

Проектовање база података — решење трећег задатка

Јун 1 2024.

Размотрити упит над табелама `Dosije` и `StudijskiProgram` као и информацију о типовима колона тих двеју табела дату испод упита. Претпоставити да је тип `DATE` 3 бајта, а величина странице 4096 бајтова.

```
SELECT d.Ime, d.Prezime
FROM   Dosije AS d JOIN
       StudijskiProgram AS sp ON d.IdPrograma = sp.Id
WHERE  sp.Naziv = 'Astronomija' AND
       d.DatDiplomiranja > '2023-10-02';
```

Indeks	Ime	Prezime	MestoRodjenja	IdPrograma	DatDiplomiranja
CHAR(9)	VARCHAR(40)	VARCHAR(40)	VARCHAR(16)	INT	DATE

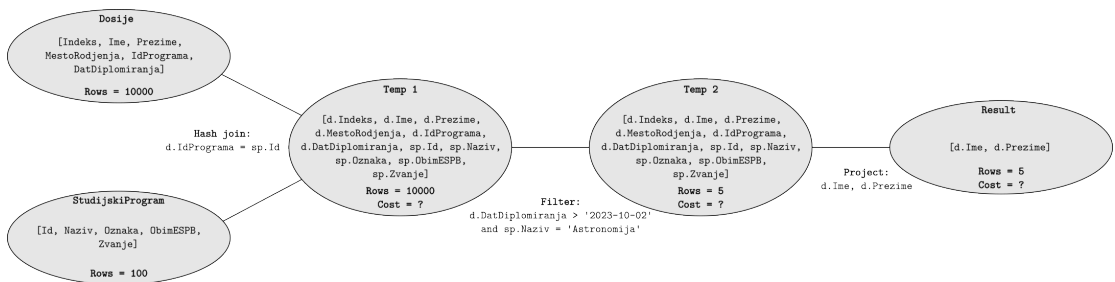
Id	Naziv	Oznaka	ObimESPB	Zvanje
INT	VARCHAR(20)	CHAR(2)	INT	VARCHAR(40)

- (а) Који тип индекса треба размотрити и евентуално додати овим двема табелама како би се упит ефикасније извршавао? Претпоставити да су колоне `Indeks` и `Id` примарни кључеви.
- Табели `Dosije` има смисла додати појединачне индексе над колонама `DatDiplomiranja` и `IdPrograma`. Због интервалног упита, `b+`-стабло је добар одабир у првом случају (мада тај индекс не би био кластерован с обзиром да већ постоји кластеровање по колони `Indeks`). У другом случају, због спајања, хеш индекс би био добра опција.
 - Табели `StudijskiProgram` има смисла додати индекс над колоном `Naziv`. Због услова једнакости, хеш индекс би могао бити кориснији.

(б) Написати пример додавања индекса табели `Dosije`.

```
CREATE INDEX DosijeProgramIndeks ON Dosije (IdPrograma) USING hash;
```

(ц) За дато стабло извршавања проценити укупан број *I/O* операција.



У просеку, у једну страницу може да стане:

- $PF(Dosije) = \lfloor \frac{4096}{9+20+20+8+4+3} \rfloor = \frac{4096}{64} = 64$ реда табеле **Dosije**, па је просечан број страница потребан за складиштење целе табеле:

$$pages(Dosije) = \lceil \frac{rows(Dosije)}{PF(Dosije)} \rceil = \lceil 156.25 \rceil = 157 \quad (1)$$

- $PF(StudijskiProgram) = \lfloor \frac{4096}{4+10+2+4+20} \rfloor = \lfloor \frac{4096}{40} \rfloor = 102$ реда табеле **StudijskiProgram**, па је просечан број страница потребан за складиштење целе табеле:

$$pages(StudijskiProgram) = \lceil \frac{rows(StudijskiProgram)}{PF(StudijskiProgram)} \rceil = \lceil 0.98 \rceil = 1 \quad (2)$$

Користимо процену цене спајања $3 \times (pages(T1) + pages(T2))$ и добијамо цену међурезултата спајања:

$$cost(Temp\ 1) = 3 \times (pages(Dosije) + pages(StudijskiProgram)) = 3 \times 158 = 474 \quad (3)$$

Да би се добио међурезултат **Temp 2**, потребно је прочитати¹ све редове из **Temp 1** и уписати 5 редова који задовољавају задати услов назад на диск:

$$PF(Temp\ 1) = \lfloor \frac{4096}{104} \rfloor = 39 \quad (4)$$

$$cost(Temp\ 2) = pages(Temp\ 1) + pages(Temp\ 2) = \lceil \frac{10000}{39} \rceil + \lceil \frac{5}{39} \rceil = 257 + 1 = 258 \quad (5)$$

За добијање цене коначног резултата **Result**, потребно је извршити пројекцију две колоне из **Temp 2**, за шта је довољно прочитати свих 5 редова (без накнадног уписа на диск):

$$cost(Result) = \lceil \frac{5}{39} \rceil = 1 \quad (6)$$

Укупна цена овог плана извршавања је $cost(Temp\ 1) + cost(Temp\ 2) + cost(Result) = 474 + 258 + 1 = 733$.

(д) За исто стабло извршити процену у случају коришћења *block nested loop join* алгоритма.

Користимо процену цене спајања $pages(T1) + pages(T1) \times pages(T2)$ и добијамо цену међурезултата:

$$cost(Temp\ 1) = pages(StudijskiProgram) + pages(StudijskiProgram) \times pages(Dosije) = 1 + 1157 = 1158 \quad (7)$$

Укупна цена плана извршавања је $cost(Temp\ 1) + cost(Temp\ 2) + cost(Result) = 1158 + 258 + 1 = 1417$.

(е) За исти упит дати скицу алтернативног стабла извршавања према којем се операције селекције и пројекције врше пре операције спајања. Објаснити да ли постоји предност таквог приступа.

Скица решења — идеја је извршити филтрирање и пројекције пре спајања. Овиме се *олакшава* најскупља операција, тј. операција спајања, јер ће међурезултати над којима она оперише заузимати мање простора. Ово би требало да доведе до мање укупне цене плана извршавања.

¹Песимистична претпоставка да се након сваке операције међурезултат враћа на диск, па је потребно поновно довлачење.