

## TIETOJEN SIIRTO INFOGLOVE - TIETOVARASTOON

### Johdanto

Infoglove -järjestelmän tarkoituksena on muodostaa yrityksen tietojärjestelmän tiedoista tietovarasto ja tarjota siihen käyttöliittymä raporttien ja kyselyjen tekemiseen.

Toimiakseen Infoglove tarvitsee tietojärjestelmästä ajantasaisia tietoja. Yrityksen tietojärjestelmästä käytetään tässä dokumentissa nimitystä "perusjärjestelmä", joka voi olla laskutus, ostojärjestelmä, tuotannonohjaus, kirjanpito, palkanlaskenta tai mikä tahansa strukturoitua tietoa sisältävä systeemi. Tyypillisesti tietoa siirretään perusjärjestelmästä tietovarastoon kuukauden välein, mutta siirto on mahdollista tehdä viikoittain tai jopa päivittäin. Perusjärjestelmiä voi olla useampiakin, esimerkiksi taloushallinto ja materiaalihallinto on usein eriytetty. Tällöin Infoglove tarjoaa mahdollisuuden tietojen yhdistämisen samaan raporttiin monesta lähteestä.

**Infoglove-tietovarasto luodaan asiakaskohtaisesti.** Tietosisältö siis on aina tapauskohtainen. Tässä dokumentissa paneudutaan esimerkkinä myynnin tietovarastoon. Samalla periaatteella luodaan tietovarastoja myös muihin tarkoituksiin.

**Siirtoerällä** tarkoitetaan yhdellä kertaa perusjärjestelmästä siirrettävää tietokokonaisuutta. Esimerkiksi myynnin tietovaraston siirtoerä voi sisältää tietyn aikavälin laskutapahtumat ja niissä esiintyvät nimikkeet, asiakkaat ja myyjät. Siirtoerän rakenteelliset vaatimukset ovat väljät. Näin on oltavakin, koska perusjärjestelmistä kätevästi ulos saatava data voi olla lähes missä muodossa tahansa. Vaatimuksena on kuitenkin, että tieto on **taulukkomuodossa**. Joustavin siirtomuoto on tekstitiedosto, jossa kenttäerottimena on TAB-merkki. Kyseinen muoto on yleensä helppo tuottaa ja tehokas lukea eikä sisällä rajoituksia esimerkiksi tiedostokohtaisen rivimäärän suhteen.

**Infoglove ETL** tarkoittaa Infoglove-järjestelmän tiedonsiirtotoimintoa kokonaisuutena. Infoglove ETL kykenee annetun ohjaustiedon perusteella luomaan siirtoerää vastaavan raportointitietokannan (eli tietovaraston) ja tallentamaan siirtoerän sisältämän datan siihen. Kun siirtoerän rakenne ja sisältö on tiedossa, niin asaan perehtynyt Infoglove-konsultti laatii siirtoa varten ohjausmäärittelyt. Lopputuloksena Infoglove ETL tuottaa myös raportointikehykset, joiden perusteella käyttäjät itse laativat tarvitsemansa raportit.

## Infoglove-tietovaraston ja siirtoerän rakenne

### **Millainen on Infoglove-tietovarasto**

Infoglove-tietovarasto on muodoltaan **tähtirakenteinen** (Star Schema). Rakenne on raportointikannoissa yleisesti käytetty, koska se antaa mahdollisuuden tehdä joustavasti halutun tasoisia kyselyitä. Tietokantatähden keskustassa sijaitsevat **faktatiedot**, eli numeerinen data. Esimerkiksi myynnin raportointikannassa faktaa ovat laskurivikohtainen myyntiarvo, kate ja myydyt yksiköt. Tähtien sakaroina toimivat **dimensiot**, eli luokittelevat tiedot. Myynnin raportointikannassa dimensioita voivat olla asiakas, tuote, myyjä, varasto ja myymälä.

**Päiväyskin** on dimensio, joka sisältyy melkein kaikkiin raportointikantoihin. Sen avulla faktatietoa luokitellaan päivittäin, viikoittain, kuukausittain, kvartaaleittain ja vuosittain. Päiväysdimensiona toimii faktarivin tapahtumapäivä, kuten laskurivin laskutuspäivä.

Kellonaikaa ei voi sisällyttää päiväysdimensioon, vaan sitä varten on tehtävä erillinen **kellonaika**dimensionsa. Kellonaikadimension tarve on harvinainen, yksi esimerkki on vähittäismyynnin kassapaineanalyysi.

Tietokannassa faktatiedot on organisoitu **riveiksi**, joista tyypillisesti yksi rivi vastaa perusjärjestelmän yhtä tapahtumaa, kuten laskuriviä. Jokaisella faktarivillä on kutakin dimensiota kohti kenttä, jolla rivi liittyy asianomaiseen dimensioon. Kenttään tallennetaan aina **dimension yksilöivä tunnus**, joka on esimerkiksi nimikkeellä tuotenumero ja asiakkaalla asiakasnumero.

Esimerkiksi ylläolevan kaavion faktatauluun tarvitaan seuraavat kentät: tapahtumapäivä (eli päiväysdimensio), myyntiarvo, kate, yksikkömäärä, tuotenumero, asiakasnumero, myyjänumero, myymälänumero ja varastonumero.

Dimensioille tarvitaan myös **attribuutteja**, joiden avulla tuotetaan lisäluokituksia. Esimerkiksi nimikkeen attribuutteina tarvitaan tuoteryhmä ja tuoteryhmän nimi.

### ***Siirtoerän rakennevaihtoehdot***

**Siirtoerän rakenne voi olla sama kuin Infoglove-tietovarastolla.** Tällöin faktatieto siirretään omassa tiedostossaan sisältäen päivämäärän, faktat ja dimensioiden yksilöivät tunnukset. Kunkin dimension attribuutit siirretään omassa tiedostossaan. Päivämäärä ei kuitenkaan tarvitse omaa dimensiokohtaista siirtotiedostoa.

**Siirtoerä voi olla myös vain yksi tiedosto**, jossa fakta- ja dimensiotieto attribuutteineen on koottu samalle riville. Myös sekamuodot, jossa osa dimensioiden attribuuteista on yhdistetty faktojen siirtotiedostoon, ovat mahdollisia.

### ***Dimensio vai attribuutti***

Mistä tiedetään, **onko jokin luokittelutieto itsenäinen dimensio vai dimension attribuutti?** Tämä on kannan suunnittelun kannalta erittäin tärkeä kysymys. Vastaus siihen saadaan esittämällä seuraavanlainen kysymys:

Jos tiedän nimikkeen tuotenumeron, niin tiedänpö varmuudella nimikkeen tuoteryhmän?  
Jos vastaus on kyllä, niin tuoteryhmä on attribuutti, muussa tapauksessa itsenäinen dimensio.

Yleistettynä kysymys kuuluu

**Jos tiedän dimension yksilöivän tunnuksen, niin tiedänpö sen perusteella luokittelutiedon x arvon?**

Vastauksen löytyminen ei aina ole helppoa. Eräässä yrityksessä uusittiin tuoteryhmäluokitus vuoden 2005 lopussa. Osa vanhoista tuoteryhmistä poistui ja uusia tuli tilalle. Monet nimikkeet vaihtoivat tuoteryhmää. Myyntijohto haluaa kuitenkin tarkastella tuoteryhmäkohtaista myyntihistoriaa toisinaan myös sellaisena kuin se on ollut **tapahtumahetkellä**, toisinaan taas nykyhetken tuoteryhmittelyn mukaan. Edellisessä tapauksessa tuotenumeroista ei voida päätellä tuoteryhmää, koska tuoteryhmällä on eri aikoina ollut samalla nimikkeellä erilaisia arvoja.

Tuoteryhmä on tässä **sekä itsenäinen dimensio että attribuutti**. Nykyhetken ryhmäjaon mukaan raportoitaessa se on selvästi attribuutti, tapahtumahetken jaolla itsenäinen dimensio. Ratkaisuna on tällöin tallentaa tuoteryhmä paitsi tapahtumalle itsenäisenä dimensiona myös nimikkeelle attribuuttina.

Raportointikantoja käsittelevässä kirjallisuudessa kuvattua tilannetta kutsutaan ”hitaasti muuttuvan dimension” ongelmaksi (slowly changing dimension). Infoglovessa ongelma siis ratkaistaan luomalla ajan mukana muuttuvasta dimensioattribuutista oma dimensionsa.

## Siirtoerän rajaus ja tietovaraston päivityspolitiikka

Siirtoerän tiedot jakaantuvat siis faktoihin ja dimensioihin. Jälkimmäiset ovat normaalisti perusjärjestelmän perustietoja.

Kaikkea perusjärjestelmän tietoa ei voi siirtää jokaisessa siirtoerässä, joten on tehtävä rajauksia.

**Kaikki dimensiokohtainen tieto voidaan yleensä sisällyttää jokaiseen siirtoerään.** Siirtoon tarvittava aika ei normaalisti paljon pitene, vaikka jokainen siirtoerä sisältäisi esimerkiksi yrityksen kaikki asiakkaat ja nimikkeet.

Joskus dimensiodataa on kuitenkin liikaa. Tällöin siirtoerän tiedon kokoaminen yhdeksi tiedostoksi rajaa luontevasti pois ylimääräisen dimensiotiedon. Tietenkin on mahdollista siirtää dimensiodata tällöinkin omissa tiedostoissaan rajaten sen sisältöä faktatiedoston tarpeiden mukaiseksi.

**Kaikkia faktatietoja ei tehokkuussyistä sisällytetä jokaiseen siirtoerään,** vaan siirretään esimerkiksi kuukauden tiedot kerrallaan. Usein perusjärjestelmän tiedot muuttuvat takautuvasti, jolloin on järkevää määritellä peräkkäisten siirtojen aikaväli osittain päällekkäiseksi. Siten esimerkiksi helmikuun siirtoerään voi sisällyttää uudelleen myös tammikuun tiedot.

Faktarivillä on aina päiväys, jonka perusteella Infoglove päättelee, sisältyykö se jo tietovarastoon. **Jos tietovarastossa on faktarivejä ennestään siirtoerän aikavälillä, Infoglove ETL poistaa saman aikavälin vanhan datan.** Tämä mahdollistaa paitsi osittain päällekkäiset aikavälit myös faktarivien uudelleen siirron miltä aikaväliltä tahansa.

## Tietojen esitysmuodot siirtoerässä

Infoglove kykenee lukemaan monen tyyppisiä tiedostomuotoja, kuten CSV, TAB-eroteltu tekstitiedosto tai Excel-tiedosto. Joustavin ratkaisu on muotoilla tiedot ANSI-merkistöllä koodatuiksi tekstitiedostoiksi, jossa kenttäerottimena toimii TAB-merkki. Excel-tiedostoissa saattaa tulla vastaan rivimäärärajoitus. Näkyvien merkkien käyttö tekstitiedostojen kenttäerottimenä saattaa taas aiheuttaa tulkintaongelmia niissä tapauksissa, joissa data sisältää erotinmerkkejä.

Siirtotiedoston ensimmäisenä rivinä on yleensä **otsikkorivi**, joka sisältää kenttien nimet. Otsikkorivin käyttö on luontevin tapa **dokumentoida** siirtotiedoston sisältö.

**Numerokentissä** (kenttätyyppi Fact) käytetään pilkkua tai pistettä desimaalierottimena. Etumerkki sijoitetaan luvun eteen. Tuhaterottimena saa olla pilkku, piste tai välilyönti, vaikka niiden käyttöä tulee välttää jos suinkin mahdollista. Esimerkiksi merkkijono 123.456,78 tulkitaan luvuksi 123456,78. Sama lopputulos saadaan merkkijonoista 123,456.78 ja 123 456,78 (jälkimmäisessä siis välilyönti). Sen sijaan merkkijonon 123.456 piste tulkitaan desimaalipilkuksi!

**Dimensioiden yksilöivät tunnukset** (kenttätyyppi Id) ovat merkkijonoja, joiden maksimipituus on 100 merkkiä.

**Dimensoiden attribuutit** (kenttätyyppi Attr) ovat myöskin merkkijonoja, joiden maksimipituus on 100 merkkiä.

**Päiväysdimensio** (Kenttätyyppi Date) on aina päivämäärä, jonka esitysmuoto on mieluiten [YYYY-MM-DD]. Jos tämän esitysmuodon tuottaminen ei onnistu, niin Infoglove ETL osaa tulkita myös muita yleisiä formaatteja, kuten DD.MM.YY tai DD.MM.YYYY. Faktataulussa voi olla vain yksi päiväysdimensiokenttä, joka perusteella Infoglove ETL laatii automaattisesti päiväysdimension attribuutit, jotka ovat viikko, kuukausi, kvartaali ja kalenterivuosi.

**Puuttuvat arvot** jätetään kokonaan tyhjäksi. Tällöin siirtotiedostoon siis tulee peräkkäisiä TAB-merkkejä puuttuvien arvojen kohdalle.

## Esimerkkisovellus: myynnin raportointi

### **Yleistä**

Tämä kuvaus toimii mallina myös muille vastaavanlaisille raportoinneille.

### **Faktatiedot: myyntitapahtumat**

Myyntitapahtumat ovat laskurivejä tai vähittäismyyntijärjestelmän kassan tapahtumarivejä. Myynnin hyvitystapahtumissa kaikki faktat, eli myyntiarvo, kate ja määrä ovat negatiivisia. Sekä dimensioita että faktakenttiä voi tarpeen mukaan poistaa ja lisätä aiemmin kuvattujen sääntöjen puitteissa.

Kenttä "Tieto" on tässä siirtotiedoston sarakeotsikko.

Tieto	Tyyppi	Kuvaus
TapPvm	Date	Tapahtuman päivämäärä, päiväysdimensio
Myyntiarvo	Fact	Rivin veroton myyntiarvo kotivaluutassa
Kate	Fact	Rivin veroton kate kotivaluutassa
Määrä	Fact	Myyty määrä varastoyksiköissä
Tuote	Id	Nimikkeen yksilöivä tunnus
Asiakas	Id	Asiakkaan yksilöivä tunnus
Myyjä	Id	Myyjän yksilöivä tunnus
Kust.paikka	Id	Kustannuspaikan yksilöivä tunnus
Myymä	Id	Myymän yksilöivä tunnus
Varasto	Id	Toimittavan varaston yksilöivä tunnus
Lasku	Id	Laskun tai kassakuitin yksilöivänumero

### **Tuotedimensio**

Tuotedimension tuotteen eli nimikkeen yksilöivä tunnus linkittää tuotedimension attribuutit faktatietoihin. Yleensä tämä tieto on tuotenumero, mutta mahdollisesti myös tuotenimi.

Tieto	Tyyppi	Kuvaus
Tuote	Id	Tuotteen yksilöivä tunnus
Nimi	Attr	Tuotteen nimitys
Tuoteryhmä	Attr	Tuoteryhmän yksilöivä tunnus
Tuoteryhmänimi	Attr	Tuoteryhmän nimi
Toim.tuote	Attr	Toimittajan käyttämä tuotenumero
Toimittaja	Attr	Toimittajan yksilöivä tunnus
Toimittajan nimi	Attr	Toimittajan nimi

## Asiakasdimensio

Asiakasdimensiossa asiakkaan yksilöivä tunnus linkittää dimension attribuutit faktatietoihin. Tämä tieto on tässä tapauksessa yleensä asiakasnumero.

Tieto	Tyyppi	Kuvaus
Asiakas	Id	Asiakkaan yksilöivä tunnus
Nimi	Attr	Asiakkaan nimi
As.ryhmä	Attr	Asiakasryhmän yksilöivä tunnus
As.ryhmänimi	Attr	Asiakasryhmän nimi
Maa	Attr	Asiakkaan maatunnus

## Myyjädimensio

Myyjän nimikirjaimet, numero tai koko nimi voivat toimia yksilöivänä tunnuksena.

Tieto	Tyyppi	Kuvaus
Myyjä	Id	Myyjän yksilöivä tunnus
Nimi	Attr	Myyjän nimi
Osasto	Attr	Myyjän osastotunnus

## Kustannuspaikkadimensio

Tieto	Tyyppi	Kuvaus
Kust.paikka	Id	Kustannuspaikan yksilöivä tunnus
Nimi	Attr	Kustannuspaikan nimi

## Myyvälädimensio

Tieto	Tyyppi	Kuvaus
Myyvälä	Id	Myyvälän yksilöivä tunnus
Nimi	Attr	Myyvälän nimi