

Aikasarjatietomalli mittaustietojen hallintaan

Antoni Wolski,
VTT Tietotekniikka

Mittaustiedot ovat luonteeltaan aikasarjoja. Nykyiset tietomallit eivät kuitenkaan tarjoa riittävästi tukea tehokkaaseen mittaustietojen tallentamiseen ja hakuun. Apua löytyy joistakin yksittäisistä tuotteista ja VTT Tietotekniikassa kehitetystä RQL-tietomallista.

Unohdetut aikasarjat

Monen sulautetun järjestelmän tehtäviin kuuluu mittaustietojen keruu ja käsittely. Pienimuotoisessa toteutuksessa asia hoituu luontevasti sovelluskohtaisen ohjelman puitteissa. Laajempi toteutus, esimerkiksi tuotantolinjan hallinta-ohjelma, vaatii jo yleisemmän ratkaisun.

Koska yksittäiset mittaukset ovat sidottuja mittausaikaan, mittaustiedoista muodostuu aikasarjoja, joilla on seuraavat yhteiset ominaisuudet:

Tietopiste on rakenteellinen tieto, joka esittää reaali maailman mittauspistettä. Osa tietopistettä kuvaavista tiedoista on ajasta riippumattomia, kuten nimet, mittauslaitteen tyyppimerkinnot ja tekniset tiedot.

Aikasarja esittää jotakin ajassa mitattua tietopisteen suurenta tai suureita. Aikasarjassa olevia tietoja yhdistää kronologinen järjestys.

Aikasarjan tietue kertoo yksittäisestä mittauksesta. Jokaisen tietueen yhteydessä on olemassa tieto mittausajasta ja mahdollisesti tietueen voimassaoloajasta (valid time). Tietueeseen kuuluu yksi tai kaksi aikaleimaa, riippuen toteutustavasta.

Aikasarjan alkio on tietueen elementti. Se esittää mitatun arvon.

Tietopistetyyppi on valmis kaava, jonka mukaan voidaan luoda uusia tietopisteitä aikasarjoihin.

Tietopisteryhmä on kokoelma tietopisteitä, jotka yhdistää jokin tekijä.

Aikasarjatietokanta on joukko tietopisteitä tyyppineen, ryhmineen ja aikasarjoihin. Sitä hallitaan ja käsitellään sopivan ohjelmiston avulla.

Tyypillisin aikasarjatietokannan toiminto on voimassaoloajan perusteella tapahtuva liitos (eli voimassaoloaikaliitos, valid time join). Sellainen tarvitaan, kun esimerkiksi halutaan tietää, *mikä oli prosessin tila ajanhetkellä t_i* . Silloin ei riitä, että järjestelmä tietää mitausarvon mitaushetken – tiedossa pitäisi olla myös hetki johon asti mittaus on voimassa. Silloin vasta voidaan muodostaa tulos kaikista mittausarvoista, joiden voimassaoloaika sisältää kyseisen ajanhetken t_i .

Aikasarjojen hallintajärjestelmälle voidaan asettaa muitakin vaatimuksia, kuten että mittausarvot interpoloidaan mitaushetkien välillä, koostearvot (esim. keskiarvot) lasketaan automaattisesti ja vanhentuneet tietueet poistetaan. Aikasarjat voivat olla myös anakin kahta eri tyyppiä. Säännöllisessä sarjassa mittausajat ovat säännöllisin aikaväleihin. Epäsäännöllisessä sarjassa mittausajat voivat olla mielivaltaisia.

Aikasarjakyselykieli tarvitaan ja on selvää, että aikapohjaisten toimintojen (kuten voimassaoloaikaliitoksen) optimointi on tärkeää.

Yleiskäyttöisen aikasarjahallinnan tarpeet olivat hyvin pitkään vailla niille kuuluvaa huomiota. Tutkimusallalla on herätty ongelmaan vasta viime aikoina [1]. Viime vuosina on kaupallisiin tietokantajärjestelmiin tuotu joukko aikasarjalaajennuksia, jotka eivät kuitenkaan tue hyvin mittausaineiston tallennustarpeita.

Ensin pohditaan, miksei relaatiotietokanta taivu tähän tarpeeseen.

Sopimaton relaatiotietokanta

Aikasarjan tuominen relaatiokantaan on kivuliasta vaikei mahdotonta.

Esimerkiksi seuraava taulu esittää lämpötila-antureiden mittauksia. Taulun joka rivillä on mitaushetken aikaleima.

Probe_id	Time	Tempr
TEMP12	10:20	24
TEMP34	12:10	99
TEMP12	13:10	20

On selvää, että aikaleima kertoo voimassaoloajan alkuhetken, kuten rivillä (TEMP12, 10:20). Mutta missä kerrotaan voimassaoloajan loppu? Se löytyy jollakin toisella rivillä, joka esittää samaan yksilöön kohdistuvan seuraavan muutoksen, eli tässä tapauksessa rivillä (TEMP12, 13:10).

Kun halutaan suorittaa eri mitaustaulujen voimassaololiitos, tulosrivi muodostetaan voimassaoloajaltaan leikkavista lähtötaulukien riveistä. Miten sitten kirjoitetaan yllä olevan taulutyypin mukaisista tauluista lähtevä SQL-kysely? Ei mitenkään, koska SQL ei tunne käsitettä *ajassa seuraava rivi*.

Ratkaisuna on koko voimassaoloaikatiedon kirjoittaminen joka riville, kuten seuraavassa esimerkissä.

Probe_id	Beg	End	Tempr
TEMP12	10:20	13:10	24
TEMP34	12:10	NULL	99
TEMP12	13:10	NULL	20

NULL-arvot tulkitaan siten, että tieto on voimassa nykyhetkelläkin.

Täten rakennettujen taulujen voimassaololiitos onnistuu nyt yhdellä SQL-komennolla, mutta lukija voi kokeilla itse, tuleeko siitä lyhyt. Voidaan myös perustellusti epäillä, kuinka hyvin tietokantatuotteen optimoija selviää tehtävästään.

Vaikka ratkaisu toimii, siinä on muitakin ongelmia. Aikatiedot on esitetty tuplasti ja tietojen päivityksissä täytyy noudattaa tiukkoja sääntöjä: rivin lisäämiseen tai muuttamiseen liittyy aina jonkun toisen rivin muuttamista. Taulun aikaheys voi helposti rikkoutua, eikä sitä voi ylläpitää automaattisesti järjestelmän toimesta. Lähteestä [2] löytyy muitakin keinoja, mutta relaatiomallin kankeudesta ei niiden avulla pääse eroon.

TSQL2-kieli

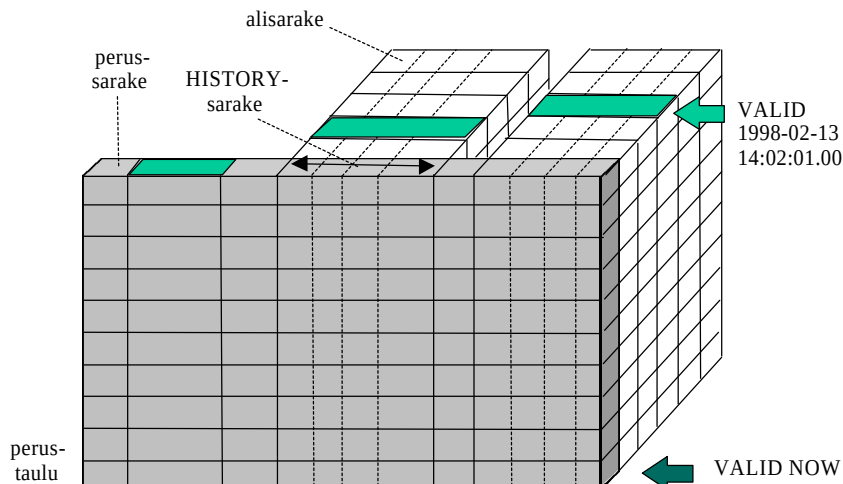
Tietokantakirjallisuutta tuntevat lukijat tietävät, että temporaaliset tietokannat ovat tulossa. Tietokantatutkijoiden yhteistyönä luotu kieli TSQL2 on kuvattu kirjassa [4] ja useissa artikkeleissa, kuten [3]. TSQL2 on otettu pohjaksi standardi SQL-kielen laajennuksessa, jonka nimi on *SQL/Temporal*. Nykytiedon mukaan laajennuksen standardi valmistuisi vuonna 2001. Kaupallisia toteutuksia ei vielä ole ja aika näyttää tuleeko niitä.

TSQL2-pohjainen tietokantajärjestelmä ratkaisisi ylhäällä esitetyt ongelmat, mutta puutteita on silti: toisaalta aikasarjakäsittelyn kannalta TSQL2-kielissä on tarpeetonta painolastia ja toisaalta joitakin ominaisuuksia puuttuu, esim. aikasarjatietueen käsite ja eri tyyppisten aikasarjojen tuki.

Aikasarjat tuotteissa

Tietokantaohjelmistojen valmistajat IBM, Informix ja Oracle ovat tuoneet tuotteisiinsa aikasarjat laajennustietotyyppin muodossa. Toteutukset on suunnattu finanssimarkkinoiden sovelluksiin. Mittausaikasarjoja voidaan niiden avulla toteuttaa, mutta käyttö on hankalaa ja tehotonta. Yhtenä haittana on tarve turvautua erityisiin funktioihin, jotka ovat erillään itse tietokantakielestä.

Teollisten valvomo-ohjelmistojen valmistajat ovat aina tukeneet mittausaikasarjoja jossakin muodossa. Tällaisia tuotteita ovat mm. Wonderwaren *InTouch*, National Instrumentsin *LabVIEW* ja Gensymin *G2*. Näillä välineillä on kuitenkin vaikea rakentaa yleiskäyt-



Kuva 1. RQL-mallin mukainen taulu.

töinen aikasarjatietokanta. Pisimmällä on Wonderware *IndustrialSQL Server* -tuotteella, jossa Microsoftin SQL Server on laajennettu aikasarjarakenteilla. Rakenteet ovat kuitenkin hyvin yksinkertaisia (tietopistetyypit ja aikasarjatietueet puuttuvat), eikä aikasarja-alkiota voi muuttaa luomisen jälkeen.

RQL-kieli ja -tietomalli

Olemme VTT Tietotekniikassa yrittäneet luoda helppotajuisen ja tehokkaasti toteutettavan aikasarjatietomallin. RapidBase-aikasarjatietokantajärjestelmässä on toteutettu RQL-kieli (RapidBase Query language) ja sen pohjalla oleva laajennettu relaatiomalli.

Kuva 1 kertoo, miten RQL-tilauksessa esitetään aikasarjoja. Taulu vastaa tietopisteen tyyppiä ja rivi yhtä tietopistettä. Tietopisteen pysyvät tiedot on esitetty tavallisilla sarakkeilla. Jokainen tietopisteeseen liittyvä aikasarja esitetään erityisellä HISTORY-tyyppisellä sarakkeella. Tällaisen sarakkeen arvona on kronologisesti järjestetty aikasarjatietueiden sekvenssi.

RQL-malli supistuu helposti tavalliseen relaatiomalliin, jos tietopisteen historia ei kiinnosta. Tavalliset SQL-komennot kohdistuvat vain aikasarjojen viimeisiin tietueisiin. Laajennetun kieliohin avulla pystytään myös helposti pureutumaan aikasarjoihin.

Kielen ilmaisuvoimasta kertoo seuraava esimerkki. Seuraavalla komenolla haetaan tietopistetaulusta tietyn tilasarakkeen arvon avulla valittujen tietopisteiden aikasarjojen arvoja viimeisen minuutin ajalta niin, että ne ovat voimassa 10 sekunnin välein olevissa aikapisteissä:

```
SELECT ... FROM table
TIMEPOINT SERIES
INTERVAL '10' SECOND
WHERE state = 'ON'
AND VALID FROM
NOW - INTERVAL '1' MINUTE
```

RQL-kielestä ja sen toteutuksesta saa lisää tietoa RapidBase-järjestelmän kotisivulta [5].

Lähteet

- [1] W. Dreyer et al. "Research Perspectives for Time Series Management Systems", *SIGMOD Record*, 23(1), March 1994, pp. 10-15.
- [2] R. Goldring. "An Introduction to Time-Series Data", *InfoDB*, 7(2), Spring 1993, pages 22 - 25
- [3] R. Snodgrass et al. "TSQL2 Language Specification", *ACM SIGMOD Record*, 23(1), March 1994, pp. 65-86.
- [4] R. Snodgrass (ed.). "The TSQL2 Temporal Query Language", Kluwer Academic Publishers, 1995, 674 p.
- [5] RapidBase-kotisivu: <http://www.vtt.fi/tte/projects/rapid>.

Antoni Wolski,
VTT Tietotekniikka,
PL 1201, 02044 VTT,
puh. (09) 456 6012,
Antoni.Wolski@vtt.fi