



Liitoksen toteutus

- Edellä esitelty putkiin perustuva lomituspöytä soveltuu hyvin tilanteisiin, joissa tiedetään että riville löytyy enintään yksi pari
 - esimerkiksi avaimen ja viiteavaimen vertailu on tällainen tilanne
 - Tällöin taululla on yleensä avaimen perustuva järjestetty indeksi tai taulu on organisoitu ISAM tai B+ -puuksi on tietueet luettavissa järjestyksessä pääavaimen mukaisesti,
 - Tarvitaan järjestäminen vain liitoksen toiselle osapuolelle
 - Esimerkissämme vain taulu suoritettiin järjestettävissä
 - jos taulu opiskelija olisi toteutettu B+ puuna tarvittaisiin $3 \cdot 2000 + 500 = 6500$ hakua
 - jos taulu opiskelija olisi kasarakenteella jolla olisi B+ -puu indeksi tarvittaisiin $3 \cdot 2000 + n \cdot 60$ (hakemistosivut) + $20 \cdot 500$ hakua = 16060

1



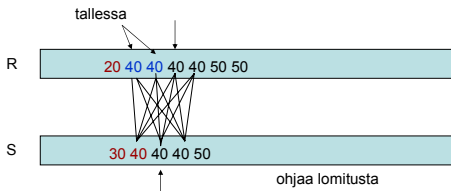
Liitoksen toteutus

- Jos tietueparien muodostuksessa vallitsee **monen suhde moneen** yhteys (yhdeällä R:n tietueella voi olla useita liittosehdon toteuttavia pareja ja myös toisinpäin) tulee lomituserästä hankalampi
 - Kun jonon tietueilla on enintään yksi pari voidaan parin teon jälkeen edetä jonossa.
 - Kun jonon tietueilla voi olla useita pareja ja liittosarakkeen arvo säilyy siirryttäessä seuraavaan tietueeseen, pitää jo paritetut tietue ottaa talteen ja käsitellä talteen otetut tietueet siinä vaiheessa kun kumppanijonossa siirrytään eteenpäin. Talteen otettujen määrä voi kasvaa suureksi, jos sama arvo toistuu usealla rivillä

2



Liitoksen toteutus



ongelmallinen paritus lomituksessa

3



Liitoksen toteutus

- Hajautusliitos (hash join)**
 - soveltuu yhtäsuuruusliitokseen
 - Hajautusliitos jakautuu kahteen vaiheeseen
 - jakovaiheeseen ja
 - tutkintavaiheeseen
 - Jakovaiheessa luetaan kumpikin tiedosto ja hajautetaan se hajauttimella h B soluun tiedostoihin HR ja HS,
 - Sivuja kirjoitetaan levyille sitä mukaa kun ne täyttyvät ja kytketään tarvittaessa ylivuotoketjuun.
 - Solussa HR[i] olevien tietueiden parit löytyvät siis solusta HS[i]

4



Liitoksen toteutus

- Tutkintavaiheessa käydään tiedostot HR ja HS kertaalleen läpi solupari kerrallaan
 - Soluparin käsittely voi tapahtua esim. hajauttamalla pienemmän tiedoston sivun tietueet keskusmuistissa olevaan hajautustaulukkoon hajauttimella h2 ja
 - käymällä sitten suuremman tiedon solun tietueet läpi siten, että kullekin vuorollaan lasketaan hajautusosoite hajauttimella h2 ja tutkimalla ko. solun tietueet ja tehdään tulostietue tarvittaessa

5



Liitoksen toteutus

- Levyoperaatioiden (luku + kirjoitus) määrä jakovaiheessa on $2 \cdot (\text{pages}(R) + \text{pages}(S)) + F$, missä F on vajaista sivuista johtuva korjaus, $F < 2B$
- Yhdistämisvaiheessa luetaan $\text{pages}(R) + \text{pages}(S) + F$ sivua,
- Eli yhteensä tarvitaan $3 \cdot (\text{pages}(R) + \text{pages}(S)) + 2F$ sivua.

6



Liitoksen toteutus

- Esim. Opiskelija & suoritus liitos vaatisi $3 \cdot (500 + 2000) + 50$ (jos $B=50$) =7550 hakua
- satunnaishakuajalla saadaan noin 75s
- jakovaiheessa voidaan hyödyntää peräkkäishakua jolloin aika hieman vähenee

7



Liitoksen toteutus

- Jos pienempi tiedostoista on niin pieni, että se mahtuu hajautettavaksi keskusmuistin hajautustaulukkoon voidaan jakovaihe jättää pois
- Tällöin levyoperaatioiden määrä on vain $\text{pages}(R) + \text{pages}(S)$,
 - eli sama kuin jaksotettujen sisäkkäisten silmukoiden tapauksessa kun pienempi voidaan lukea yhtenä jaksona
 - hajautustaulun käyttö on todennäköisesti nopeampi tapa muodostaa parit kuin etsiä niitä silmukassa.

8



Joukko-operaatioiden toteutus

- Relaatioalgebran joukko-operaatioissa tyypillinen ratkaisu on käyttää lomitusliitoksen tai hajautusliitoksen muunnelmia
 - yhdisteen lomitusvaiheessa otetaan kaikki mukaan ja karsitaan toistuvat rivit
 - leikkauksen lomitusvaihe valitsee yhteiset
 - erotukset lomitusvaiheessa valitaan ne, joita ei löydy toisesta tiedostosta.
 - levyoperaatioiden määrät vastaavat lomitus- ja hajautusliitoksen määriä.

9



Koostefunktioiden toteutus

- SQL:n koostefunktioita voidaan soveltaa koko aineistoon tai ryhmiin.
- Jos koostefunktion argumentti esiintyy hakemistosarakkeena, voidaan koostefunktion arvo laskea pelkästään tiheän hakemiston läpikäynnillä tutkimatta tietueita.
- Min ja max saadaan selville harvastakin hakemistosta hakemalla ensimmäinen / viimeinen lehtisivu
- Lukumäärä voidaan selvittää pääavainhakemiston avulla

10



Ryhmittely

- SQL:n ryhmittely on toteutettavissa
 - järjestämällä tiedosto ryhmittelysarakkeiden perusteella
 - samaan ryhmään kuuluvat tulevat peräkkäin, jolloin tunnusluvut on laskettavissa yhdellä läpikäynnillä

11