

TÖL303G Gagnasafnsfræði

Lokapróf

**Kennarar: Hjálmtýr Hafsteinsson og
Jóhann Haukur Gunnarsson**

18. desember, 2009

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 rekstur bifreiðaverkstæðis. Viðskiptavinur bókar tíma fyrir viðgerð á bifreið sinni. Fyrir hvern viðskiptavin þarf að skrá kennitölu, nafn og síma. Fyrir bifreiðarnar þarf að geyma skráningarnúmer, gerð, tegund og árgerð. Hver viðgerð þarf tiltekna varahluti. Halda þarf skrá yfir varahluti, lýsingu á þeim og verð. Hver bifreið getur aðeins komið einu sinni inni til viðgerðar á einum degi. Viðskiptavinur getur komið með margar mismunandi bifreiðar til viðgerðar. Auk þess hefur verkstæðið tiltekinn fjölda viðgerðarmanna sem hver sér um eina viðgerð á einni bifreið.

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. Sýnið auk þess hvaða **töflur** yrðu í þessu gagnasafni. Þið þurfið ekki að sýna fullar skilgreiningar á töflunum, heldur bara tengsl þeirra við einindavenslaritið.

2. Gefnar eru töflurnar

```
Has_seen( fan, movie )
Plays_in( actor, movie )
Likes( fan, actor )
```

Töflurnar geyma upplýsingar um tengsl á milli kvikmynda (**movie**), leikara (**actor**) og áhorfenda (**fan**). Þar er skráð hvaða kvikmyndir áhorfendur hafa séð (**Has_seen**), hvaða leikarar leika í hvaða kvikmyndum (**Plays_in**), og hvernir eru uppáhaldsleikarar einstakra áhorfenda (**Likes**).

Sýnið SQL skipanir til að svara eftirfarandi fyrirspurnum á þetta gagnasafn:

- Sýnið þær kvikmyndir sem aðdáendur leikarans Bruce Willis hafa séð.
- Sýnið fjölda þeirra áhorfenda sem hafa séð þær kvikmyndir sem bæði Sean Connery og Harrison Ford léku í.
- Fyrir hvern leikara sýnið fjölda áhorfenda sem hafa séð kvikmynd sem þeir hafa leikið í.
- Sýnið þá áhorfendur sem hafa séð kvikmyndir sem innihalda enga leikara sem þeim líkar.

3. Gefin er taflan **students** með upplýsingum um nemendur:

sid	name	credits	year
5004	Sigurður Jónsson	45	2
5015	Anna Halldórsdóttir	NULL	1
5021	Gunnar Bjarnason	80	3
5023	Axel Axelsson	NULL	2
5037	Jenný Jónsdóttir	120	3
5044	Birna Bjarnadóttir	90	2

Finna á alla þá nemendur sem hafa færri einingar en allir nemendur sem eru á þriðja ári. Það hafa komið uppástungur um tvær ólíkar SQL fyrirspurnir til að leysa málið:

```
select name from students s1
  where not exists ( select * from students s2
                    where s2.year = 3 and s2.credits <= s1.credits );
```

```
select name from students s1
  where s1.credits < all ( select s2.credits from students s2
                        where s2.year = 3 );
```

- Útskýrið í örfáum orðum hvernig fyrirspurnirnar tvær vinna.
- Miðað við gögnin í töflunni að ofan hvert er úttakið úr hvorri fyrirspurn fyrir sig?
- Útskýrið niðurstöðuna í b)-lið.

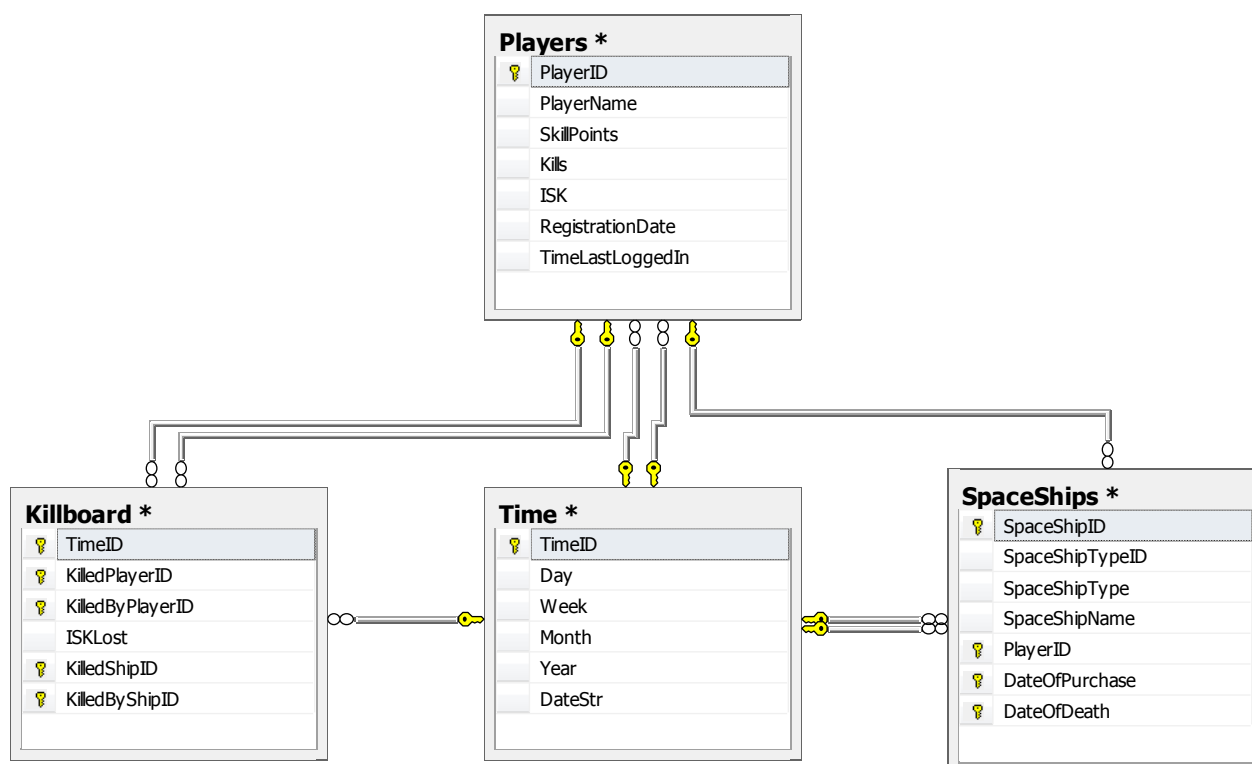
4. Lát $R(A, B, C, D, E, F)$ vera vensl með fallákveðunum $AC \rightarrow B, BD \rightarrow F, F \rightarrow CE$.
- a) Finnið alla lykla venslanna R .
 - b) Finnið $(ABF)^+$, þ.e. lokun sviðanna (e. attribute closure) $\{A, B, F\}$ með tilliti til fallákveðana að ofan.
 - c) Ef venslin R eru ekki á BCNF bendið þá á hvers vegna ekki og brjótið venslin upp þannig að þau uppfylli BCNF.

5. Gefnar eru töflurnar **registration(sid, course, semester, compl)** og **prereqs(course, prereq)**. Í töflunni **registration** eru námskeiðsskráningar nemenda (**sid**) í HÍ. Sviðið **compl** er af taginu **boolean** og segir til um það hvort nemandinn hafi lokið námskeiðinu. Í upphafi er gildi þess **false**. Í töflunni **prereqs** eru geymdar allar forkröfur allra námskeiða. Taflan inniheldur eina línu fyrir hverja forkröfu sem námskeið hefur.

Þið eigið að skrifa *kveikju* (e. trigger) sem passar að öllum forkröfum sé fullnægt við námskeiðsskráningu (eða breytingu á skráningu). Nemandi þarf að hafa lokið öllum þeim námskeiðum sem eru skráð sem forkröfur fyrir námskeiðið sem á að skrá hann í. Kveikjan á að gefa villuskilaboð með númerum þeirra forkröfunámskeiða sem vantar.

6. a) Eitt helsta öryggisvandamálið í gagnasafnskerfum er SQL innstunga (e. SQL injection). Í Java forritum sem nota JDBC er forriturum ráðlagt að nota **PreparedStatement** til að framkvæma fyrirspurnir til þess að koma í veg fyrir þessa öryggishættu. Útskýrið hvernig **PreparedStatement** kemur í veg fyrir SQL innstungu og lýsið því hvaða aðra kosti **PreparedStatement** hefur fram yfir **Statement** viðmótið í JDBC.

b) Eftirfarandi skema er hvorki staðlað (e. normalized) né raunverulegt vöruhús gagna. Tilgangur þessarar skemu er þó sá að geta séð eftir degi hverjir drápu hvern í geimbardaga, hversu mikið tapaðast og eignir viðkomandi í gjaldmiðlinum ISK.



i) Breytið þessu skema í stjörnuskema. Það má líta á alla dálkana sem stök í sameiginlegu mengi og endurraða þeim innan taflna og búa til nýjar töflur, svo lengi sem allir dálkarnir í menginu séu enn fyrir hendi. Einnig má bæta við nýjum dálkum. Í nýja skemanu á að vera hægt að summera upp eftir PlayerID, TimeID, SpaceShipID o.s.frv. þannig að niðurstaðan sé rétt miðað við gefnar forsendur.

ii) Skriðið fyrirspurn á nýja skemað sem skilar topp 10 spilurunum sem valda mestu tjóni per mánuð frá upphafi. Útkoman ætti að vera á eftirfarandi formi:

PlayerName	TotalValueDestroyed	Month	Year

7. Verkröð hreyfinga (e. transaction schedule) er sögð vera *laus við flæðandi vinnsluslit* (e. free of cascading rollbacks) ef sérhver hreyfing les aðeins gildi sem hafa verið staðfest (e. committed).

- a) Er eftirfarandi verkröð hreyfinganna T_1 og T_2 laus við flæðandi vinnsluslit? Hér þýðir $R_2(A)$ að hreyfing T_2 er að lesa hlutinn A .

$R_1(A), W_2(B), R_2(A), W_2(A), R_1(B), \text{Commit}_2, W_1(B), \text{Commit}_1$

- b) Getur flæðandi vinnsluslit átt sér stað þegar notuð er stíf 2ja fasa læsing (e. strict 2PL)? Útskýrið.
- c) Helsti *kosturinn* við verkröð sem er laus við flæðandi vinnsluslit er að það er ekki lengur vandamál að þegar hætt er við eina hreyfingu þá verði að hætta við margar aðrar hreyfingar í framhaldi af því. Hver er hins vegar helsti *gallinn* við slíka verkröð?
- d) Eru allar verkraðir sem eru lausar við flæðandi vinnsluslit líka raðbindanlegar (e. serializable)? Rökstyðjið eða sýnið mótdæmi.