

# TÖL303G Gagnasafnsfræði

Upptökupróf

Kennarar: Hjálmtýr Hafsteinsson og  
Jóhann Haukur Gunnarsson

17. Janúar, 2009

kl. 9<sup>00</sup> – 12<sup>00</sup>

*Öll dæmin hafa sama vægi. Aðeins þarf að leysa 6 dæmi af 7. Sex bestu dæmin gilda. Öll skrifleg hjálpargögn og reiknivél leyfileg.*

- Athugið að svar án rökstuðnings er einskis virði. Rökstyðjið því öll svör og munið að það er óþarfi að skrifa upp skilgreiningar sem eru í bókinni.

1. Þið eigið að hanna gagnasafn fyrir hótél. Hótelið hefur mörg herbergi. Hvert herbergi hefur númer, staðsetningu, mögulegan gestafjölda og hvort það sé reyklaust. Viðskiptavinir geta pantað herbergi fyrir ákveðið tímabil. Einstökum herbergjum er ekki úthlutað til viðskiptavina við pöntun, heldur þegar þeir koma. Viðskiptavinirnir gista síðan í tilteknu herbergi í ákveðna daga. Greiðsla fyrir dvölinu getur farið fram að hluta fyrirfram, eða við komu, en síðan er gert upp í lokin. Halda þarf utanum hversu mikið hefur verið greitt inná dvölinu og hvenær.

Teiknið þessar upplýsingar upp í einindavenslarit (e. entity-relationship diagram) með öllum þeim skorðum sem fram koma í lýsingunni. Fyrir hver einindavensl sem koma fyrir í ritinu réttlætið eiginleika þeirra með vísun í texta dæmisins eða forsendu sem þið gefið ykkur.

## 2. Gefnar eru töflurnar

```
Movies( mid, mtitle, year, type )
Ratings( mid, cid, rating )
Customers( cid, cname, age )
```

Töflurnar geyma upplýsingar um kvikmyndir (**Movies**), viðskiptavinir (**Customers**) og einkunnir sem viðskiptavinir gefa myndunum (**Ratings**). Hver kvikmynd er af tiltekinni gerð (**type**), sem gæti verið "Action", "Drama", "Comedy", o.s.frv. Viðskiptavinir geta gefið myndum einkunn (**rating**), sem er heiltala frá 0 til 10.

Þið eigið að búa til SQL skipanir til að svara eftirfarandi fyrirspurnum á þetta gagnasafn:

- Sýnið þær kvikmyndir sem viðskiptavinurinn "John Smith" hefur gefið einkunn í lækkandi röð eftir einkunn og síðan í stafrófsröð innan einkunnar.
- Finnið fjölda þeirra kvikmynda sem enginn hefur gefið einkunn.
- Sýnið þær kvikmyndir sem fengu hæstu meðaleinkunn fyrir hvert ár, af þeim myndum sem fengu fleiri en 5 einkunnir. Sýnið ár, nafn kvikmynda, fjölda einkunna og meðaleinkunn.
- Sýnið þær kvikmyndir af gerðinni "Action" sem fá hæstu meðaleinkunn frá viðskiptavinum á aldursbilinu 15 til 25 ára.

## 3. Gefnar eru töflurnar

```
songs( title, performer, length, albumname, type)
albums( aname, artist, pub_year)
```

- Sýnið hvernig eftirfarandi skorða (e. constraint) væri skilgreind: Lag má ekki vera lengra en 10 mínútur, nema ef tegundin (**type**) er '**classical**'.
- Útskýrið hvernig hægt væri að tryggja að öll lögin sem eru á sömu hljómplötu (e. album) séu af sömu tegund. Miðið við að notað sé PostgreSQL. Þið þurfið ekki að sýna allan kóða, en gefa þarf ítarlega lýsingu.
- Búið til sýnina (e. view) **rockalbums(name, artist, length, year)**, sem inniheldur aðeins hljómplötur af tegundinni '**rock**' og sviðið **length** er samanlögð lengd allra laganna á plötunni.

4. Lát  $R(A, B, C, D)$  vera vensl með fallákveðunum  $A \rightarrow B, A \rightarrow C$ .

a) Finnið alla lykla venslanna  $R$ .

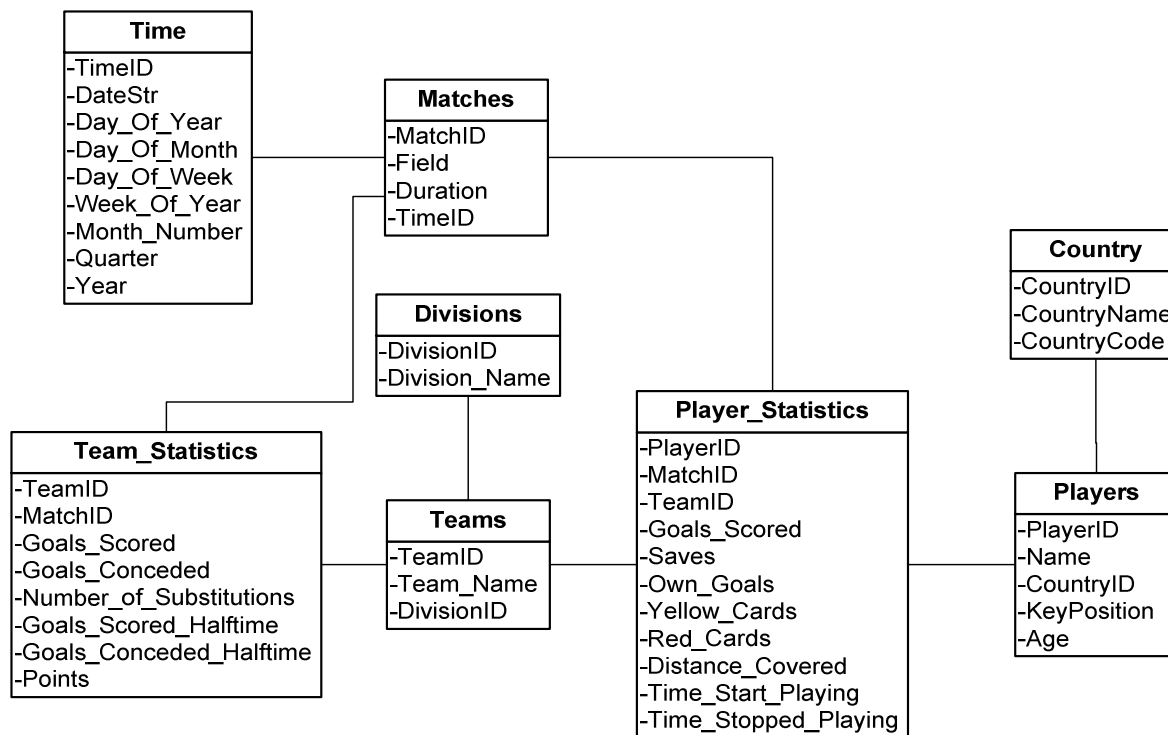
b) Venslin  $R$  eru ekki á BCNF, útskýrið hvers vegna.

c) Hægt er að skrifa fallákveðurnar tvær sem eina fallákveðu  $A \rightarrow BC$ . Skiptir máli hvort venslin  $R$  eru brotin upp samkvæmt  $A \rightarrow B$  eða samkvæmt  $A \rightarrow BC$ , til að koma þeim á BCNF? Útskýrið.

5. Gefnar eru töflurnar **sailors**(*sid*, *sname*, *rating*, *age*) og **reserves**(*sid*, *bid*, *day*). Skrifið PL/pgSQL fallið **comp\_expr**(*sid*), sem fær sem viðfang auðkenni sjómanns (*sid*) og flokkar hann eftir því hversu oft hann hefur pantað báta á síðustu 5 árum. Ef sjómaðurinn hefur pantað 100 sinnum eða oftar skilar fallið strengnum **'experienced'**, ef hann hefur pantað sjaldnar en 10 sinnum skilar það **'inexperienced'**, en annars er skilað strengnum **'average'**.

PostgreSQL hefur breytuna **CURRENT\_DATE**, sem inniheldur alltaf núverandi dagsetningu á **date**-formi. Hægt er að nota venjulegan frádrátt á **date** gildi og útkoman er fjöldi daga sem er á milli dagsetninganna.

6. Eftirfarandi mynd sýnir skema á einföldu vöruhúsi fyrir ýmsa tölfræði úr Enska boltanum þar sem upplýsingar um alla leiki í öllum deildum Enska boltans eru geymdar fyrir síðastliðin ár.



- Hvaða töflur eru gagnatöflur (e. fact table) og hvað má kalla svona skema?
- Augljóst er að engin hugsun er í hönnuninni varðandi hægbreytilegar víddir (e. slowly changing dimensions). Sýnið eitt dæmi um það hvernig hönnunin bregst eftir því sem tíminn líður og hvernig hægt væri að betrubæta hönnunina í því tilviki.
- Gerið SQL fyrirspurn sem skilar stigatöflu í lok tímabils í réttri röð fyrir árið 2005 og deild X.
- Gerið SQL fyrirspurn sem tilgreinir fyrir lið X hvaða leikmaður hefur skorað oftast þegar liðið hefur unnið leiki.

7. Gefnar eru eftirfarandi tvær hreyfingar,  $T_1$  og  $T_2$

$T_1$	$T_2$
Read(A)	Read(B)
$A = A * 2$	Read(A)
Write(A)	$A = A + B$
Read(B)	Write(A)
$B = B - 1$	
Write(B)	

Gerið ráð fyrir því að í upphafi hafi  $A$  gildið 10 og  $B$  gildið 20.

- Ef hreyfingarnar eru framkvæmdar raðbundið (e. serial) hver eru þá lokagildin á  $A$  og  $B$ ? Sýnið báða möguleikana, þ.e. bæði  $T_1, T_2$  og  $T_2, T_1$ .
- Sýnið aðra fléttun skipana  $T_1$  og  $T_2$  sem gefur önnur lokagildi á  $A$  eða  $B$ .
- Útskýrið hvernig 2ja fasa læsing (2PL) kemur í veg fyrir vandamálið í  $b$ )-lið.
- Er einhver möguleiki á að sjálfhelda komi upp í  $c$ )-lið?