

Tietokantakyselyjen optimointi relaatiotietokannassa (vai niiden hallintajärjestelmissä?)

Olli Rissanen

Kandidaatintutkielma
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Helsinki, 24. helmikuuta 2013

Tiedekunta — Fakultet — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Olli Rissanen			
Työn nimi — Arbetets titel — Title			
Tietokantakyselyjen optimointi relaatiotietokannassa (vai niiden hallintajärjestelmissä?)			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Kandidaatintutkielma		24. helmikuuta 2013	0
Tiivistelmä — Referat — Abstract			
<p>Tutkielmassa tutustutaan tietokantakyselyjen optimointiin relaatiotietokantojen hallintajärjestelmien osalta sekä optimoinnin vaikutukseen kyselyjen suorituskyyssä. Tärkeimmät suorituskyykymittarit ovat prosessorin ja muistin käyttö.</p>			
Avainsanat — Nyckelord — Keywords			
Information systems Query optimization			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

Sisältö

1 Johdanto

Tietokantojen suorituskyky on yhä tärkeämpää tiedon määrän kasvaessa. Automatisoidulla optimoinnilla voidaan kasvattaa ohjelmistojen suorituskykyä ja nopeutta tietokantojen käyttäjien tiedonhakua (data-access?).

Tutkielman tarkoituksena on ..

Tietokannalla tarkoitetaan sen sisältämää dataa ja tietorakenteita. Tietokantajärjestelmä sisältää tietokannan hallintajärjestelmän sekä tietokannan.

Tietokannan hallintajärjestelmä on kokoelma ohjelmia tiedon tallentamiseen, muokkaamiseen sekä tiedon keräämiseen tietokannasta. Tietokannan rakenteen monimutkaistuessa tieto voidaan kerätä usealla eri tavalla, ja samankaltaisten kyselyjen erot voivat olla erittäin suuria. Hallintajärjestelmän sisältämän kyselyoptimoijan tehtävänä on löytää kyselylle suorituskykyisin kyselysuunnitelma mahdollisimman nopeasti. Optimoinnilla saavutettu prosessorin ja muistin käytön väheneminen kyselyjä suoritettaessa parantaa tietokannan suorituskykyä huomattavasti.

Kyselyoptimoijan tavoitteena on minimoida itse optimointiin käytetty aika ja maksimoida optimoinnista saatu hyöty. Kyselyoptimoija toimii etsien kyselyä vastaavat mahdolliset kyselysuunnitelmat ja valitsemalla niistä tehokkaimman. Tietokantakyselyä vastaavia kyselysuunnitelmia voi olla useita, sillä kyselyiden algebralliset esitykset voidaan usein esittää monena loogisesti vastaavana esityksenä. Algebrallista operaatiota kohden voi myös löytyä useita toteutuksia, kuten join-operaatiota toteuttavat merge join ja hash join. Haasteena on tehokkaan kyselysuunnitelman luominen ja kyselysuunnitelmien suorituskyvyn etukäteinen ennustaminen.

Tutkielman rakenteesta

2 workname: Taustaluku

Tietokannan hallintajärjestelmien jako: query optimizer ja query execution engine.

Tietokantakyselyiden optimoinnilla viitataan tietokantakyselyn suorittamiseen mahdollisimman tehokkaasti. Optimoinnin tavoitteena on joko maksimoida suorituskyky annetuilla resursseilla tai minimoida resurssien käyttö. Mitattavia resursseja ovat suorittimen ja muistin käyttö sekä kommunikointikustannukset. Muistin käyttö jakautuu tallennuskustannukseen sekä ulkomuistiin pääsyn kustannukseen. Tallennuskustannuksella tarkoitetaan ulkomuistin sekä puskurimuistin käyttöä, ja se tulee aiheelliseksi kun muistin käyttö aiheutuu pullonkaulaksi.

Resurssin merkitys riippuu tietokantatyypistä. Hajautetuissa tietokannoissa hitailla yhteysväylillä kommunikointikustannukset hallitsevat kustannuksia. Paikallisesti hajautetuissa tietokannoissa kaikilla resursseilla on sama painoarvo. Keskitetyissä tietokannoissa ulkomuistiin pääsyn kustannus ja

prosessorin käyttö ovat oleellisia. Tämän tutkielman aihepiiriin kuuluu vain keskitettyjen tietokantojen optimointi.

todo: liitoskohta

Relaatiotietokanta on relaatiomalliin perustuva tietokanta. Relaatiomallin keskeinen piirre on kaiken datan esittäminen n-paikkaisen karteesisen tulon osajoukkona, ja se tarjoaa deklaraatiivisen menetelmän datan ja kyselyjen määrittämiseen. Relaatiomalli koostuu attribuuteista, monikoista ja relaatioista. Matemaattisessa määritelmässä attribuutti on pari joka sisältää attribuutin nimen ja tyyppin sekä jokaiseen attribuuttiin liittyy sen arvojoukko. Monikko on järjestetty joukko attribuuttien arvoja. Relaatio koostuu otsakkeesta ja sisällöstä(body?), jossa otsake on joukko attribuutteja ja keho on joukko monikkoja. Relaation otsake on myös jokaisen monikon otsake. Visuaalisessa esityksissä relaatio on taulukko ja monikko taulukon rivi.

todo: SQL ja relaatiomalli

todo: relaatiotietokanta vs no-sql

todo: optimoi menee

[?]

3 menetelmä 1

4 menetelmä 2

5 menetelmä n

6 menetelmien vertailu

7 yhteenveto

Lähteet