Hajautettujen tietokantojen ongelmat

Jan Lindström jplindst@cs.helsinki.fi

1. helmikuuta 2001

Hajautettu tietokanta (distributed database, DDS) on kokoelma loogisesti yhteenliittyviä tietokantoja, jotka ovat hajautettu yli tietoverkon. *Hajautettu tietokannanhallintajärjestelmä* (distributed database management system, DDBS) on ohjelmisto, joka hallinnoi hajautettua tietokantaa ja mahdollistaa käyttäjälle hajautusriippumattoman rajapinnan.

Tietokannan hajauttaminen useille tietokoneille verkossa aiheuttaa uusia ongelmia, jotka pitää ratkaista. Hajautettujen tietokantojen tutkimuksessa on käsitelty seuraavia ongelmakohtia:

- Hajautetun tietokannan suunnittelu: kuinka tiedot talletetaan hajautettuun tietokantaan. Tietokanta voi olla ositettu tai toisinnettu.
- Hajautettu kyselynsuoritus: kuinka valita tehokas tai vaihtoehtoisia tapoja kyselyn suorittamiselle verkon yli.
- Hajautetun hakemiston käsittely: miten löydetään missä tiedot sijaitsevat. Onko hakemisto keskitetty vai hajautettu.
- Hajautettu samanaikaisuudenhallinta: kuinka taataan useaan tietokantaan operoivan tapahtuman oikeellisuus.
- Hajautettujen lukkiumien hallinta: kuinka estetään tai löydetään lukituskäytäntöä käytettäessä lukkiumat.
- Hajautetun tietokannan luetettavuus: kuinka taataan hajautetun tietokannan eheys. Kuinka havaitaan virheet ja toivutaan niistä.
- Käyttöjärjestelmätuki: millaisia käyttöjärjestelmäominaisuuksia hajautettu tietokanta vaatisi.

- Heterogeeniset tietokannat: kuinka yhdistää tietoa useita eri tietomalliin perustuvista tietokannoista.
- Turvallisuus: tapa varmistaa tietokantaa käyttävän asiakkaan identiteetti.

Moniin edellä esitettyihin ongelmiin on kirjallisuudessa esitetty joko vain osittaisia tai heuristisia ratkaisuja. Esityksessä tavoitteena on esitellä eri ongelmien perusteet ja ainakin osaan jokin kirjallisuudessa esiimtynyt ratkaisu.

Tietokannanhallintajärjestelmät ovat suuria ja monimutkaisia järjestelmiä. Järjestelmän käytön ja ylläpidon opiskelu ja osaamisen ylläpito vie paljon aikaa. Tämän vuoksi kirjallisuudessa on esitetty aivan uusi lähestymistapa tietokantahallintajärjestelmille. Tätä arkkitehtuurimallia kutsutaan adaptoituvaksi RISC-tietokantamalliksi. Esityksessä pyritään antamaan perustelut miksi uutta tietokanta-arkkitehtuuria tarvitaan ja millainen on RISC-tietokantamalli. Koska tutkimus on tällä alueella hyvin nuorta, luodaan katsaus olemassaoleviin tutkimusongelmiin.

Viitteet

[1] D. Bell and J. Grimson. *Distributed Database System*. Addison-Wesley, 1992.

[2] P. A. Bernstein and N. Goodman. Concurrency control in distributed database systems. *ACM Computing Surveys*, 13(2), June 1981.

[3] S. Chaudhuri and G. Weikum. Rethinking database system architecture: Towards a self-tuning RISC-style database system. In *Proceedings of the 26th International Conference on Very Large Databases*, Cairo, Egypt, 2000. Morgan Kaufmann Publishers.

[4] M. T. Özsu and P. Valduriez. *Principles of Distributed Database System*. Prentice Hall, second edition, 1999.