

Helsingin yliopisto/tietojenkäsittelytieteen laitos
Tietokannan hallinta, kurssikoe 16.12.2004 / H. Laine

Kirjoita jokaiseen erilliseen vastauspaperiin kurssin nimi, tenttipäivämäärä, nimikirjoituksesi ja nimen selvennys sekä syntymäaikasi.

Anna kunkin tehtävän vastaus erillisellä vastauspaperilla.

1. a) Selvitä lyhyesti, mitkä ovat B+ -puun ja ISAM-rakenteen merkittävimmät erot. (6p)
Kummassakin rakenteessa on harva puumainen hakemisto järjestetyn perustiedoston 'päällä'. Haku rakenteista tapahtuu periaatteessa samalla tavalla. ISAM-rakenteen hakemisto on staattinen: hakemisto luodaan tiedoston uudelleenorganisoinnin tai tiedoston perustamisen yhteydessä. ISAM-rakenteen lehtisivut ovat fyysisesti peräkkäisiä. Jos sivulle tulee enemmän tietueita kuin mahtuu, käytetään ylivuotoketjuja. B+ -puu uudelleenorganisoiuu automaattisesti. Myös hakemisto kasvaa ja kutistuu. Lehtitason solmut on kytketty kaksisuuntaisesksi linkitetyksi listaksi.
avainsanoja alleviivattu staattisuus vs. dynaamisuus 3p, muista lisäpisteitä

b) Esitä B+ -puun lisäysalgoritmin periaate. (6p)

Hae lisäyksen kohdesivu (=sivu, jolle tietue pitäisi sijoittaa)

Olkoon kohdesivu P ja hakemistosivu, jolta sinne päädyttiin R.

Jos sivulla P on tilaa, lisää tietue sivulle.

Jos sivulla P ei ole tilaa:

ota käyttöön uusi sivu S

jaa sivun P tietueet puoliksi sivun P ja sivun S kesken

lisää uusi tietue joko sivulle P tai S avaimen arvosta riippuen

kytke sivu S linkitettyyn lehtisivujen listaan sivun P jälkeen

lisää sivun S pienin avain ja S:n osoite hakemistosivulle R oikeaan kohtaan

jos avain ja osoite -pari ei mahdu sivulle R:

Ota käyttöön uusi hakemistosivu R'

Olkoon K lisäyksen jälkeen keskimäinen sivun R erottimista

K:ta suuremmat erottimet osoittimiseen siirretään sivulta R sivulle R'.

Eroin K ja osoite sivulle R' lisätään hakupolulla ylempänä olevaan hakemistotietueeseen

Tämä voi jälleen johtaa hakemistosivun jakoon.

jos puolituksen idea käy riittävän hyvin esiin vähintään 3p.

c) Tiedosto toteutetaan B+ -puu rakenteena. Montako tietuetta enintään ja vähintään voidaan tallettaa B+ -puuhun, jonka korkeus on 3, kun puun lehtisivulle mahtuu enintään 20 tietuetta ja hakemistosivulle enintään 200 osoitinta. (6p)

vähintään: $2 \cdot 100 \cdot 10 + 1 = 2001$ (myös $2 \cdot 100 \cdot 10$, $2 \cdot 101 \cdot 10 + 1$ ja $2 \cdot 101 \cdot 10$ käyvät)

enintään: $200 \cdot 200 \cdot 20 = 800000$

2. Tarkastellaan tauluja

Tuote(tuotenumero, tuotenimi, paino, hinta,..)

Tilaus(tilausnumero, tilaaja->asiakas, tilauspäivä, toimituspäivä,...)

Asiakas(asiakasnumero, nimi,..., muuta),

Tilausrivi(tilausNo->tilaus, rivinumero, tuote->tuote, määrä)

Tietokannassa käytettävä tiedostojen sivukoko on 4kB. Taulujen koot ovat seuraavat:

Taulu	Rivejä	Rivejä sivulla keskimäärin
Tuote	4000	10
Tilaus	40000	10
Asiakas	2000	5
Tilausrivi	160000	100

Toimittamattomia tilauksia on noin 1000. Näiden toimituspäivä-sarakkeessa on tyhjäarvo (null). Tilaukset jakautuvat tasaisesti asiakkaiden ja tuotteiden suhteen. Taulut on toteutettu järjestämättöminä peräkkäisrakenteina (kasa). Niillä on B+ -puina toteutetut oheishakemistot pääavaimen ja viiteavaimien (osoitettu nuolilla '->') perusteella. Voit olettaa, että hakemistotietueita mahtuu yhdelle hakemistosivulle enintään noin 400.

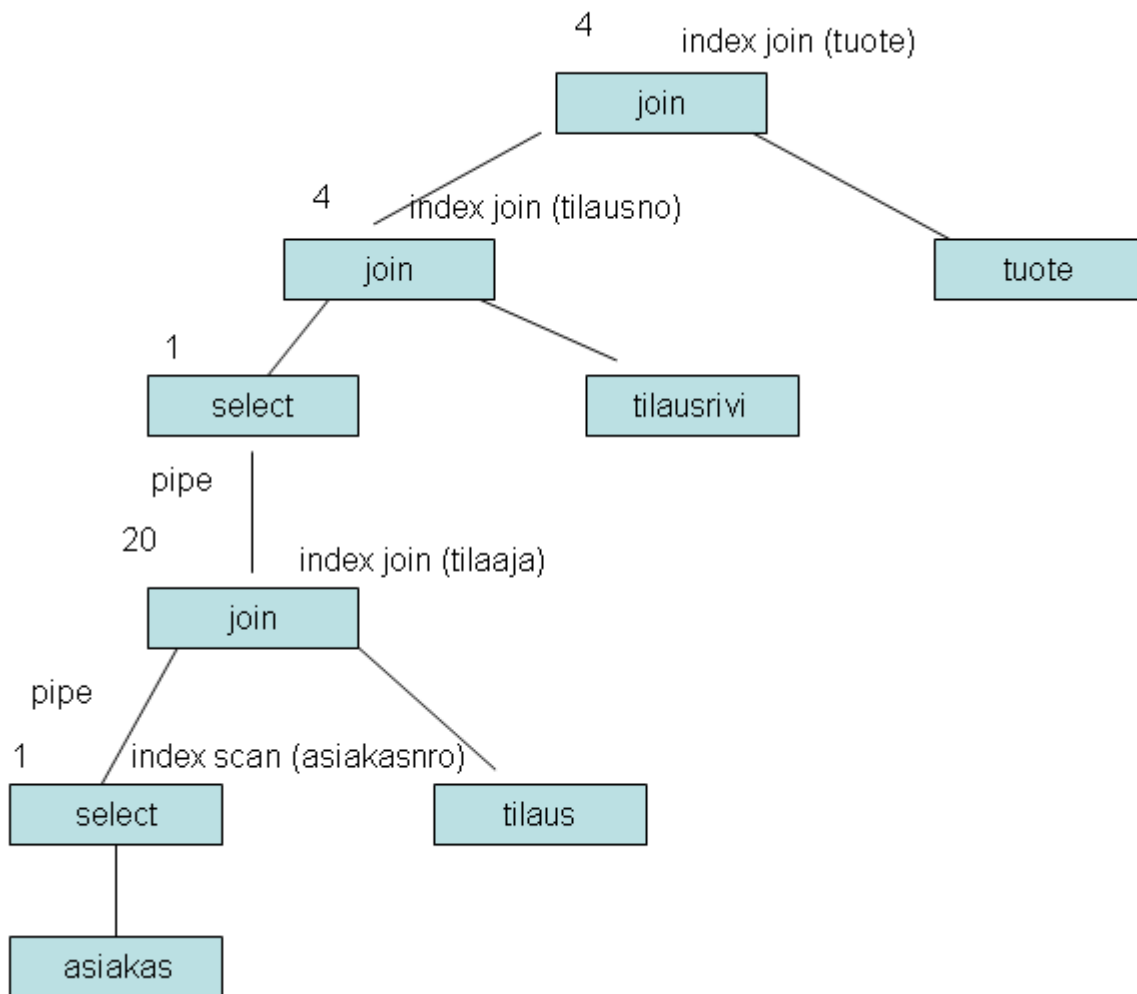
Tarkastellaan kyselyä

```
select asiakasnumero,tilausnumero,tuotenumero,tuotenimi, hinta
from tilaus, asiakas, tilausrivi, tuote
where asiakasnumero=12345 and
      toimituspäivä is null and
      tilaus.tilaja=asiakas.asiakasnumero and
      tilaus.tilausnumero=tilausrivi.tilausNo and
      tilausrivi.tuote=tuote.tuotenumero
order by asiakasnumero, tilausnumero;
```

a) Esitä arviosi vastauksen koosta. (6p)

Asiakkaalla keskimäärin $40000/2000 = 20$ tilausta, joista $1/40$ osa = 0.5 toimittamattomia. Pyöristetään yhteen. Tilaukseen liittyy keskimäärin 4 riviä, eli odotettu vastauksen koko olisi 4 riviä.

b) Miten kysely olisi tehokkaimmin suoritettavissa ja montako levyhakua tällöin tarvitaan. Perustele vastauksesi. Keskusmuistitilaa käytettävissä n. 50 sivun verran. (12p)



Hakemistojen korkeudet: 1 taso <400, 2 tasoa $400 - 400*400=160000$, 3 tasoa: $2*200*200=80000- 64000000$

Haku asiakasnumeron perusteella asiakas taulusta: 2000 tietuetta => 2 tasoinen hakemisto 2+1 hakua

Liitos asiakas/tilaus: Tilauksia 40000 => 2 tasoinen hakemisto => 2+20 hakua

Valinta: toimittamattomat: ei levyhakuja, jäljellä 1 tilaus

Liitos tilaus/tilausrivi: Tilausrivejä 160000 => 3 hakemistotasoa => 3+4 hakua

Liitos tilausrivi/tuote: Tuotteita 4000 => 2 tasoinen hakemisto => $4*(2+1)$ hakua

yhteensä $(2+1)+(2+20)+(3+4)+(4*(2+1))= 44$.

Arvosteltavat: toteutusidea (0-6p), tehottomista ratkaisuksista sakotetaan, mitä huonommin toimiva sitä enemmän.

Laskenta: (0-6p), jos laskettu oikein huonommallekin ratkaisumallille saa täydet laskentapisteet.

3. a) Mikä on luku- ja kirjoituslukon ero? (4p)

Lukulukko antaa oikeuden lukea, kirjoituslukko kirjoittaa.

Usealla prosessilla voi olla lukulukko samaan tietoaalkioon. Kirjoituslukko on poissulkeva.

Jos yhdellä prosessilla on kirjoituslukko, muilla ei voi olla mitään lukkoja samaan tietoaalkioon.

b) Selitä ankaran kaksivaiheisen lukituksen (strict 2 phase locking, 2PL) periaate. (5p)

2PL: kaikki lukot varataan ennen kuin yhtään vapautetaan.

Strict 2PL: 2PL + lukot vapautetaan vasta tapahtuman sitoutuessa.

c) Selitä, mitä tarkoitetaan lukkiutumalla (deadlock) ja miten lukkiutumia voidaan estää tai purkaa? (5p)

Lukkiutuma: Tapahtuma ei pysty etenemään, koska se odottaa lukon vapautumista toiselta tapahtumalta. Tämä puolestaan ei pysty etenemään ennen kuin siltä lukon vapautumista odottava pääsee etenemään.

Estämiskeinoja:

- Kaikki lukot varataan heti transaktion alussa.
- Tapahtuma ei jää koskaan odottamaan vaan keskeytetään ja käynnistetään myöhemmin uudelleen.
- Kun lukkiutuma havaitaan, keskeytetään jokin lukkiutumaketjussa oleva ja käynnistetään se myöhemmin uudelleen.