

Tietokantakyselyjen optimointi relaatiotietokannassa

Olli Rissanen

Kandidaatintutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Helsinki, 24. maaliskuuta 2013

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Olli Rissanen			
Työn nimi — Arbetets titel — Title			
Tietokantakyselyjen optimointi relaatiotietokannassa			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Kandidaatintutkielma		24. maaliskuuta 2013	4
Tiivistelmä — Referat — Abstract			
<p>Tutkielmassa tutustutaan tietokantakyselyjen optimointiin relaatiotietokantojen hallintajärjestelmien osalta sekä optimoinnin vaikutukseen kyselyjen suorituskyyvyssä.</p>			
Avainsanat — Nyckelord — Keywords			
Information systems Query optimization			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

1 Johdanto

Modernit järjestelmät lisäävät jatkuvasti tietokantojen työtaakkaa tiedon määrän kasvaessa. Jotta tiedosta saadaan mahdollisimman paljon irti, tarvitaan tiedon hallitsemiseen yhä tehokkaampia työkaluja. Tietokannan suorituskyky on tärkeää koko järjestelmän suorituskyvyn osalta, sillä tiedon lukeminen massamuistista on hidasta verrattuna rekistereiden tai välimuistin käyttöön. Optimoimalla tietokantakyselyjen suoritusta voidaan vaikuttaa suoritettujen operaatioiden määrään sekä muistialueen kokoon ja siten vähentää tietokannan vasteaikaa sekä resurssien käyttöä. [MKR12]

Tietokantaa käytetään tietokannan hallintajärjestelmällä, joka on koelma ohjelmia tiedon tallentamiseen, muokkaamiseen, analysointiin ja keräämiseen tietokannasta. Hallintajärjestelmää käytetään kyselykielellä, joista esimerkiksi SQL on suunniteltu relaatiotietokantojen hallintajärjestelmille. Hallintajärjestelmän vastuulla on kyselyn muuttaminen tietokannan ymmärtämään muotoon säilyttäen kyselyn alkuperäisen tarkoituksen. Kyselyn optimointi on toteutettu automaattisena toimenpiteenä tietokannan hallintojärjestelmän sisältämässä kyselyn optimoijassa, ja kaikista hallintajärjestelmän komponenteista optimoijalla on suurin merkitys tietokannan suorituskykyyn. [MKR12] Kyselyn optimoijan tavoitteena on minimoida itse optimointiin käytetty aika ja maksimoida optimoinnista saatu hyöty. [JK84]

Optimoija toimii etsien kyselyä vastaavat kyselysuunnitelmat ja valitsemalla niistä tehokkaimman. Kyselysuunnitelma sisältää sarjan algebrallisia operaatioita tietokannan relaatioille jotka tuottavat tulokseksi halutun vastauksen. Tietokantakyselyä vastaavia kyselysuunnitelmia voi olla useita, sillä kyselyjen algebralliset esitykset voidaan usein esittää monena loogisesti vastaavana esityksenä. [JK84] Algebrallista operaatiota kohden voi myös löytyä useita toteutuksia, kuten join-operaatiota toteuttavat merge join ja hash join.

Saman kyselyn tuottamat kyselysuunnitelmat voivat olla suorituskyvyltään jopa eri suuruusluokassa. [Ora13]

Optimointi on vaikea hakuongelma, jossa hakualue voi nousta erittäin suureksi. [Cha98] Haasteeksi nousee kyselysuunnitelmien luominen ja niiden suorituskyvyn ennustaminen. Kaikkien mahdollisten kyselysuunnitelmien luominen on usein liian hidasta, joten optimoijan tulee valita pienin mahdollinen hakualue joka pitää sisällään halvimmat suunnitelmat. Suorituskyvyn ennustamisen ja hakualueen rajauksen lisäksi optimoija tarvitsee tehokkaan algoritmin koko hakualueen läpikäymiseen. On epärealistista odottaa kyselyn optimoijan aina löytävän parhaan kyselysuunnitelman, ja onkin tärkeämpää välttää huonoimpia suunnitelmia ja löytää hyvä suunnitelma. [RG03]

Kappale 2 sisältää esitiedot kyselyn optimoijan toiminnalle, jonka lisäksi kappaleessa perehdytään optimoijan rakenteeseen hallintajärjestelmän sisällä. Kappale 3 käsittelee optimointiprosessin vaiheita yksityiskohtaisesti. Kappaleissa 4 ja 5 esitellään kyselyn optimoijan käyttöä MySQL-hallintajärjestelmällä.

Lähteet

- [Cha98] Chaudhuri, Surajit: *An overview of query optimization in relational systems*. Teoksessa *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, sivut 34–43. ACM, 1998.
- [JK84] Jarke, Matthias ja Jurgen Koch: *Query optimization in database systems*. ACM Computing surveys (CsUR), 16(2):111–152, 1984.
- [MKR12] Mor, Jyoti, Indu Kashyap ja RK Rathy: *Analysis of Query Optimization Techniques in Databases*. International Journal, 47, 2012.

- [Ora13] Oracle: *MySQL 5.0 Reference Manual*, 2013. http://docs.oracle.com/cd/E17952_01/refman-5.0-en/controlling-optimizer.html.
- [RG03] Ramakrishnan, R. ja J. Gehrke: *Database Management Systems*. McGraw-Hill international editions: Computer science series. McGraw-Hill Education, 2003, ISBN 9780072465631. <http://books.google.fi/books?id=JSVhe-WLGZOC>.