

CS–A1150 Tietokannat

Tentti 20.5.2020

- (8 p) Tee seuraavan kuvauksen pohjalta UML-kaavio kuljetusyrityksen tietokantaa varten. Käytä kurssilla opetettua notaatiota ja merkitse avainattribuutit kurssilla opetetulla tavalla.
Kuljetusyrityksellä on joukko toimipisteitä, joihin asiakkaat voivat tuoda paketteja kuljetettavaksi ja noutaa niitä. Yritys kuljettaa paketteja toimipisteestä toiseen. Yksi kuljetus on aina yhdestä toimipisteestä toiseen toimipisteeseen, joten sama paketti voi joutua useaan eri kuljetukseen ennen kuin se päätyy haluttuun lopulliseen määräpaikkaan.
Toimipisteistä tiedetään niiden tunnus (yksikäsitteinen), nimi ja osoite. Kuljetukset noudattavat vakioreittejä, joista tiedetään lähtötoimipiste ja tulotoimipiste. Reitti on aina kahden toimipisteen välinen. Reitillä on yksikäsitteinen reittitunnus. Tietokantaan tallennetaan etukäteen tieto siitä, mitä kuljetuksia kullakin reitillä on tulossa. Lisäksi tietokannassa säilytetään historiatietoa jo tapahtuneista kuljetuksista. Kuljetuksista tiedetään niiden reitin lisäksi suunniteltu tai toteutunut lähtöaika (päivämäärä ja kellonaika) sekä käytetyn ajoneuvon tyyppi.
Kuljetettavista paketeista tallennetaan paketin tunnus (yksikäsitteinen), lähettäjän ja vastaanottajan yhteystiedot, tilavuus, paino, peritty maksu, toimipiste, johon paketti on alunperin jätetty kuljetettavaksi sekä paketin lopullinen määränpäätoimipiste. Lisäksi tietokannassa on tieto jokaisesta kuljetuksesta, johon paketti on kuulunut sekä paketin tämänhetkinen tila (esim. kuljetuksessa, odottaa seuraavaa kuljetusta tai odottaa vastaanottajan noutoa).
Paketin lähettäjällä on mahdollisuus valita erilaisia lisäpalveluja paketin yhteyteen (esimerkiksi vakuutus, lisäsäilytysaika määränpäässä, varovainen käsittely tai ylimääräinen ilmoitus vastaanottajalle). Tietokantaan on tallennettu tiedot erilaisista mahdollisista lisäpalveluista (tunnus, kuvaus ja palvelusta perittävä maksu) sekä kullekin paketille tieto siitä, mitä lisäpalveluja tälle paketille on valittu.
 - (2 p) Muuta a-kohdan UML-kaavio relaatiomalliin. Esitä relaatiokaaviot ja alleviivaa niistä avainattribuutit.
- Tehtävä liittyy seuraavaan tietokantakaavioon, joka kuvaa taiteilijoita, maalauksia, taidemuseoita ja sitä, mikä museo omistaa minkin maalauksen.

Tietokantakaavio:

```
Artists(name, born, died, country)
Paintings(id, name, painter, year, type, value)
Museums(name, city, webpage)
Owns(museumName, paintingID, since, until)
```

Relaatio `Artists` sisältää tiedot tietokannassa olevista taiteilijoista. Taiteilijoista on tallennettu nimi, syntymävuosi, kuolinvuosi ja taiteilijan kotimaa. Jos taiteilija on vielä elossa, kuolinvuosi on `NULL`. Oletetaan, että tietokannassa ei ole kahta samannimistä taiteilijaa.

Relaatio `Paintings` sisältää tiedot tietokannassa olevista maalauksista. Jokaisella maalauksella on yksikäsitteinen tunnus. Sen lisäksi maalauksesta tiedetään sen nimi, taiteilijan nimi, valmistumisvuosi, tyyppi (esimerkiksi öljyväri tai akvarelli) ja maalauksen arvo euroina.

Relaatio `Museums` sisältää tiedot museoista: museon nimen, kotikaupungin ja kotisivun osoitteen. Tietokannassa ei ole kahta samannimistä museota.

Relaatio `Owns` kertoo, mikä museo omistaa minkin maalauksen. Museon nimen ja maalauksen tunnuksen lisäksi tietoihin sisältyy se, mistä vuodesta lähtien mihin vuoteen asti museo on omistanut maalauksen. Tietokannassa säilytetään tietoa myös entisistä omistuksista, vaikka museo ei enää omistaisi teosta. Jos teos on vielä museon omistuksessa, attribuutin `until` arvo on `NULL`. Oletetaan, että sama museo ei ole omistanut samaa teosta kahteen eri kertaan.

Maalausten arvot ovat desimaalilukuja ja vuosiluvut kokonaislukuja. Muiden attribuuttien arvot ovat merkkijonoja. Voit olettaa, että relaatioiden monikoiden attribuuteilla ei ole `NULL`-arvoja attribuutteja `died` ja `until` lukuunottamatta.

Kirjoita seuraavat kyselyt SQL-kielillä:

Vinkki: voit tutkia, onko attribuutin `until` arvo `NULL` kirjoittamalla

```
WHERE until IS NULL
```

tai muu kuin `NULL` kirjoittamalla

```
WHERE until IS NOT NULL
```

- a) (2 p) Kaikkien niiden maalausten nimi, taiteilijan nimi ja valmistumisvuosi, jotka maalannut taiteilija on syntynyt vuoden 1990 jälkeen ja joiden arvo on yli 10 000 euroa.
- b) (2 p) Niiden museoiden nimet ja kotikaupungit, jotka omistavat tai ovat omistaneet yli miljoonan euron arvoisen teoksen.
- c) (2 p) Niiden museoiden nimet ja kotikaupungit, jotka omistavat tai ovat omistaneet jonkin Pekka Halosen maalaaman maalauksen, mutta eivät ole koskaan omistaneet yhtään vuoden 2000 jälkeen valmistunutta maalausta.
- d) (2 p) Haetaan museot, jotka omistavat tällä hetkellä vähintään neljä Vincent van Goghin maalaamaa taulua. Kyselyn pitää tuottaa jokaisesta tällaisesta museosta nimi, museon tällä hetkellä omistamien Vincent van Goghin taulujen lukumäärä sekä näiden taulujen yhteisarvo (taulujen määrä ja yhteisarvo lasketaan siis erikseen jokaiselle museolle vain sen omistamista tauluista).

Selitä, mihin kyselyyn seuraavat relaatioalgebran lausekkeet tuottavat vastauksen (esimerkiksi: "lauseke tuottaa kaikkien Pekka Halosen maalaamien taulujen nimet ja valmistumisvuodet")

- e) (2 p)

$$\pi_{year, country}(\text{Artists} \bowtie (\sigma_{value > 5000}(\text{Paintings})))$$

- f) (2 p) (lauseke koostuu kolmesta rivistä)

$$\begin{aligned} & \pi_{O1.museumName}(\rho_{O1}(\text{Owns} \bowtie_{paintingID=ID} (\sigma_{painter='Dali'}(\text{Paintings}))) \\ & \bowtie_{O1.ID < O2.ID \text{ AND } O1.museumName = O2.museumName} \\ & (\rho_{O2}(\text{Owns} \bowtie_{paintingID=ID} (\sigma_{painter='Dali'}(\text{Paintings})))) \end{aligned}$$

3. a) (1 p) Tee kurssin SQL-lopputesti ja -kysely A+-järjestelmässä. Testi ja kysely ovat auki 27.5. asti.
- b) (3 p) Oletetaan, että tehtävän 2 tietokanta on niin suuri, että se ei mahdu kokonaan keskusmuistiin. Oletetaan edelleen, että tietokannassa on tiedot noin 10 000 maalauksesta ja 1000 taiteilijasta, joilla on 15 eri kotimaata. Taiteilijat jakautuvat suhteellisen tasaisesti eri kotimaiden välille. Taulu `Artists` on tallennettu 30 levysivulle taiteilijoiden nimen mukaan järjestyksessä ja taulu `Paintings` 700 levysivulle niin, että maalaukset ovat siinä täysin satunnaisessa järjestyksessä. Tietokannassa tehdään hyvin usein esimerkiksi seuraavia kyselyjä:

- Haetaan määrätyn taiteilijan maalaamia tauluja, esimerkiksi

```
SELECT *
FROM Paintings
WHERE painter = 'Pekka Halonen';
```

- Haetaan taiteilijoita, joilla on määrätty kotimaa, esimerkiksi

```
SELECT *
FROM Artists
WHERE country = 'Finland';
```

Taiteilijoiden nimet ja kotimaat kyselyissä vaihtelevat, eivätkä ne ole joka kerralla samat kuin esimerkki-kyselyssä. Kannattaako taululle `Paintings` tehdä hakemisto (engl. index) attribuutin `painter` perusteella? Entä taululle `Artists` attribuutin `country` perusteella? Perustele vastauksesi muutamalla lauseella. Mitään tarkkoja laskutoimituksia ei tarvitse esittää.

4. Tarkastellaan relaatiota $R(A, B, C, D, E)$, jossa on voimassa funktionaaliset riippuvuudet $A \rightarrow B$, $B \rightarrow C$, $D \rightarrow E$ ja $D \rightarrow E$.

- a) (1 p) Perustele, miksi relaatio ei ole Boyce-Codd-normaali muodossa (BCNF:ssä).
- b) (6 p) Osita relaatio Boyce-Codd-normaali muotoon käyttämällä kurssilla (ja oppikirjassa) esitettyä algoritmia. Perustele lyhyesti jokainen muodostamasi uusi relaatio. Jatka osittamista niin pitkälle, että jäljellä on vain BCNF:ssä olevia relaatioita. Perustele, miksi lopulliset relaatiot ovat BCNF:ssä.

5. Tarkastellaan verkkokaupan varaston tilannetta kuvaavaa tietokantaa, jossa on mm. taulu `Products(productNo, name, count, price)` jossa `productNo` on tuotteen yksikäsitteinen tunnus, `name` sen nimi, `count` tuotteen kappalemäärä varastossa ja `price` tuotteen hinta verkkokaupassa. Oletetaan myös, että ennen alla kerrottujen transaktioiden suoritusta taulussa on ainakin rivi

`(R-300, superball, 20, 75.0)`

Taululle ei ole määritelty sellaisia eheysehtoja, joilla olisi merkitystä alla annettujen käskyjen onnistumisen kannalta.

Oletaan, että tietokannassa suoritetaan osittain samanaikaisesti seuraavia transaktioita T1, T2, T3 ja T4. Transaktioiden keskinäistä järjestystä ei ole määrätty, mutta jos yhden transaktion sisällä on useita käskyjä, suoritetaan ne keskenään siinä järjestyksessä kuin ne on alla annettu.

- T1 – Lukee tuotteen R-300 kappalemäärän (attribuutin `count` arvon) taulun vastaavalta riviltä.
– Lisää lukemaansa kappalemäärään 500 ja päivittää tauluun uuden kappalemäärän tuotteelle R-300.
– Päivittää taulussa tuotteen R-300 hinnaksi (attribuutin `price` uudeksi arvoksi) 80.0.
- T2 – Lisää tauluun uuden rivin
`(T-200, superballmachine, 50, 400.0)`
- T3 – Laskee, kuinka monen eri tuotteen hinta taulussa on yli 100 euroa.
– Laskee, kaikkien taulussa olevien yli 100 euron hintaisten tuotteiden kappalemäärien summan.
- T4 – Päivittää tuotteen R-300 hinnaksi (attribuutin `price` uudeksi arvoksi) taulussa 150.0.

a) (3 p)

Oletetaan, että kaikki transaktoilta vaadittavat ACID-ominaisuudet ovat voimassa. Mitkä seuraavista tilanteista ovat siinä tapauksessa mahdollisia? Merkitse kaikki oikeat vaihtoehdot.

- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 75.0 euroa.
- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 80.0 euroa.
- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 150.0 euroa.
- Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 520 ja hinta 75.0.
- Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 20 ja hinta 80.0.
- Tuotetta R-300 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten eri tuotteiden määrän ensimmäisessä käskyssä, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.
- Tuotetta T-200 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten tuotteiden hinnan ensimmäisellä kerralla, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.

b) (2 p) Oletetaan, että kaikki transaktoilta vaadittavat ACID-ominaisuudet ovat muuten voimassa, mutta ominaisuus atomisuus (engl. *atomicity*) ei mahdollisesti toteudu. Mitkä seuraavista tilanteista ovat siinä tapauksessa mahdollisia? Merkitse kaikki oikeat vaihtoehdot.

- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 75.0 euroa.
- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 80.0 euroa.
- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 150.0 euroa.
- Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 520 ja hinta 75.0.
- Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 20 ja hinta 80.0.
- Tuotetta R-300 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten eri tuotteiden määrän ensimmäisessä käskyssä, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.
- Tuotetta T-200 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten tuotteiden hinnan ensimmäisellä kerralla, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.

- c) (1 p) Oletetaan, että kaikki transaktoilta vaadittavat ACID-ominaisuudet ovat muuten voimassa, mutta sarjallistuvuuden sijaan kaikkien transaktioiden eristyvyystaso (engl. isolation level) on READ COMMITTED. Mitkä seuraavista tilanteista ovat siinä tapauksessa mahdollisia? Merkitse kaikki oikeat vaihtoehdot.
- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 75.0 euroa.
 - Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 80.0 euroa.
 - Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 150.0 euroa.
 - Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 520 ja hinta 75.0.
 - Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 20 ja hinta 80.0.
 - Tuotetta R-300 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten eri tuotteiden määrän ensimmäisessä käskyssä, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.
 - Tuotetta T-200 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten tuotteiden hinnan ensimmäisellä kerralla, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.
- d) (1 p) Oletetaan, että kaikki transaktoilta vaadittavat ACID-ominaisuudet ovat muuten voimassa, mutta sarjallistuvuuden sijaan kaikkien transaktioiden eristyvyystaso on REPEATABLE READ. Mitkä seuraavista tilanteista ovat siinä tapauksessa mahdollisia? Merkitse kaikki oikeat vaihtoehdot.
- Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 75.0 euroa.
 - Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 80.0 euroa.
 - Kaikkien transaktioiden suorittamisen jälkeen tuotteen R-300 hinta on 150.0 euroa.
 - Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 520 ja hinta 75.0.
 - Tietokanta jää tilaan, jossa tuotteen R-300 lukumäärä taulussa on 20 ja hinta 80.0.
 - Tuotetta R-300 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten eri tuotteiden määrän ensimmäisessä käskyssä, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.
 - Tuotetta T-200 ei lasketa mukaan, kun T3 laskee yli 100 euron hintaisten tuotteiden hinnan ensimmäisellä kerralla, mutta se on mukana, kun T3 laskee yli 100 euron hintaisten tuotteiden kappalemäärän summan.

Muista vastata kurssin palautekyselyyn viimeistään 2.6.2020. Kyselyyn vastaamisesta saa 20 harjoitustehtäväpistettä. Linkki kyselyyn on lähetetty kurssille ilmoittautuneille 13.5. sähköpostitse. Huomaa, että tämä on eri kysely kuin tehtävässä 3a mainittu lopputesti ja -kysely.