

TÖL303G Gagnasafnsfræði

Lokapróf

17. desember, 2008

Kennarar: Hjálmtýr Hafsteinsson og
Jóhann Haukur Gunnarsson

kl. 13³⁰ – 16³⁰

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 ákveðna þætti í rekstri sjúkrahúss. Sjúkrahúsið hefur marga skráða lækna. Fyrir hvern lækni er skráð auðkenni hans og sérsvið. Sjúklingar eru við komuna á spítalann skoðaðir af einhverjum tilteknum lækni. Eftir innlögn er hver sjúklingur meðhöndlaður af einum eða fleiri læknum. Þegar sjúklingur er meðhöndlaður af lækni þá er meðferðin skráð með dagsetningu, lýsingu og niðurstöðum. Sem hluta af tiltekinni meðferð getur læknir pantað rannsóknir á sjúklingi. Fyrir hverja rannsókn þarf að skrá auðkenni rannsóknarinnar, tegund hennar (blóðrannsókn, þvagrannsókn, o.s.frv), dagsetningu, starfsmann sem framkvæmdi rannsóknina og útkomu hennar. Auk þess þarf að skrá persónuupplýsingar um hvern sjúkling.

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

```
Emp( eid, ename, salary, works_for_did )
Project( pid, descr, budget, hosted_by_did )
Takes_part( eid, pid, hours_per_month )
Dept( did, dname, manager )
```

Töflurnar geyma upplýsingar um starfsmenn (**Emp**), verkefni (**Project**) og deildir (**Dept**) í fyrirtæki nokkru. Það er skráð hvaða starfsmenn vinna í hvaða verkefni og hversu marga tíma á mánuði (**Takes_part**). Einnig er skráð í hvaða deild hver starfsmaður er (**works_for_did** sviðið í töflunni **Emp**) og hvaða hvaða deild hvert verkefni tilheyrir (**hosted_by_did** sviðið í töflunni **Project**).

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

- a) Sýnið það verkefni sem hefur mesta fjármagnið (**budget**) af þeim verkefnum sem tilheyra "Sales" deildinni.
- b) Sýnið þau verkefni sem hafa enga starfsmenn skráða á sig.
- c) Sýnið þá starfsmenn sem aðeins vinna við verkefni sem tilheyra deildinni sem þeir sjálfir eru í.
- d) Sýnið öll þau verkefni sem starfsmaðurinn "John" vinnur við, í röð eftir umfangi, þ.e. heildarmagni tíma sem settir eru í verkefnið (ekki aðeins tíma "John"). Sýnið þó aðeins þau verkefni þar sem umfangið er meira en 100 tímar.

3. Í venjulegri tengingu (e. join) tafla skiptir röð taflanna ekki máli, svo að venjuleg tenging er bæði víxlin (e. commutative) og tengin (e. associative). *Vinstri ytri tenging* (e. left outer join) er augljóslega ekki víxlin, þannig að (**A natural left outer join B**) skilar ekki sömu línunum og (**B natural left outer join A**). Í þessu dæmi eigið þið að sýna að aðgerðin er heldur ekki tengin.

Sýnið því dæmi (og **útskýrið það**) um þrjár töflur A, B og C, þannig að SQL-skipunin

```
select *
  from (A natural left outer join B) natural left outer join C
```

gefur ekki sömu niðurstöðu og skipunin

```
select *
  from A natural left outer join (B natural left outer join C)
```

Vísbending: Gerið til hægðarauka ráð fyrir því að allar töflurnar þrjár hafi einn dálk sem er skilgreindur alveg eins. Töflurnar þurfa ekki að hafa nema eina línu hver.

4. Lát $R(A, B, C, D, E)$ vera vensl með fallákveðunum $AB \rightarrow C$, $AC \rightarrow B$, $AD \rightarrow E$, $B \rightarrow D$, $BC \rightarrow A$.

- a) Finnið alla lykla venslanna R .
- b) Ef venslin R eru ekki á BCNF bendið þá á hvers vegna ekki og brjótið venslin upp þannig að þau uppfylli BCNF.
- c) Ef venslin R væru brotin upp í venslin $R1(A, B, C)$ og $R2(A, B, D, E)$ væri það uppbrót þá taplaust (e. lossless)? Útskýrið.

5. Gefin er taflan `emp(eid, ename, salary)`. Skrifðu kveikju (e. trigger) fyrir PostgreSQL sem vinnur þannig að ef einhver reynir að lækka launin (þ.e. `salary`) í fyrirtækinu með UPDATE-skipun, þá gerist sú breyting ekki hjá ykkur (`eid = 234`) eða hjá besta vini ykkar (`eid = 543`). Úskýrið einnig hvernig hægt væri að breyta laununum hjá þessum tveimur starfsmönnum án þess að taka kveikjuna úr sambandi.

6. a) Hvaða þrjú megin þættir eru í ETL ferli? Lýsið hverjum þætti ferlisins fyrir sig, þ.e. upphafi og endi hvers þáttar, og tilgangi hans.
- b) Hægbreytilegar víddir (e. slowly changing dimensions) eru ákveðið vandamál í vöruhúsum gagna. Hvaða þrjár leiðir eru í boði til að takast á við þetta vandamál og hver þeirra er skást með tilliti til varðveislu gagna.
- c) Teiknið stjörnuskemu (e. star schema) fyrir útlánagrunn (e. exposure database), þar sem vitað er á hverjum degi hversu mikið hver einstaklingur skuldar (eftirstöðvar), upphæð vanskila, fjölda daga sem lán hefur verið í vanskilum og verðmat ábyrgða á bakvið lánin ef einhver eru. Lánin tilheyra mismunandi útibúum og eru af mismunandi tegundum og sum þeirra geta verið fryst. Sýnið þetta sem ER-rit eða UML-klasarit þar sem dálkar hvernar töflu sjást og vensl þeirra á milli.

7. Segjum að hreyfing T þurfi aðeins að skrifa í gagnasafnshlutinn A í lok vinnslu sinnar, en þurfi að lesa A strax í upphafi. Ef 2ja fasa læsing (2PL) er notuð þá þyrfti T að fá X -lás á A strax í upphafi og þar með að útiloka aðrar hreyfingar frá því að lesa A þar til T hefur lokið keyrslu. Betra væri ef hreyfingin mætti *uppfæra* (e. *upgrade*) lásinn frá S -lás uppí X -lás. Þá gætu aðrar hreyfingar líka lesið A á meðan T hefur S -lás á honum, en síðan myndi T biðja um uppfærslu uppí X -lás (og þyrfti þá að bíða á meðan aðrar hreyfingar hafa S -lás á A).

- a) Sýnið að ef þessi uppfærsla er leyfð þá er möguleiki á sjálfheldu (e. deadlock) milli tveggja hreyfinga. Sýnið dæmi um slíkt.
- b) Til að forðast möguleikan á sjálfheldu væri hægt að hafa þriðju lásagerðina, U -lás (uppfærslulás), sem væri þannig að einungis væri hægt að uppfæra hluti með U -lás uppí X -lás. Hreyfing getur fengið U -lás á hlut þó aðrar hreyfingar hafi S -lása á honum, en eftir að kominn er U -lás á hlut þá er ekki hægt að fá neina aðra lása á hann (hvorki S -, U - né X -lása). Úskýrið hvernig þetta kemur í veg fyrir sjálfhelduna úr a)-lið. Sýnið í töflu öll tilfelli af því hvernig U -lásinn passar við hinar lásagerðirnar og rökstyðjið hvert tilfelli.