

2. međuispit iz Baza podataka

13. svibnja 2011.

Zadaci 1 do 7 se odnose na relacije opisane na **slici 1**. Na slici **nisu** prikazane sve n-torke koje su sadržane u relacijama. U relacije sa slike 1 se pohranjuju podaci o trkačima (relacija **trkac**) koji sudjeluju na utrkama (relacija **utrka**). Vrijeme (u sekundama) koje pojedini trkač ostvari u pojedinoj utrci evidentirano je u relaciji **rezultat**. Trkačima koji nisu završili utrku ostvareno vrijeme nije evidentirano (ima NULL vrijednost). Države iz kojih dolaze trkači odnosno države u kojima se održavaju utrke evidentirane su u relaciji **drzava**. Osim atributa **rezVrijeme** u relaciji **rezultat** niti jedan atribut ne može poprimiti NULL vrijednost.

Slika 1.

trkac					rezultat		
sifTrkac	prezime	ime	datRodj	sifDrzava	sifTrkac	sifUtrka	rezVrijeme
101	Petrović	Ivan	04.06.1987	1	101	21	24927
102	Jurić	Ante	12.05.1989	1	102	21	25395
103	Varga	Tamas	04.06.1991	3	102	22	35956
104	Molnar	Antal	25.11.1990	3	103	22	37937
105	Korošec	Janez	05.11.1987	2	105	22	NULL
					101	23	28299

utrka						drzava	
sifUtrka	nazUtrka	datUtrka	sifDrzava	brojKm	indeksTezine	sifDrzava	nazDrzava
21	Istra1	25.10.2009	1	75	6	1	Hrvatska
22	Velebit	15.06.2010	1	100	7	2	Slovenija
23	Red Bull	10.07.2010	4	80	3	3	Mađarska
24	Alpine	13.09.2010	2	50	1	4	Austrija

Napišite po jednu SQL naredbu kojom će se obaviti sljedeće:

1. Za svaki ostvareni rezultat u kojem je prosječna brzina trkača bila veća od 2 m/s ispisati šifru i naziv te utrke te šifru i prezime tog trkača. **Zadatak riješiti bez podupita.** (1 bod)
2. Za **svaku** državu ispisati šifru, naziv i broj utrka (nula ili više) koje su se u njoj održavale u drugoj polovici (7.-12. mjesec) 2010. godine. Ispis poredati silazno po broju održanih utrka. **Zadatak riješiti bez podupita.** (2 boda)
3. Ispisati šifru i naziv svake utrke u kojoj nije sudjelovao nijedan natjecatelj iz države s nazivom 'Italija'. (1,5 bod)
4. Za održanu utrku ispisati šifru i naziv, šifru osobe s najboljim vremenom utrke i spomenuto ostvareno najbolje vrijeme. Napomena: više osoba može ostvariti najbolje vrijeme. (2 boda)
5. Svakoju utrci koju je **završilo** više od 50 natjecatelja smanjiti indeks težine za 1, pri čemu indeks težine ne smije poprimiti vrijednost manju od 1. (2 boda)
6. Napisati SQL naredbe koje će kreirati najmanji mogući broj indeksa koji će omogućiti efikasno obavljanje (pomoću B+ stabla) svih dolje navedenih upita. (2 boda)

```
SELECT * FROM utrka WHERE datUtrka = TODAY AND nazUtrka = 'Velebit';
SELECT * FROM utrka WHERE nazUtrka = 'Red Bull';
SELECT * FROM utrka ORDER BY brojKm;
SELECT * FROM utrka ORDER BY sifUtrka, nazUtrka;
SELECT * FROM utrka ORDER BY sifUtrka DESC;
SELECT * FROM utrka ORDER BY brojKm DESC, sifUtrka ASC, nazUtrka ASC;
```

7. Napisati jedan **izraz relacijske algebre** (ne SQL upit) koji odgovara sljedećem:

- Za sve trkače ispisati šifru, prezime i broj utrka (nula ili više) na kojima su sudjelovali.

(1,5 bod)

8. Uz pretpostavku da na relacijskoj shemi $R = PQRSTUV$ vrijede funkcijske zavisnosti iz skupa: $F = \{PQ \rightarrow RT, R \rightarrow QS, Q \rightarrow UV, V \rightarrow U\}$, ispitajte je li **skup atributa PQ ključ** relacijske sheme? U svakom koraku obavezno navesti koji aksiom ili pravilo se koristi. **(2,5 boda)**
9. U bazi podataka evidentiraju se utrke Formule 1 koje se održavaju tijekom više godina. Relacijska shema FORMULA sastoji se od sljedećih atributa: **(3 boda)**
- $sifVN$ – šifra Velike nagrade
 - $nazVN$ – naziv Velike nagrade (npr. Velika nagrada Australije)
 - $sifVozac$ – šifra vozača
 - $imeVozac$ – ime vozača
 - $prezVozac$ – prezime vozača
 - $sifMomcad$ – šifra momčadi za koju vozač nastupa u utrci
 - $nazMomcad$ – naziv momčadi za koju vozač nastupa u utrci
 - $datUtrka$ – datum održavanja utrke
 - $vrijeme$ – rezultat vozača u utrci (vrijeme od označenog starta do ulaska vozača u cilj)

Vrijede sljedeća pravila:

- istoga datuma održava se samo jedna utrka, a svaka se utrka vozi za jednu Veliku nagradu (npr. 14.3.2010. se održava Velika nagrada Bahreina, 28.3.2010. se održava Velika nagrada Australije,..., 27.3.2011. se održava Velika nagrada Australije ...)
- svaki vozač vozi utrku u automobilu neke momčadi i može mijenjati momčadi tijekom karijere
- moguće je da dvojica vozača u nekoj utrci ostvare posve jednako vrijeme

Određiti ključ relacijske sheme FORMULA tako da ona bude u 1NF, a zatim postupno normalizirati relacijsku shemu na 2NF i 3NF.

10. Zadano je B^+ stablo reda 25 u koje je zapisano 190 000 zapisa. Koliko je UI operacija potrebno obaviti prilikom traženja jednog zapisa u **najlošijem** slučaju (uračunati i jednu UI operaciju koja je potrebna za dohvat bloka s podacima)? Objasniti kako ste došli do rezultata. **(2 boda)**
11. Napisati SQL naredbe koje će kreirati relacije **klub** i **sportas** prema relacijskim shemama $KLUB = \{sifKlub, nazivKlub\}$ i $SPORTAS = \{sifSportas, sifKlub, imeSportas, prezimeSportas, spol\}$. Smisleno odaberite tipove podataka. Prilikom kreiranja relacija osigurati da:
- u relaciji **sportas** atribut **sifSportas** bude primarni ključ
 - ime i prezime sportaša ne mogu poprimiti NULL vrijednost
 - vrijednost atributa spol može biti samo Z ili M
 - šifra kluba u relaciji **sportas** poprima samo vrijednosti atributa **sifKlub** u relaciji **klub**
 - naziv kluba ne može poprimiti NULL vrijednost
 - dva kluba ne mogu imati isti naziv
 - prilikom brisanja zapisa iz relacije **klub** budu obrisani i svi zapisi o sportašima (u relaciji **sportas**) koji se nalaze u obrisanom klubu
- (3,5 boda)**
12. Koristeći **relacijsku algebru** napišite uvjet koji mora biti zadovoljen da bi se dekompozicija relacije $r(R)$ na relacije $r_1(R_1), r_2(R_2), \dots, r_n(R_n)$ obavila bez gubitka informacija. **(2 boda)**

1. Za svaki ostvareni rezultat u kojem je prosječna brzina trkača bila veća od 2 m/s ispisati šifru i naziv te utrke te šifru i prezime tog trkača. **Zadatak riješiti bez podupita.** (1 bod)

```
SELECT rezultat.sifutrka, nazutrka, rezultat.siftrkac, prezime
FROM trkac JOIN rezultat ON trkac.siftrkac=rezultat.siftrkac
      JOIN utrka ON rezultat.sifutrka=utrka.sifutrka
WHERE (brojKm*1000)/rezVrijeme>2;
```

2. Za **svaku** državu ispisati šifru i naziv države te broj utrka (nula ili više) koje su se u njoj održale u drugoj polovici 2010. godine. Rezultat poredati silazno po broju održanih utrka. **Zadatak riješiti bez podupita.** (2 boda)

```
SELECT drzava.sifdrzava, nazdrzava, COUNT(utrka.sifutrka) AS broj_utrka
FROM drzava
LEFT JOIN utrka ON drzava.sifdrzava = utrka.sifdrzava
      AND month(datutrka)>6
      AND year(datutrka)=2010
GROUP BY drzava.sifdrzava , nazdrzava
ORDER BY broj_utrka DESC;
```

3. Ispisati šifru i naziv svake utrke u kojoj nije sudjelovao nijedan natjecatelj iz države s nazivom 'Italija'. (1.5 boda)

```
SELECT sifutrka, nazutrka
FROM utrka
WHERE sifUtrka NOT IN (
      SELECT sifutrka
      FROM rezultat JOIN trkac ON rezultat.siftrkac=trkac.siftrkac
            JOIN drzava ON drzava.sifdrzava=trkac.sifdrzava
      WHERE nazdrzava='Italija');
```

ili

```
SELECT sifutrka, nazutrka
FROM utrka
WHERE NOT EXISTS (
      SELECT *
      FROM rezultat JOIN trkac ON rezultat.siftrkac=trkac.siftrkac
            JOIN drzava ON drzava.sifdrzava=trkac.sifdrzava
      WHERE nazdrzava='Italija'
      AND rezultat.sifutrka=utrka.sifutrka);
```

ili

```
SELECT utrka.sifutrka, nazutrka
FROM rezultat JOIN trkac ON rezultat.siftrkac=trkac.siftrkac
      JOIN drzava ON drzava.sifdrzava=trkac.sifdrzava
      RIGHT JOIN utrka ON rezultat.sifutrka=utrka.sifutrka
      AND nazdrzava='Italija'
GROUP BY utrka.sifutrka, nazutrka
HAVING COUNT(rezultat.sifUtrka)=0;
```

4. Za održanu utrku ispisati šifru i naziv, šifru osobe (jedne ili više njih) s najboljim vremenom utrke i spomenuto ostvareno najbolje vrijeme. (2 boda)

```
SELECT utrka.sifutrka, utrka.nazutrka, siftrkac, rezvrijeme
FROM rezultat r1 JOIN utrka ON r1.sifutrka=utrka.sifutrka
WHERE r1.rezvrijeme=(
      SELECT MAX(r2.rezvrijeme)
      FROM rezultat r2
      WHERE r1.sifutrka=r2.sifutrka);
```

ili

```
SELECT utrka.sifutrka, utrka.nazutrka, siftrkac, rezvrijeme
FROM rezultat r1 JOIN utrka ON r1.sifutrka=utrka.sifutrka
WHERE (
      SELECT COUNT(*)
      FROM rezultat r2
      WHERE r1.sifutrka=r2.sifutrka AND r2.rezvrijeme>r1.rezvrijeme
)=0;
```

```
SELECT r1.sifUtrka, nazutrka, r1.sifTrkac, r1.rezvrijeme
FROM utrka JOIN rezultat r1 ON utrka.sifutrka=r1.sifutrka
LEFT JOIN rezultat r2 ON r1.sifutrka=r2.sifutrka
AND r2.rezvrijeme>r1.rezvrijeme
GROUP BY r1.sifUtrka, r1.sifTrkac, nazutrka, r1.rezvrijeme
HAVING COUNT(r2.sifUtrka)=0;
```

5. Svakoj utrci koju je **završilo** više od 50 natjecatelja smanjiti indeks težine za 1, pri čemu indeks težine ne smije poprimiti vrijednost manju od 1. (2 boda)

```
UPDATE utrka
SET indeksTezine = indeksTezine - 1
WHERE indeksTezine > 1
AND sifUtrka IN
(
SELECT sifUtrka
FROM rezultat
WHERE rezVrijeme IS NOT NULL
GROUP BY sifUtrka
HAVING COUNT(*)>50);
```

6. Napisati SQL naredbe koje će kreirati najmanji mogući broj indeksa koji će omogućiti efikasno obavljanje (pomoću B+ stabla) svih dolje navedenih upita. (2 boda)

```
CREATE INDEX i1 ON utrka (brojKm, sifUtrka DESC, nazUtrka DESC);
CREATE INDEX i2 ON utrka (nazUtrka, datUtrka);
CREATE INDEX i3 ON utrka (sifUtrka, nazUtrka);
```

7. Za sve trkače ispisati šifru, prezime i broj utrka na kojima su sudjelovali. (1,5 bod)

$P_{\text{REZULTAT(sifra, prezime, brojUtrka)}}(\text{sifTrkac, prezime, } G_{\text{COUNT(sifUtrka)}(\text{trkac}^* > \text{< rezultat})}$

8. Uz pretpostavku da na relacijskoj shemi $R = \mathbf{PQRSTUV}$ vrijede funkcijske zavisnosti iz skupa: $F = \{\mathbf{PQ} \rightarrow \mathbf{RT}, \mathbf{R} \rightarrow \mathbf{QS}, \mathbf{Q} \rightarrow \mathbf{UV}, \mathbf{V} \rightarrow \mathbf{U}\}$, ispitati je li **PQ ključ** relacijske sheme? U svakom koraku obavezno navedite koji aksiom ili pravilo koristite. (2.5 boda)

Provjeriti može li **PQ** biti **ključ** relacijske sheme?

$\mathbf{PQ} \rightarrow \mathbf{PQ}$ (refleksivnost)
 $\mathbf{PQ} \rightarrow \mathbf{PQ}$ i $\mathbf{PQ} \rightarrow \mathbf{RT} \Rightarrow \mathbf{PQ} \rightarrow \mathbf{PQRT}$ (akumulacija)
 $\mathbf{PQ} \rightarrow \mathbf{PQRT}$ i $\mathbf{Q} \rightarrow \mathbf{UV} \Rightarrow \mathbf{PQ} \rightarrow \mathbf{PQRTUV}$ (akumulacija)
 $\mathbf{PQ} \rightarrow \mathbf{PQRTUV}$ i $\mathbf{R} \rightarrow \mathbf{QS} \Rightarrow \mathbf{PQ} \rightarrow \mathbf{PQRSTUV}$ (akumulacija)

PQ funkcijski određuje cijelu relacijsku shemu; potrebno je provjeriti određuju li je i podskupovi **P** odnosno **Q**.

$\mathbf{P} \rightarrow \mathbf{P}$ (refleksivnost)

Nije moguće izraz dalje dopuniti s desne strane da uključi **QRSTUV** pa funkcijska zavisnost $\mathbf{P} \rightarrow \mathbf{QRSTUV}$ ne vrijedi tj. **P** nije ključ.

$\mathbf{Q} \rightarrow \mathbf{Q}$ (refleksivnost)
 $\mathbf{Q} \rightarrow \mathbf{Q}$ i $\mathbf{Q} \rightarrow \mathbf{UV} \Rightarrow \mathbf{Q} \rightarrow \mathbf{QUV}$ (akumulacija)

Nije moguće izraz dalje dopuniti s desne strane da uključi **PRSTUV** pa funkcijska zavisnost $\mathbf{Q} \rightarrow \mathbf{PRSTUV}$ ne vrijedi tj. **Q** nije ključ.

9. U bazi podataka evidentiraju se utrke Formule 1 koje se održavaju tijekom više godina. Relacijska shema FORMULA sastoji se od sljedećih atributa: **(3 boda)**

sifVN	– šifra Velike nagrade
nazVN	– naziv Velike nagrade (npr. Velika nagrada Australije)
sifVozac	– šifra vozača
imeVozac	– ime vozača
prezVozac	– prezime vozača
sifMomcad	– šifra momčadi za koju vozač nastupa u utrci
nazMomcad	– naziv momčadi za koju vozač nastupa u utrci
datUtrka	– datum održavanja utrke
vrijeme	– rezultat vozača u utrci (vrijeme od označenog starta do ulaska vozača u cilj)

Vrijede sljedeća pravila:

- istoga datuma održava se samo jedno natjecanje
- svaki vozač vozi utrku u automobilu neke momčadi i može mijenjati momčadi tijekom karijere
- moguće je da dvojica vozača u nekoj utrci ostvare posve jednako vrijeme

Odaberite ključ relacijske sheme FORMULA tako da ona bude u 1NF a zatim postupno normalizirajte relacijsku shemu na 2NF i 3NF.

1.NF

KLJUČ: (sifVozac, datUtrka)

2.NF

VOZAC (sifVozac, imeVozac, prezVozac)

TERMIN (datUtrka, sifVN, nazVN)

UTRKA (sifVozac, datUtrka, sifMomcad, nazMomcad, vrijeme)

3.NF

VOZAC1 (sifVozac, imeVozac, prezVozac)

TERMIN1 (datUtrka, sifVN)

NATJECANJE(sifNatj, nazVNj)

UTRKA1 (sifVozac, datUtrka, sifMomcad, vrijeme)

MOMCAD (sifMomcad, nazMomcad)

10. Zadano je B+ stablo reda 25 u koje je zapisano 190 000 zapisa. Koliko će biti potrebno obaviti UI operacija prilikom traženja jednog zapisa u najgorem slučaju (uračunati i jednu UI operaciju koja je potrebna za dohvrat bloka s podacima)? Objasniti kako ste došli do rezultata. **(2 boda)**

Najgori slučaj:

Broj kazaljki u internim čvorovima minimalno 13 (osim u korijenu 2), u listovima minimalno 12.

Za 5 razina: $2 \cdot 13 \cdot 13 \cdot 13 \cdot 12 = 52\,728$

Za 6 razina: $2 \cdot 13 \cdot 13 \cdot 13 \cdot 13 \cdot 12 = 685\,464$

Sa 6 razina stablo mora obuhvatiti BAREM 685 464 zapisa, što je previše; s 5 razina stablo mora obuhvatiti BAREM 52 728 zapisa, ali može imati i više od toga, tj. potrebnih 190 000.

Stablo treba imati 5 razina; ukupno $5+1 = 6$ UI operacija

11. Napisati SQL naredbe koje će kreirati relacije **klub** i **sportas** prema relacijskim shemama KLUB {sifKlub, nazivKlub} i SPORTAS {sifSportas, sifKlub, imeSportas, prezimeSportas, spol}. Smisleno odaberite tipove podataka. Prilikom kreiranja relacija osigurati da:

- u relaciji **sportas** atribut **sifSportas** bude primarni ključ
 - ime i prezime sportaša ne mogu poprimiti NULL vrijednost
 - vrijednost atributa spol može biti samo Z ili M
 - šifra kluba u relaciji **sportas** poprima samo vrijednosti atributa **sifKlub** u relaciji **klub**
 - naziv kluba ne može poprimiti NULL vrijednost
 - dva kluba ne mogu imati isti naziv
 - prilikom brisanja zapisa iz relacije **klub** budu obrisani i svi zapisi o sportašima (u relaciji **sportas**) koji se nalaze u obrisanom klubu
- (3.5 boda)**

```
CREATE TABLE klub(  
    sifKlub INTEGER PRIMARY KEY  
    , nazivKlub NCHAR(25) NOT NULL UNIQUE  
);
```

```
CREATE TABLE sportas(  
    sifKlub INTEGER NOT NULL REFERENCES klub(sifKlub) ON DELETE CASCADE  
    , imeSportas NCHAR(25) NOT NULL  
    , prezimeSportas NCHAR (25) NOT NULL  
    , spol CHAR(1) CHECK (spol IN ('Z', 'M')) -  
    , mbrSportas CHAR(11) PRIMARY KEY);
```

12. . Koristeći **relacijsku algebru** napišite uvjet koji mora biti zadovoljen da bi se dekompozicija relacije **r(R)** na relacije **r₁(R₁)**, **r₂(R₂)**, ..., **r_n(R_n)** obavila bez gubitka informacija. **(2 boda)**

- Relacija **r(R)** se dekomponira na relacije **r₁(R₁)**, **r₂(R₂)**, ..., **r_n(R_n)** bez gubitaka informacija (lossless decomposition) ako vrijedi:

$$r_1(R_1) \bowtie r_2(R_2) \bowtie \dots \bowtie r_n(R_n) = r(R)$$

odnosno

$$\pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \dots \pi_{R_n}(r) = r(R)$$