilnosti potrebujejo množico poizvedb, ki ne vsebujejo disjunkcije in negacije!

Procesiranje poizvedbe (7)

Dekompozicija » Semantična analiza (2)

- Metodi za ugotavljanje semantičnih nepravilnosti:
 - Graf povezav relacij → preverjanje napačnih formulacij
 - Ograf povezav normaliziranih atributov → preverjanje kontradikcij

Procesiranje poizvedbe (8)

Dekompozicija » Semantična analiza (3)

• Graf povezav relacij:

- Za vsako relacijo in za rezultat kreiramo vozlišče.
- Za relacije v stiku kreiraj povezavo med vozlišči.
- Za izvor v operacijah projekcije kreiraj povezave med vozlišči.
- Če graf ni povezan (t.j. obstajajo nepovezani deli), je poizvedba napačno formulirana!

Procesiranje poizvedbe (9)

Dekompozicija » Semantična analiza (4)

• Graf normaliziranih atributov:

- Za vsako referenco na atribut ali konstanto 0 kreiraj vozlišče.
- Med vozlišči, ki predstavljajo stik, kreiraj usmerjeno povezavo.
- Za vsako selekcijo kreiraj usmerjeno povezavo med vozliščema, ki predstavlja atribut selekcije in vozliščem, ki predstavlja konstanto 0.

Procesiranje poizvedbe (10)

Dekompozicija » Semantična analiza (5)

• Graf normaliziranih atributov:

- Povezave a → b uteži z vrednostjo c, če gre za pogoj neenakosti tipa a ≤ b + c.
- Povezave 0 → a uteži z vrednostjo –c, če gre za pogoj neenakosti tipa a ≥ c.
- Če graf vsebuje cikel z negativnim seštevkom uteži, potem je poizvedba kontradiktorna!

Procesiranje poizvedbe (11)

Dekompozicija » Poenostavitev

- Pri poenostavitvi se ukvarjamo z:
 - detekcijo odvečnih kvalifikatorjev,
 - o odpravo ponavljajočih sklopov in
 - transformacijo poizvedbe v semantično ekvivalentno, vendar enostavnejšo in za računanje učinkovitejšo obliko!
- V tej fazi se upoštevajo omejitve dostopa, definicije pogledov ter omejitve skladnosti, kar včasih posledično vpliva na pojav redundance.

Procesiranje poizvedbe (12)

Dekompozicija » Reorganizacija

 Pri reorganizaciji se poizvedba reorganizira z namenom zagotavljanja učinkovitejše implementacije.

Naloga #1

Ugotavljanje ali je poizvedba semantično pravilna?

Ugotavljanje semantične pravilnosti

HOTEL (<u>hotelNo: number</u>, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo</u>: number, <u>hotelNo</u>: number, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

GUEST (guestNo: number, guestName: varchar2(30),

guestAddress: varchar2(30))

 Ugotovite, ali je naslednja poizvedba semantično pravilna?

SELECT r.type, r.price

FROM ROOM r, HOTEL h

WHERE r.hotel_number = h.hotel_number **AND**

h.hotel_name = 'Habakuk' **AND**

r.type > 100;

Ugotavljanje semantične pravilnosti (2)

- Potrebno je izvesti vse korake:
 - o analiza,
 - o normalizacija,
 - o semantična analiza,
 - poenostavitev in
 - o reorganizacija poizvedbe.

Ugotavljanje semantične pravilnosti (3)

HOTEL (<u>hotelNo: number</u>, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo: number</u>, <u>hotelNo: number</u>, type: varchar2(30), price: real)

BOOKING (<u>hotelNo</u>: number, <u>guestNo</u>: number, <u>dateFrom</u>: date,

dateTo: date, roomNo: number)

GUEST (guestNo: number, guestName: varchar2(30),

guestAddress: varchar2(30))

SELECT r.type, r.price

FROM ROOM r, HOTEL h

WHERE r.hotel_number = h.hotel_number **AND**

h.hotel_name = 'Habakuk' AND

r.type > 100;

 Pri analizi najprej preverimo leksikalno in sintaktično analizo (podobno kot naredijo prevajalniki).

Ugotavljanje semantične pravilnosti (4)

HOTEL (<u>hotelNo: number</u>, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo: number</u>, <u>hotelNo: number</u>, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

GUEST (guestNo: number, guestName: varchar2(30),

guestAddress: varchar2(30))

SELECT r.type, r.price

FROM ROOM r, HOTEL h

WHERE r.hotel_number = h.hotel_number AND

h.hotel_name = 'Habakuk' AND

r.type \rightarrow 100;

 Dodatno preveri ali so relacije in atributi zapisani v sistemskem katalogu in če so operacije primerne glede na tip objekta.

Naloga #2

Ugotavljanje ali je poizvedba semantično pravilna?

Ugotavljanje semantične pravilnosti

HOTEL (<u>hotelNo: number</u>, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo</u>: number, <u>hotelNo</u>: number, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

GUEST (guestNo: number, guestName: varchar2(30),

guestAddress: varchar2(30))

 Ugotovite, ali je poizvedba, ki za vsak hotel iz Pirana izpiše podatke o gostih in sobah, ki so jim pripadale, semantično pravilna?

SELECT g.guestName, g.guestAddress

FROM ROOM r, HOTEL h, BOOKING b, GUEST g

WHERE r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND

h.city = 'Piran' **AND** b.dateTo < '15.08.2006';

Ugotavljanje semantične pravilnosti (2)

HOTEL (<u>hotelNo: number</u>, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo</u>: number, <u>hotelNo</u>: number, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

GUEST (<u>guestNo</u>: number, guestName: varchar2(30),

guestAddress: varchar2(30))

SELECT g.guestName, g.guestAddress

FROM ROOM r, HOTEL h, BOOKING b, GUEST g

WHERE r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND

h.city = 'Piran' **AND** b.dateTo < '15.08.2006';

 Pri analizi najprej preverimo leksikalno in sintaktično analizo (podobno kot naredijo prevajalniki).

Ugotavljanje semantične pravilnosti (3)

HOTEL (hotelNo: number, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo: number</u>, <u>hotelNo: number</u>, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

(guestNo: number, guestName: varchar2(30), **GUEST**

guestAddress: varchar2(30))

SELECT g.guestName, g.guestAddress

ROOM r, HOTEL h, BOOKING b, GUEST g FROM

r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND WHERE

h.city = 'Piran' **AND** b.dateTo < '15.08.2006';

 Dodatno preveri ali so relacije in atributi zapisani v sistemskem katalogu in če so operacije primerne glede na tip objekta. 🗸



Ugotavljanje semantične pravilnosti (4)

HOTEL (<u>hotelNo: number</u>, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo</u>: number, <u>hotelNo</u>: number, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

GUEST (<u>guestNo</u>: number, guestName: varchar2(30),

guestAddress: varchar2(30))

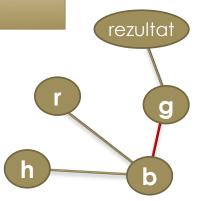
SELECT g.guestName, g.guestAddress

FROM ROOM r, HOTEL h, BOOKING b, GUEST g

WHERE r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND

h.city = 'Piran' **AND** b.dateTo < '15.08.2006';

 Pri semantični analizi preverimo graf povezav relacij.



Procesiranje poizvedbe (13)

Optimizacija (1)

o Procesiranje poizvedbe sestavlja:

Optimizacija

- Izbira najučinkovitejše izvedbene strategije za procesiranje poizvedbe, kjer obstaja dva pogleda:
 - doseči želimo najkrajši čas izvajanja poizvedbe,
 - želimo čim bolj izkoristiti razpoložljiva sredstva (npr. vzporedne operacije).

Procesiranje poizvedbe (14)

Optimizacija (2)

- Spoznali bomo hevristično optimizacijo poizvedb, kjer uporabljamo hevristična pravila.
 - Na ta način izraz v relacijski algebri pretvorimo, kjer je končni rezultat učinkovitejši.

Procesiranje poizvedbe (15)

Optimizacija (3)

Transformacijsko pravilo 1

- Kaskada selekcije: $\sigma_{p \wedge q \wedge r}(R) = \sigma_p(\sigma_q(\sigma_r(R)))$
 - Primer: $\sigma_{\text{branchNo} = 'B003' \land salary > 15000}(\text{Staff}) = \sigma_{\text{branchNo} = 'B003'}(\sigma_{\text{salary} > 15000}(\text{Staff}))$

Transformacijsko pravilo 2

- Komutativnost oper. selek.: $\sigma_p(\sigma_q(R)) = \sigma_q(\sigma_p(R))$
 - Primer: $\sigma_{\text{branchNo} = 'B003'}(\sigma_{\text{salary} > 15000}(\text{Staff})) = \sigma_{\text{salary} > 15000}(\sigma_{\text{branchNo} = 'B003'}(\text{Staff}))$

Procesiranje poizvedbe (16)

Optimizacija (4)

o Transformacijsko pravilo 3

- Pri zaporedju operacij projekcije je pomembna le zadnja: $\Pi_{I}\Pi_{M}$... $\Pi_{N}(R) = \Pi_{I}(R)$
 - Primer: $\Pi_{\text{IName}}\Pi_{\text{branchNo, IName}}$ (Staff) = Π_{IName} (Staff)

o Transformacijsko pravilo 4

• Komutativnost operacij selekcije in projekcije:

$$\Pi_{Ai, ..., Am}(\sigma_p(R)) = \sigma_p(\Pi_{Ai, ..., Am}(R)), \text{ kjer } p \in \{A_1, A_2, ..., A_m\}$$

• Primer:
$$\Pi_{\text{fName, IName}}$$
 ($\sigma_{\text{IName='Beech'}}$ (Staff)) = $\sigma_{\text{IName='Beech'}}$ ($\Pi_{\text{fName, IName}}$ (Staff))

Procesiranje poizvedbe (17)

Optimizacija (6)

- Transformacijsko pravilo 5
 - Komutativnost theta stika, kartezijskega produkta, equijoin in naravni stik: $\mathbb{R} \bowtie_{p} \mathbb{S} = \mathbb{S} \bowtie_{p} \mathbb{R}$
 - Primer: Staff⋈_{staff.branchNo} = branch.branchNo</sub> Branch = Branch⋈_{staff.branchNo} = branch.branchNo

Procesiranje poizvedbe (18)

Optimizacija (7)

Transformacijsko pravilo 6

• Komutativnost selekcije in theta stika (ali kartezijskega produkta): $\sigma_p(R \bowtie_r S) = (\sigma_p(R))\bowtie_r S$, kjer $p \in \{A_1, A_2, ..., A_n\}$ in $\sigma_{p \land q}(R \bowtie_r S) = (\sigma_p(R))\bowtie_r (\sigma_q(S))$ • Primer: $\sigma_{position = 'Manager' \land city = 'London'}$ (Staff $\bowtie_{Staff.branchNo}$ = Branch.branchNo Branch) = $(\sigma_{position = 'Manager'}$ (Staff)) $\bowtie_{Staff.branchNo} = (\sigma_{position = 'Manager'})$ (Staff)) $\bowtie_{Staff.branchNo} = (\sigma_{position = Manager'})$ (Staff)

Procesiranje poizvedbe (19)

Optimizacija (8)

Transformacijsko pravilo 7

- Komutativnost projekcije in theta stika (ali kartezijskega produkta): $\Pi_{L_1 \cup L_2}(R \bowtie_r S) = (\Pi_{L_1}(R)) \bowtie_r (\Pi_{L_2}(S))$, kjer je $L = L_1 \cup L_2$ in L_1 atributi relacije R in L_2 atributi relacije S
 - Primer: $\Pi_{\text{position, city, branchNo}}$ (Staff $\bowtie_{\text{Staff.branchNo}} = Branch.branchNo}$ Branch) = $(\Pi_{\text{position, branchNo}}$ (Staff)) $\bowtie_{\text{Staff.branchNo}} = Branch.branchNo}$ ($\Pi_{\text{city, branchNo}}$ (Branch))

Procesiranje poizvedbe (20)

Optimizacija (9)

Transformacijsko pravilo 7 (dodatek)

• Če pogoj stika vsebuje atribute, ki jih ni v L ($M = M_1 \cup M_2$ in M_1 atributi relacije R in M_2 atributi relacije S): $\Pi_{L_1 \cup L_2}(R \bowtie_r S) = \Pi_{L_1 \cup L_2}(\Pi_{L_1 \cup M_1}(R)) \bowtie_r (\Pi_{L_2 \cup M_2}(S))$

• Primer: $\Pi_{\text{position, city}}$ (Staff $\bowtie_{\text{Staff.branchNo}} = Branch.branchNo}$ Branch) = $\Pi_{\text{position, city}}$ (($\Pi_{\text{position, branchNo}}$ Staff)) $\bowtie_{\text{Staff.branchNo}} = Branch.branchNo}$ ($\Pi_{\text{city, branchNo}}$ (Branch))

Procesiranje poizvedbe (21)

Optimizacija (10)

Transformacijsko pravilo 8

Komutativnost operacij unije in preseka:

$$R \cup S = S \cup R \text{ in } R \cap S = S \cap R$$

Transformacijsko pravilo 9

 Komutativnost operacij selekcije in operacij množic:

$$\sigma_{p}(R \cup S) = \sigma_{p}(S) \cup \sigma_{p}(R)$$

$$\sigma_{p}(R \cap S) = \sigma_{p}(S) \cap \sigma_{p}(R)$$

$$\sigma_{p}(R - S) = \sigma_{p}(S) - \sigma_{p}(R)$$

Procesiranje poizvedbe (22)

Optimizacija (11)

- o Transformacijsko pravilo 10
 - Komutativnost projekcije in unije:

$$\Pi_{l}(R \cup S) = \Pi_{l}(S) \cup \Pi_{l}(R)$$

- Transformacijsko pravilo 11
 - Asociativnost theta stika in kartezijskega produkta: $(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$
 - Primer: (Staff⋈_{Staff.staffNo} = PropertyForRent.staffNo
 PropertyForRent
 MownerNo = Owner.ownerNo ∧ staff.lName = Owner.lName
 Owner =
 Staff⋈_{staff.staffNo} = PropertyForRent.staffNo ∧
 staff.lName=|Name
 (PropertyForRent⋈_{ownerNo}Owner)

Procesiranje poizvedbe (23)

Optimizacija (12)

- Transformacijsko pravilo 12
 - Asociativnost unije in preseka:

$$(R \cup S) \cup T = S \cup (R \cup T)$$
 in $(R \cap S) \cap T = S \cap (R \cap T)$

Procesiranje poizvedbe (24)

Optimizacija (13)

- o Poznamo 5 strategij hevrističnega procesiranja:
 - Strategija 1
 - Selekcijo izvedi čim prej (zmanjša se kardinalnost)
 - Selekcije nad isto relacijo naj bodo skupaj
 - Uporabi transformacije 2, 4, 6 in 9
 - o Strategija 2
 - Združi kartezijski produkt s selekcijami, ki so hkrati stični pogoj

Procesiranje poizvedbe (25)

Optimizacija (14)

Strategija 3

- Relacije čim bolj zreduciramo, preden uporabimo binarne operacije – najbolj omejevalne selekcije se upoštevajo na začetku
- Podobno asociativnost uporabimo na theta stiku, kjer se manjši stiki izvedejo prej

Strategija 4

- Projekcije uredimo kaskadno (pravilo 3) in projekcijo premaknemo bliže listom drevesa (pravila 4, 7 in 10)
- Projekcijo izvedemo na samem začetku
- Projekcije nad istimi relacijami naj bodo skupaj

Procesiranje poizvedbe (26)

Optimizacija (15)

Strategija 5

- Če se rezultat pojavi večkrat in ni prevelik, ga shrani za nadaljnjo uporabo
- Če je možno se rezultat shrani v primarni pomnilnik, sicer se izračuna kaj stane več: branje iz sekundarnega pomnilnika ali ponovni izračun izraza

Naloga #3

Optimizirajte poizvedbo z uporabo hevrističnih pravil

Optimizacije poizvedbe (1)

HOTEL (hotelNo: number, hotelName:varchar2(30), city:varchar2(30))

ROOM (<u>roomNo</u>: number, <u>hotelNo</u>: number, type: varchar2(30), price: real)

BOOKING (hotelNo: number, guestNo: number, dateFrom: date,

dateTo: date, roomNo: number)

GUEST (guestNo: number, guestName: varchar2(30),

guestAddress: varchar2(30))

 Z uporabo hevrističnih pravil optimizirajte naslednjo poizvedbo!

SELECT r.roomNo, r.type, r.price

FROM ROOM r, BOOKING b, HOTEL h

WHERE r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND

h.hotelName = 'Habakuk' **AND** r.price > 100;

Optimizacija poizvedbe (2)

 Najprej SQL poizvedbo pretvorimo v relacijsko algebro:

SELECT r.roomNo, r.type, r.price

FROM ROOM r, BOOKING b, HOTEL h

WHERE r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND

h.hotelName = 'Habakuk' AND r.price > 100;



 $\Pi_{r.roomNo,\ r.type,\ r.price} (\sigma_{r.roomNo} = b.roomNo \land b.hotelNo = h.hotelNo \land h.hotelName = 'Habakuk' \land r.price > 100 ((r x b) x h))$

Optimizacija poizvedbe (3)

- Nato poizvedbo v relacijski algebri predstavimo kot kanonično drevo relacijske algebre:
 - Osnovne relacije postanejo listi drevesa,
 - Vmesne relacije so vozlišča drevesa,
 - Rezultat je v korenu drevesa,
 - Zaporedje operacij si sledi od listov do korena.

Optimizacija poizvedbe (4)

 $\Pi_{r.roomNo,\,r.type,\,r.price}(\sigma_{r.roomNo} = b.roomNo \land b.hotelNo} = h.hotelNo \land h.hotelName = 'Habakuk' \land r.price > 100((r x b) x h))$

Tr.roomNo, r.type, r.price

Tr.roomNo = b.roomNo \

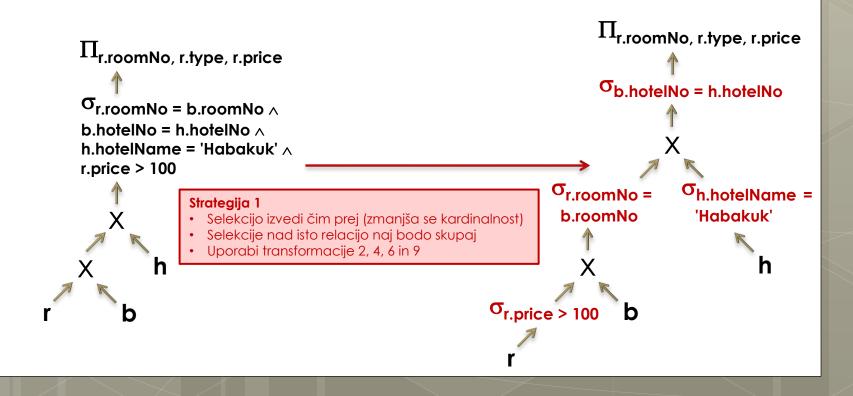
b.hotelNo = h.hotelNo \
h.hotelName = 'Habakuk' \
r.price > 100

X

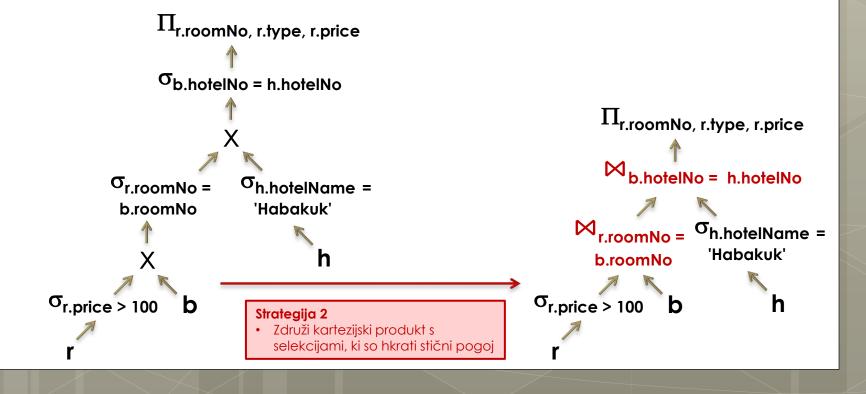
X

h

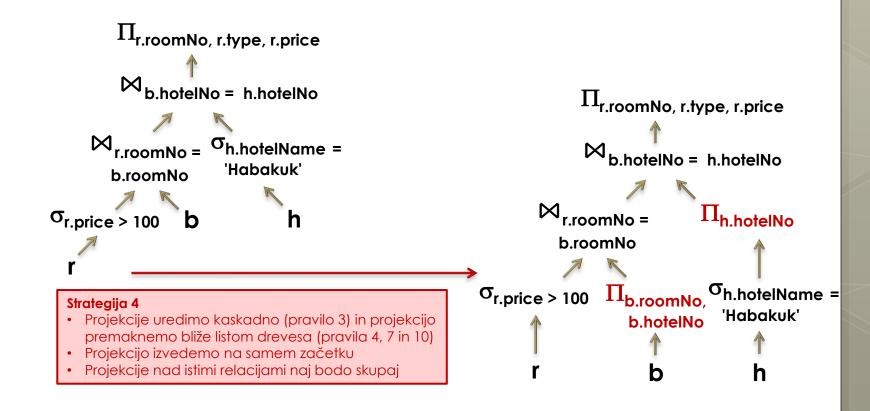
Optimizacija poizvedbe (5)



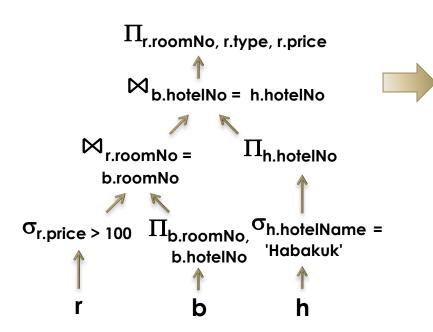
Optimizacija poizvedbe (6)



Optimizacija poizvedbe (7)



Optimizacija poizvedbe (8)



Optimizirana rešitev

```
\begin{split} &\Pi_{r.roomNo,\,r.type,\,r.price}(\\ &(\,(\sigma_{r.price}>_{100}(r))\\ &\stackrel{\bowtie}{\bowtie}_{r.roomNo}=\text{b.roomNo}\\ &(\Pi_{r.roomNo,\,b.hotelNo}(b))\,\,)\\ &\stackrel{\bowtie}{\bowtie}_{b.hotelNo}=\text{h.hotelNo}\\ &(\,\Pi_{h.hotelNo}(\sigma_{h.hotelName}=_{'Habakuk'}(h))\,\,)\\ &) \end{split}
```

Prvotna rešitev

```
\Pi_{r.roomNo, r.type, r.price} ( \\ \sigma_{r.roomNo} = b.roomNo \land \\ b.hotelNo = h.hotelNo \land \\ h.hotelName = 'Habakuk' \land r.price > 100 ( (r x b) x h) \\ )
```