

Jan Landén

Käsitteelliset tietovarastojen moniulotteisessa suunnittelussa

Tietojärjestelmätieteen

kandidaatintutkielma

31.3.2008

Jyväskylän yliopisto

Tietojenkäsittelytieteiden laitos

Jyväskylä

SISÄLLYSLUETTELO

1 JOHDANTO	3
2 TIETOVARASTO	4
2.1 Tietovarastointiprosessi.....	4
3 TIETOMALLIT	6
3.1 Mallit, kielet ja abstraktiotasot.....	7
3.2 Tietomallien luokittelua	10
3.3 Käsittemallien yleisiä vaatimuksia	12
4 TIETOVARASTOJEN MONIULOTTEINEN MALLINNUS JA KÄSITEMALLIT	13
4.1 Tietovarastojen moniulotteinen mallinnus.....	14
4.2 Tietovarastojen käsittemallit	16
4.3 Moniulotteisen mallinnuksen tietomallien vaatimuksia	18
5 MONIULOTTEISEN MALLINNUKSEN KÄSITEMALLIT.....	21
5.1 A UML profile for multidimensional modeling in data warehouses (Luján-Mora ym., 2006)	21
5.1.1 Tietomallin tavoitteita	21
5.1.2 Tietomallin elementit.....	22
5.1.3 Tietomallin ominaisuuksia.....	31
5.1.4 Tietomallin arviointia	31
6 YHTEENVETO.....	33
LÄHDELUETTELO	34

1 JOHDANTO

Tietovarasto on kokoelma teknologioita, jonka tarkoituksena on mahdollistaa tietotyöläisten parempi ja nopeampi päätöksenteko (Chaudhuri & Dayal, 1997). Tietovaraston tehtävänä on tukea monimutkaista tiedon analysointia ja siihen perustuvaa päätöksentekoa (online analytical processing, OLAP). Operatiivisten tietokantojen, jotka usein myös toimivat tietovaraston pääasiallisina tietolähteinä, tehtävänä sitä vastoin on transaktioiden tallentaminen (online transaction processing, OLTP). Sekä operatiivisten tietokantojen että tietovarastojen suunnittelussa käsitteellinen mallintaminen on tärkeää laadukkaan lopputuloksen takaamiseksi. Käsitteellisessä mallintamisessa keskitytään mallintamaan tallennettavan tiedon ominaisuuksia ottamatta kantaa varsinaiseen toteutustapaan.

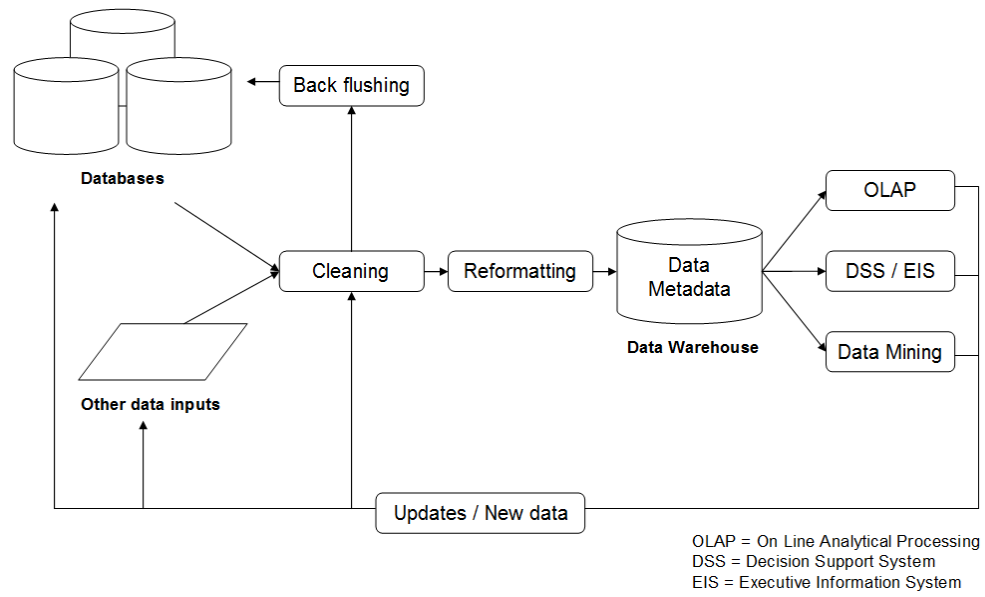
Tietovarastot suunnitellaan usein moniulotteisiksi kokoelmiksi mittari- ja dimensiotietoja. Moniulotteinen mallintaminen ja transaktiotietokantojen suunnittelu edellyttävät erilaisia suunnittelutekniikoita. Tämän tutkielman tutkimusongelmana on selvittää, minkälaisia käsitteitä moniulotteisten käsitemallien tulisi kattaa ja minkälaisia yleisiä vaatimuksia käsitemalleille on asetettu. Tutkimusongelmana on myös arvioida tarkasteltavana olevan moniulotteisen käsitemallin soveltuvuutta tietovarastojen moniulotteiseen käsitemallintamiseen. Tutkimusmenetelmä on käsitteellis-teoreettinen.

2 TIETOVARASTO

Yleisesti ajatellaan, että tietovarasto on keskeinen komponentti organisaation liiketoimintaprosessien tiedon varastoinnissa. Tietovarasto organisoii ja varastoi tietoa, jota tarvitaan informatiiviseen ja analyyttiseen tietojenkäsittelyyn (Inmon & Hackathorn, 1994). Laajalti viitatus Inmonin määritelmän mukaan tietovarasto on aihekeskeinen (Subject-Oriented), yhtenäinen (Integrated), aikasidonnainen (Time-variant) ja pysyvä (Non-volatile) kokoelma tietoa, joka mahdollistaa johdon päätöksenteon ja lähitulevaisuuden ennustamisen perustuen olemassa olevaan tietoon. Aihekeskeisyydellä tarkoitetaan, että toisiinsa liittyviä reaali maailman kohteita kuvaavat tietoelementit ovat loogisesti yhdistetty myös tietovarastossa. Yhtenäisyydellä tarkoitetaan, että tietovarasto kattaa tietoja organisaation kaikista, tai ainakin tärkeimmistä liiketoimintaprosesseista ja tietojen esitystapa on yhtenäinen. Aikasidonnaisuus ilmenee tietovarastossa siten, että tietovarasto sisältää dataa pitkältä aikaväliltä ja aika sisältyy aina jollain tavalla avainrakenteeseen. Pysyvyydellä viitataan siihen, että tietovarastoon kerran ladattuja tietoja ei myöhemmin päivitetä vaan tietojen käyttö on tietojen lukemista. (Inmon, 1996)

2.1 Tietovarastointiprosessi

Kuvassa 1. on esitetty yksinkertaistettu kuvaus tietovarastointiprosessista. Tietoa ladataan operatiivisista järjestelmistä ja muista tietolähteistä tietovarastoon. Samalla se puhdistetaan ja muokataan yhtenäiseen, analyyttistä käyttöä tukevaan muotoon. Tämän jälkeen tieto on erilaisten analyyttisten järjestelmien saatavilla. Kuvaan on piirretty myös takaisinkytkentä *OLAP*, *EIS/DSS* ja *Data Mining* järjestelmistä tietovarastoon ja tietolähteisiin. Tämä kuvaa tiedon analysoinnin tuottaman relevantin tiedon, kuten liiketoimintasääntöjen vaikutusta tietovarastointiprosessiin. (Elmasri & Navathe, 2000)

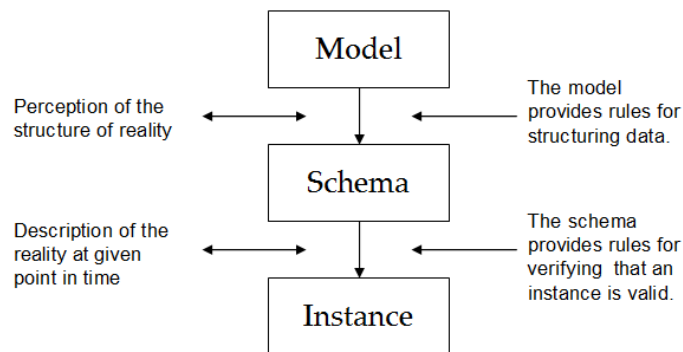


Kuva1. Tietovarastointiprosessi (Elmasri & Navathe, 2001)

3 TIETOMALLIT

Tietokannan perustavanlaatuinen ominaisuus on tarjota tietynlainen abstraktiotaso piilottaen tallennusmedian yksityiskohdat, joita suurin osa käyttäjistä ei tarvitse. *Tietomalli (data model)*, kokoelma tietokannan rakennetta kuvaavia käsitteitä, tarjoaa tarvittavat keinot tämän abstraktiotason saavuttamiseksi. (Elmasri & Navathe, 2000)

Skeema (schema) kuvaa tiettyä reaali maailman osaa käyttäen tiettyä tietomallia. Skeema on staattinen, aikariippumaton kokoelma *kielellisiä tai graafisia esityksiä*, jotka kuvaavat tiedon rakennetta. *Skeeman ilmentymä (instance of schema)* on dynaaminen, aikasidonnainen kokoelma *dataa*, joka on rakenteeltaan analoginen skeeman kanssa. (Batini ym., 1992)



Kuva 2. Mallin skeeman ja ilmentymän väliset suhteet (Batini ym, 1992)

Kuvassa 2 on esitetty tietomallin, skeeman ja ilmentymän väliset käsitteelliset suhteet. Tietomalli määrittää säännöt tiedon strukturoinnille. Skeema ilmentää pysyvää reaali maailman rakennetta ja sen ilmentymä kuvaa reaali maailman tilaa tietynä ajankohtana.

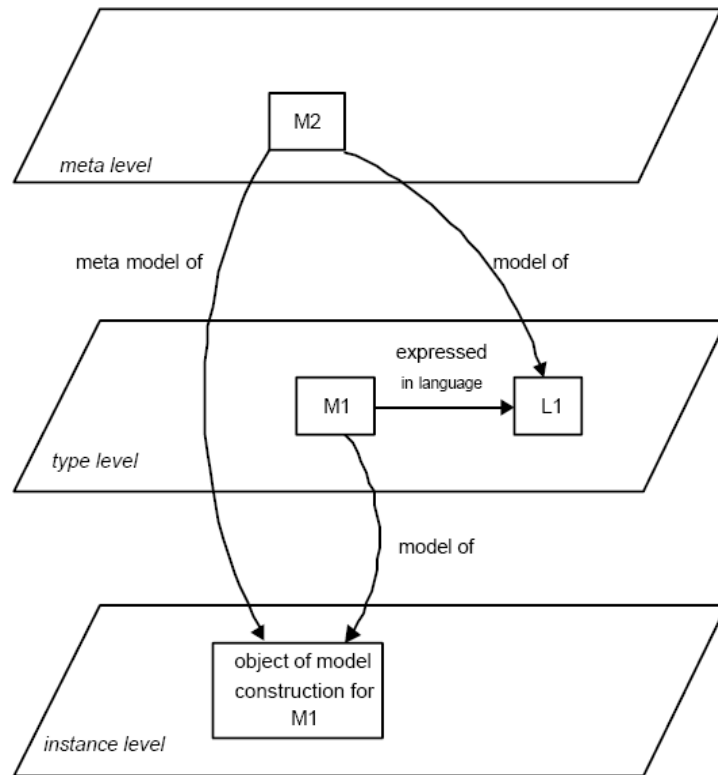
3.1 Mallit, kielet ja abstraktiotasot

Roland Holten (Holten, 2001) esittelee alkujaan tieteen filosofiasta peräisin olevan ja tietojärjestelmätieteen kirjallisuudessakin laajalti käsitellyn abstraktiotasohierarkian tietovarastokontekstissa siten, että eri abstraktiotasot kuvaavat seuraavien elementtien loogisia suhteita:

1. Liiketoiminnan johtamisen kannalta merkitykselliset liiketoimintaprosessien osat, jotka ovat kuvattu malleissa
2. Kielet, joilla malleja luodaan määriteltäessä näkymiä liiketoimintaprosesseihin
3. Kielien varsinaiset esitykset eli mallit

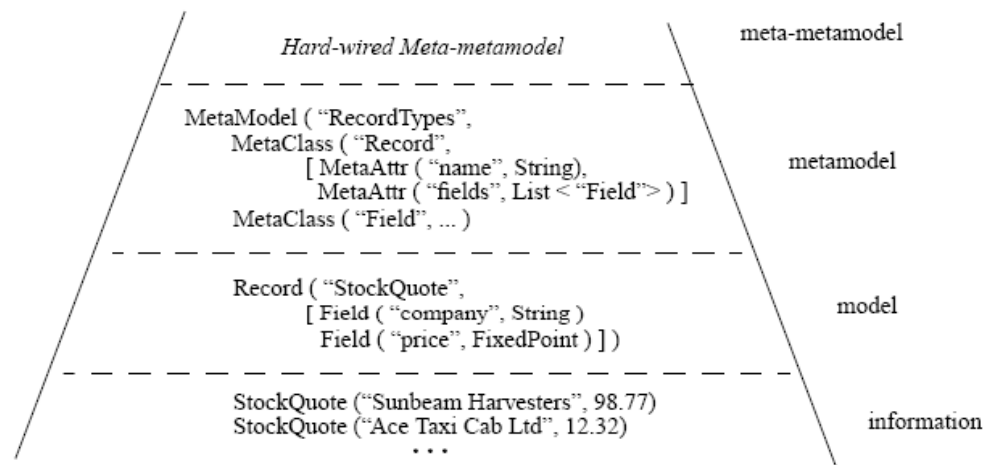
Kieltä, joka on tieteellisen tarkastelun kohteena, sanotaan *kohdekieleksi* (*object language*). Kieltä, jota tässä tarkastelussa käytetään, sanotaan *metakieleksi* (*meta language*). Erityisesti formaaleja kieliä käytetään mallien luomiseen. Kieli L1, jota käytetään mallin M1 luomiseen, voidaan kuvata mallilla M2. Näin ollen voidaan sanoa, että malli M2 on mallin M1 kuvaaman kohteen *metamalli* (*meta model*).

Kuvassa 3 on esitetty seuraavat abstraktiotasot: *ilmentymätaso* (*instance level*), *tyyppitaso* (*type level*) ja *metataso* (*meta level*). Ilmentymätasolla sijaitsevat varsinaiset liiketoimintaprosessit ja niiden osat. Tyyppitasolla on mallinnettu (M1) ilmentymätason kohteet käyttäen jotain tähän soveltuvaa kieltä (L1). Metatasolla määritellään tämän kielen käsitteet ja termit. (Holten, 2001)



Kuva 3. Mallit ja abstraktiotasot. (Holten 2001)

OMG:n (Object Management Group) MOF (Meta Object Facility) spesifikaatiossa esitetään edellisen kanssa analoginen abstraktiotasohierarkia, mutta lisätään vielä yksi, ylin taso: *Meta-metataso (meta meta level)*. Kuvassa 4. on esimerkki nelitasoisesta metadata arkkitehtuurista. (OMG, 2002)



Kuva 4. Nelitasoinen metadata arkkitehtuuri (OMG, 2002)

Nelitasoinen metadata arkkitehtuuri koostuu seuraavista tasoista:

1. *Informaatio - tasolla* sijaitsee data, jota halutaan kuvata.
2. *Malli - tasolla* on metadataa, joka kuvaa informaatio - tason datan rakennetta. Tämä metadata on koostettu malliksi.
3. *Metamalli - tasolla* määritellään metadatan rakenne ja semantiikka. Tämä *meta-metadata* on koostettu meta-malliksi.
4. *Meta-metamalli - tasolla* kuvataan metamalli - tason rakenne ja semantiikka.

Nelitasoisen metadata arkkitehtuurin tarkoituksena on, että yhdellä meta-metamallilla voidaan kuvata useita meta-malleja. Kuten kuvan 4. esimerkin StockQuote - tietuetyypillä (määritelty malli-tasolla) voidaan kuvata useita informaatiotason ilmentymiä, RecordTypes - meta-mallilla (määritelty metamalli - tasolla) voidaan kuvata useita tietuetyyppejä. Samalla tavoin meta-metamalli - tasolla voidaan kuvata useita metamalleja, jotka puolestaan kuvaavat erityyppistä informaation metadataa (OMG, 2002).

3.2 Tietomallien luokittelua

Elmasri ja Navathe jakavat tietomallit niiden sisältämien tietokannan rakennetta kuvaavien käsitteiden tyyppin mukaan kolmeen luokkaan: *Käsittellinen tietomalli (conceptual data model / high-level)* sisältää käsitteitä, jotka ovat lähellä useimpien käyttäjien tapaa hahmottaa tallennettavaa tietoa. *Fyysinen tietomalli (physical data model / low-level)* sisältää käsitteitä, jotka kuvaavat tiedon fyysisestä tallennustapaa tietokoneelle. Näiden kahden ääripään väliin jää *kuvaava tietomalli (representative data model / intermediate-level)*, jonka käsitteet ovat loppukäyttäjän ymmärrettävissä, mutta eivät ole liian kaukana tavasta, jonka mukaan tiedot ovat organisoitu tietokoneelle. (Elmasri & Navathe, 2000)

Batini, Ceri ja Navathe jakavat tietomallit niin ikään kolmeen tasoon niiden sisältämien rakenteiden ja käsitteiden mukaan, sekä sen mukaan, mihin tietokannan suunnitteluvaiheeseen ne parhaiten sopivat: *Käsitteellisellä tasolla (Conceptual level)* tietomallit ovat lähellä loppukäyttäjää ja toteutustapariippumattomia. *Loogisen tason (Logical level)* tietomallit ovat riippuvaisia valitun tietokannanhallintajärjestelmän tyypistä, mutta silti loppukäyttäjän ymmärrettävissä. *Fyysisen tason (Physical level)* tietomallit ovat tietokannanhallintajärjestelmäkohtaisia ja kuvaavat kuinka data on varsinaisesti tallennettu tietokoneelle. (Batini ym, 1992)

Taulukossa 1. on kuvattu yhteenvedona tietomallien jako käsitteellisiin, loogisiin ja fyysisiin malleihin sekä näiden eri abstraktiotasot. Tämän tutkielman aiheena olevat käsittemallit sijoittuvat taulukossa merkittyyn kohtaan.

	meta-metamalli	metamalli	malli	ilmentymä
käsitteellinen		X		
looginen				
fyysinen				

Taulukko 1. Tietomallien luokittelu ja abstraktiotasot.

3.3 Käsittemallien yleisiä vaatimuksia

Batini, Ceri ja Navathe (1992) esittävät neljä laatuvaatimusta käsitteellisille tietomalleille:

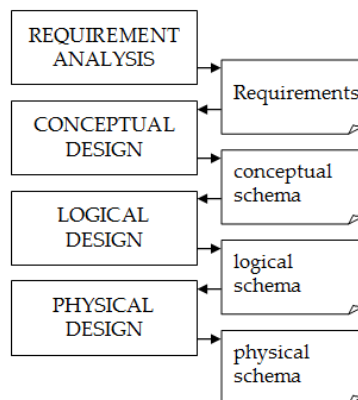
1. *Ilmaisuvoimaisuus (expressiveness)*. Tietomallin sisältämä erilaisten käsitteiden laaja kirjo mahdollistaa reaali maailman kokonaisvaltaisen mallintamisen. Näin ollen tietomallit, joilla on rikas käsitteistö, ovat myös hyvin ilmaisuvoimaisia.
2. *Yksinkertaisuus (simplicity)*. Käsitteellisen tietomallin tulee olla yksinkertainen, jotta tietomallin mukaan mallinnettu skeema olisi helposti kehittäjien ja loppukäyttäjien ymmärrettävissä. Ilmaisuvoimaisuus ja yksinkertaisuus ovat keskenään ristiriitaisia vaatimuksia.
3. *Minimaalisuus (minimality)*. Tämä vaatimus saavutetaan kun jokaisella tietomallin käsitteellä on erillinen tarkoitus suhteessa muihin tietomallin käsitteisiin. Toisin sanoen mitään tietomallin käsitettä ei voida esittää muiden käsitteiden koosteella.
4. *Formaalius (formality)*. Formaalius edellyttää, että jokaisella mallin käsitteellä on erillinen, tarkka ja hyvin määritelty selitys. Formaaleja käsitteitä voidaan käsitellä matemaattisesti.

Tietomallin graafiselle notaatiolle esitetään seuraavat kaksi vaatimusta:

1. *Graafinen täydellisyys (graphic completeness)*. Tietomalli on graafisesti täydellinen, jos sen jokaisella käsitteellä on oma graafinen esitystapa. Muutoin graafista esitystä tulee täydentää kirjallisesti.
2. *Helppolukuisuus (ease of reading)*. Malli on helppolukuinen, jos sen jokainen käsite on esitetty selvästi toisistaan eroavilla symboleilla.

4 TIETOVARASTOJEN MONIULOTTEINEN MALLINNUS JA KÄSITEMALLIT

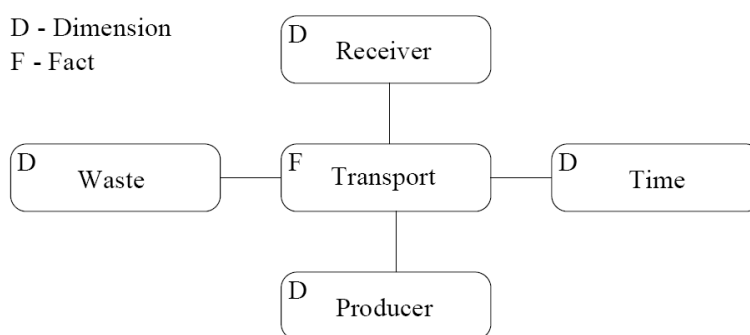
Tietovarastot suunnitellaan usein moniulotteisiksi kokoelmiksi mittari- ja dimensiotietoja. Moniulotteinen mallintaminen ja transaktiotietokantojen suunnittelu edellyttävät erilaisia suunnittelutekniikoita. Paljon on kirjoitettu siitä, kuinka tietovarasto tulisi mallintaa, mutta suunnittelumenetelmästä ei olla yhtä mieltä. Suurin osa menetelmistä kuitenkin hyväksyy käsitteellisen- ja loogisen suunnitteluvaiheen erottamisesta syntyvän hyödyn mahdollisuuden. Käsitteellinen mallintaminen pyrkii rakentamaan toteutustapariippumattoman käsitteellisen skeeman tietovarastolle, menetelmästä riippuen, perustuen joko käyttäjävaatimukseen tai lähdejärjestelmien rakenteeseen. Loogisessa suunnittelussa rakennetaan vastaava looginen skeema valittuun ympäristöön ottaen huomioon tiettyjä rajoitteita kuten tallennusmedian rajallisuus ja kyselyjen vasteaika. Jotkin menetelmät sisältävät oman vaiheensa tietovaraston fyysiseen suunnitteluun ja vaatimusmäärittelyyn. Kuvassa 5 on esitetty tietovarastojen suunnitteluvaiheet ja niiden väliset suhteet toiminnallisen näkökulman mukaan. Tosielämässä prosessi sisältäisi takaisinkytkentöjä antaen mahdollisuuden palata edeltäviin vaiheisiin. (Rizzi ym., 2006)



Kuva 5. Tietovaraston suunnittelun päävaiheet (Rizzi ym., 2006)

4.1 Tietovarastojen moniulotteinen mallinnus

Moniulotteisesti mallinnetun tietovaraston pääasiallisina etuina ovat sen mahdollistamien tehokkaiden tietokantakyselyiden lisäksi sen analoginen rakenne analyytikoiden ajattelutavan kanssa. Moniulotteinen mallinnus perustuu moniulotteisen kuution käsitteeseen. Moniulotteinen kuutio koostuu *faktasta* ja joukosta siihen liitettyjä *dimensioita*. Fakta kuvaa analyysin kohteena olevaa dataa ja dimensiot kuvaavat sen eri analysointinäkökulmia. Fakta koostuu edelleen *tietosoluista* (*fact cells*), jotka voidaan identifioida kuution dimensioiden *tasojen* (*dimension levels*) suhteen. Kuvassa 6 on esitetty esimerkki neliulotteisesta jätteenkuljetusliiketoiminnan analysointikuutiosta. Kuutio koostuu kuljetus - faktasta ja neljästä siihen liitetystä dimensiosta: kuljetettava jäte, jätteen tuottaja, jätteen vastaanottaja ja tapahtuma-aika. (Abelló ym., 2001) (Abelló ym., 2002)

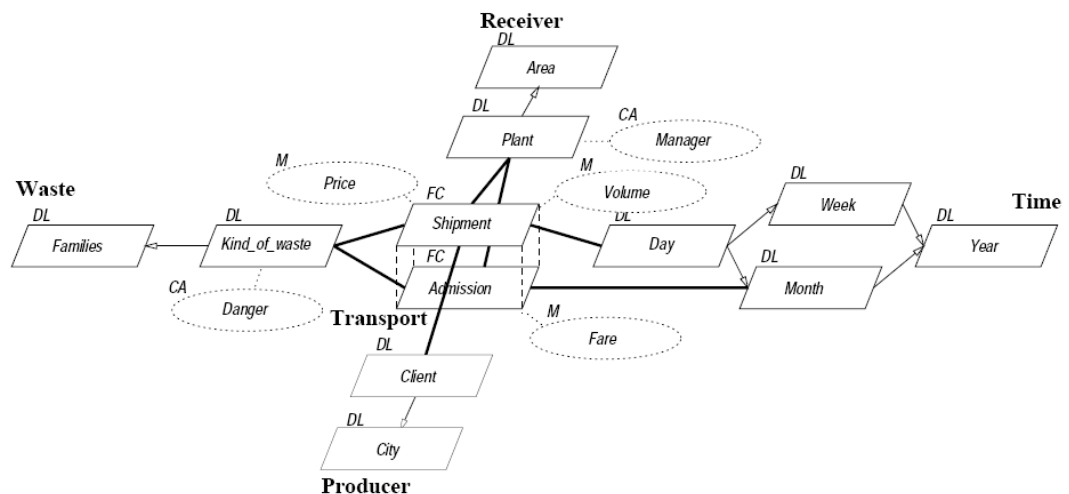


Kuva 6. Esimerkki moniulotteisesta skeemasta (Abelló ym., 2001)

Abelló ym. (2001) jakavat moniulotteisen mallin kolmeen tarkkuustasoon:

1. *Ylätasolla (Upper Level, UL)* on olemassa *dimensioita (dimension, D)* ja *faktoja (fact, F)*. Dimensiot määrittävät faktojen analysointinäkökulmat. Liittämällä faktaan joukko dimensioita luodaan tähtimallin mukainen rakenne. Navigointi tähtimallista toiseen on mahdollista näiden jakamien dimensioiden kautta.
2. *Välitasolla (Intermediate Level, IL)* dimensiot ja faktat ovat jaettu dimension tasoiksi (*dimension level, DL*) ja tietosoluiksi (*fact cell, FC*). Dimensioiden tasot muodostavat hierarkioita. Jokainen faktan tietosolu sisältää dataa, joukon samalla tarkkuustasolla (*granularity*) olevia mittareita, jokaista tähän liitettyä dimension tasoa kohti.
3. *Alatasolla (Lower Level, LL)* kuvataan dimension tasojen ja tietosolujen attribuutit eli *luokitteluattribuutit (Classification Attributes, CA)* ja *mittarit (Measures, M)*.

Kuva 7 esittää saman moniulotteisen mallin tarkemmalla tarkkuustasolla, kuin kuva 6. Jokainen dimensio on jaettu hierarkioihin ja fakta on jaettu tietosoluihin. Kuvaan on piirretty myös luokitteluattribuutit ja mittarit.



Kuva 7. Esimerkki tarkennetusta moniulotteisesta skeemasta (Abello ym., 2001)

4.2 Tietovarastojen käsitelmallit

Käsitelmalleilla pyritään mallintamaan tietovarasto tai sen osa ottamatta kantaa varsinaiseen toteutustapaan. Käsitelmallit ovat laajasti hyväksytyt tärkeäksi perustaksi suunnitellessa hyvin dokumentoituja ja asiakkaiden tarpeet toteuttavia tietokantoja. Yleensä käsitelmallit perustuvat graafiseen notaatioon, joka tukee käsitteellisten määritysten kirjoittamista, lukemista ja niiden hallintaa sekä kehittäjien että käyttäjien näkökulmista. Kirjallisuudessa käsitelmallinnus on tähän mennessä keskittynyt lähinnä tietokannan moniulotteiseen mallintamiseen ja ETL - prosessin (extract, transform, load) mallintamiseen. ETL - prosessin mallintamista on tutkittu vielä selvästi vähemmän kuin moniulotteista mallinnusta, mutta tutkimus voi tulla vahvasti vaikuttamaan tietovarastojen suunnitteluprosessin luotettavuuteen ja kestoan. (Rizzi ym., 2006)

ER - mallia (entity/relationship) on laajasti käytetty relaatiokantojen käsitteellisenä formalismina. Moniulotteisen mallintamisen näkökulmasta ongelmana kuitenkin on, että malli soveltuu paremmin kuvaamaan tietojen välisiä suhteita, kuin näiden synteisiä (Kimbal, 1996). Vaikka ER - malli onkin ilmaisuvoimaltaan riittävä kuvaamaan keskeisimmät tietovaraston käsitteet, ei se perusmuodossaan, kykene riittävästi korostamaan moniulotteisen mallinnuksen avainkäsitteitä. (Rizzi ym., 2007)

Monet suunnittelijat painottavat *tähtimallin* (*star schemata*), joka on pääasiainen moniulotteisen mallin toteutustapa relaatiokannoissa, käyttöä käsitteellisessä suunnittelussa. Tähtimalli on kuitenkin vain denormalisoitu relaationaalinen malli, joka määrittelee joukon relaatioita ja eheysvaatimuksia. Käsitteellinen mallintaminen ilman staattisen, funktionaalisen ja dynaamisen mallin tukea johtaa usein huonoon lopputulokseen. (Rizzi, 2007)

Taulukkoon 2 on koottu erityisesti tietovarastojen moniulotteiseen mallintamiseen esitetyt käsitelmallit. Tietomallit voidaan jakaa karkeasti kolmeen osaan: ER - mallien laajennukset, UML - mallien laajennukset ja tapauskohtaiset mallit. Taulukosta ilmenee myös, mikäli tietomallin yhteydessä on esitetty jokin käsitteellisen mallintamisen metodologia.

	E/R extension	object-oriented	ad hoc
no method	Franconi and Kamble (2004); Sapia et al. (1998); Tryfona et al. (1999)	Abelló et al. (2002); Nguyen, Tjoa and Wagner (2000)	Tsois et al. (2001)
method		Luján-Mora et al. (2002)	Golfarelli et al. (1998); Hüsemann et al. (2000)

Taulukko 2. Käsitteelliset tietomallit (Rizzi, 2007).

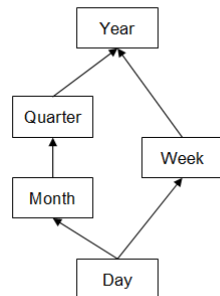
4.3 Moniulotteisen mallinnuksen tietomallien vaatimuksia

Tietovarastojen moniulotteisille käsitelmille on esitetty suuri joukko vaatimuksia. Useimmat vaatimukset liittyvät tietomallien käsitteistöön ja näiden ilmaisuvoimaan. Pedersen (2000) esittää yksitoista vaatimusta kompleksisen tietovaraston moniulotteiselle tietomallille. Kirjoittaja esittää väitöskirjassansa myös tietomallin, joka täyttää moniulotteisille tietomalleille asettamansa vaatimukset:

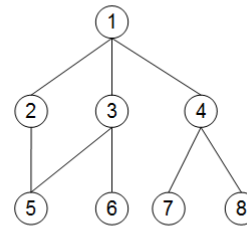
1. *Eksplisiittiset dimensioiden hierarkiat.* Dimensioiden tasohierarkiat tulisi voida kuvata eksplisiittisesti skeemassa. Dimensiohierarkian voidaan ajatella olevan suunnattu asyklinen verkko (*directed acyclic graph*), jossa solmut kuvaavat dimension tasoja ja välit kuvaavat kohdetason ilmentymien koostumista lähdetason ilmentymistä. Kuvan 8a aikadimensiossa on viisi tasoa (*dimension level*).
2. *Symmetrinen dimensioiden ja mittareiden käsittely.* Tietomallin tulisi mahdollistaa mittareiden käsittelyn kuin ne olisivat dimensioita ja päinvastoin. Esimerkiksi jos skeemassa on määritelty asiakkaan ikä mittaritiedoksi, tulisi olla mahdollista myös määritellä dimensio, joka luokittelee asiakkaat eri ikäryhmiin.
3. *Dimensioiden vaihtoehtoiset hierarkiat.* Yksittäisen dimension tulee voida sisältää useita, vaihtoehtoisia tasohierarkioita. Esimerkiksi kuvan 8a aikadimensiossa on kaksi vaihtoehtoista tasohierarkiaa: Day < Month < Quarter < Year ja Day < Week < Year.
4. *Tuki koostefunktioiden semantiikalle.* Mittareiden merkitykselliset koostefunktiot tulisi voida määritellä eksplisiittisesti skeemassa ja estää näin niiden virheellinen tulkinta. Esimerkiksi yksikköhinta - mittaria ei voida mielekkäästi summata minkään dimension suhteen, mutta sen keskiarvo-, minimi- ja maksimi-koostefunktiot voivat olla merkityksellisiä.
5. *Ei-tiukat (non-strict) hierarkiat.* Dimensioiden hierarkiat eivät aina ole tiukat. Toisin sanoen dimension tasojen välillä voi olla monesta moneen suhteita. Esimerkiksi tuotedimension alimmalla tasolla oleva yksittäinen tuote voi kuulua useampaan tuoteryhmään. Kuvassa 8b solmu 5 on liitetty ylemmän tason solmuihin 2 ja 3. Tietomallin tulee voida käsitellä ei-tiukkoja hierarkioita siinä missä tiukkojakin.
6. *Tasapainottamattomat (non-onto) hierarkiat.* Usein dimensioiden hierarkiat eivät ole tasapainotettuja. Tasapainottamattomassa hierarkiassa etäisyys dimensiohierarkian juurisolmusta lehtisolmuun vaihtelee. Tietomallin tulee

voida käsitellä myös tasapainottamattomia hierarkioita. Kuvassa 8c juurisolmun 1 ja lehtisolmujen 2, 5, 6 ja 7 etäisyys vaihtelee.

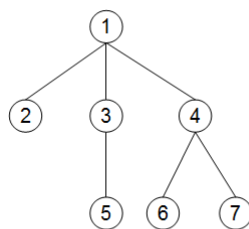
7. *Ei-kattavat (non-covering) hierarkiat.* Usein myös solmujen väliset linkit hyppäävät yhden tai useamman tason yli dimensiohierarkiassa. Kuvassa 8d solmu 4 on liitetty suoraan juurisolmuun 1.
8. *Dimensioiden ja faktojen monesta-moneen (many-to-many) suhteet.* Faktan ja dimension suhde ei ole aina perinteinen monesta-yhteen. Tietomallin tulee sallia myös monesta-moneen suhteet faktan ja dimension välillä.
9. *Muutoksen ja ajan hallinta.* Vaikka dimensiotieto muuttuu ajan myötä, tulisi olla mahdollista suorittaa mielekkäitä analyysejä aikaa vasten. Tämä vaatimus liittyy hitaasti muuttuvien dimensioiden käsitteeseen (slowly changing dimensions).
10. *Eri tarkkuustasojen hallinta.* Tietomallin tulisi mahdollistaa mittaritiedon tallentaminen eri tarkkuustasolla. Vaatimus on osittain yhteydessä kohdan 6. vaatimuksen kanssa (tasapainottamattomat hierarkiat).
11. *Epätarkkuuden hallinta.* Kun mittaritietoa on tallennettu eri tarkkuustasolla, on tietomallin tuettava epätarkkuutta sisältäviä kyselyjä.



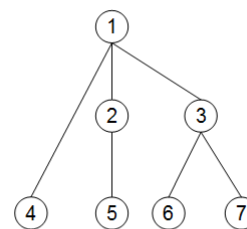
a) aikadimension tasohierarkia



b) ei-tiukka ("non-strict") hierarkia



c) tasapainottaamaton ("non-onto") hierarkia



d) ei-kattava ("non-covering") hierarkia

Kuva 8. Dimensiohierarkiat.

Abelló ym. (2001) tarkastelevat moniulotteisen mallinnuksen tietomalleja luokittelmalle ne neljään kategoriaan pääasiallisen käyttökohteensa mukaan (käsitteelliset -, loogiset -, fyysiset - ja formaalit tietomallit) ja tutkimalla näiden ilmaisuvoimaa eri tarkkuustasoilla vastaamalla seuraaviin kysymyksiin:

Alataso (LL)	
Mittari (M)	Voidaanko skeemassa esittää ylipäätään mittareita?
Luokitteluattribuutti (CA)	Voiko dimensiolla olla attribuutteja vai sisältyykö kaikki informaatio korkeintaan dimension hierarkiaan?
Suhteet alatasolla	Onko skeemassa mahdollisuutta esittää mittareiden ja / tai luokitteluattribuuttien välisiä suhteita?
Välitaso (IL)	
Taso (DL)	Onko dimensioihin mahdollista mallintaa eksplisiittisiä dimension tasoja?
Tietosolu (FC)	Voidaanko mittarit ryhmitellä tietosoluihin ja nämä liittää niitä vastaaviin dimension tasoihin (eikä esimerkiksi vain dimensioiden alimmalle tasolle)?
Suhteet välitasolla	Onko skeemassa mitään mahdollisuutta kuvata tietosolujen ja / tai dimension tasojen välisiä suhteita? (esim. eksplisiittisiä dimensiohierarkioita)
Ylätaso (UL)	
Fakta (F)	Voidaanko tietosoluja yhdistää, vaikka ne olisivatkin eri tarkkuustasolla, mikäli niitä käytetään samassa päätöksentekoprosessissa?
Dimensio (D)	Voidaanko mallintaa useampia dimension tasoja ja liittää niitä skeeman muihin elementteihin.
Suhteet ylätasolla	Onko skeemassa mitään mahdollisuutta kuvata faktojen ja / tai dimensioiden välisiä suhteita?

Taulukko 1. Moniulotteisten tietomallien ilmaisuvoima eri tarkkuustasolla (Abelló ym., 2001)

5 MONIULOTTEISEN MALLINNUKSEN KÄSITEMALLIT

Tässä kappaleessa esitetään UML:ään perustuva tietokantojen moniulotteiseen mallinnukseen esitetty tietomalli, sen keskeisimmät käsitteet ja näiden graafiset esitykset. Tämän jälkeen tietomallia verrataan käsitemallien yleisiin vaatimuksiin sekä tarkastellaan tietomallin ilmaisuvoimaa moniulotteisille tietomalleille esitettyjen vaatimusten valossa.

5.1 A UML profile for multidimensional modeling in data warehouses (Luján-Mora ym., 2006)

Luján-Mora, Trujillo ja Song (2006) esittelevät käsitteellisen, oliosuuntautuneen tietomallin tietovarastojen moniulotteiseen mallintamiseen. Tietomalli perustuu heidän aiemmin esittämiinsä tietomalleihinsa (Trujillo ym., 2001) ja (Luján-Mora ym., 2002). Tietomalli laajentaa UML - kieltä (Unified Modeling Language) tarkoituksenaan paremmin tukea moniulotteisen mallinnuksen keskeisiä rakenteellisia ja dynaamisia käsitteitä. Laajennukset ovat toteutettu määrittäen joukko stereotyyppejä (*stereotype*), rajoitteita (*constraints*) ja arvoja (*tagged values*), UML:n omia tietomallin laajentamiseen tarkoitettuja elementtejä. Tietomallin perustaksi on valittu UML - kieli kahdesta syystä: 1) Tietomallin perustuessa laajalti käytettyyn standardiin mallinnuskieleen, käyttäjän ei tarvitse opetella erityistä notaatiota tietovarastojen moniulotteiseen mallinnukseen. 2) UML on helposti laajennettava mallinnuskieli ja tämä on mahdollista räätälöidä kattamaan tietyn aihealueen erityispiirteet, kuten esimerkiksi tietovarastojen moniulotteisen mallinnuksen.

5.1.1 Tietomallin tavoitteita

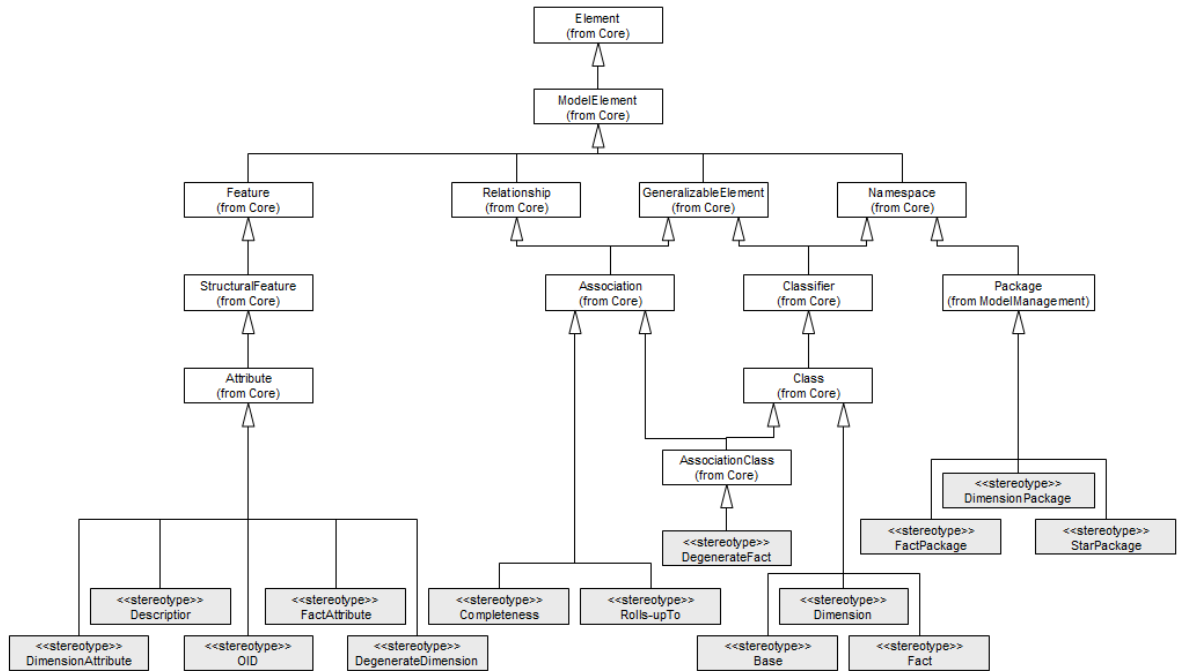
Luján-Mora ym. (2006) pyrkivät esittämään tietomallin, joka olisi *tarkka* (*accurate*), *toistamaton* (*non-redundant*), *yhtenäinen* (*consistent*), *yksinkertainen*

(*simple*) ja *ymmärrettävä* (*understandable*). Tarkkuudella tarkoitetaan, että tietomalli mahdollistaa tärkeimpien moniulotteisten ominaisuuksien kuvaamisen käsitteellisellä tasolla. Toistamattomuudella estetään skeemassa esitetyn käsitteen rinnakkaiset määrittelyt. Yhtenäisyydellä kirjoittajat viittaavat tietomallin määrittelemien elementtien olevan erillisiä ja kohdealueen kattavia. Yksinkertaisuuteen ja ymmärrettävyyteen pyritään jakamalla skeema eri tarkkuustasoihin hyödyntäen UML:n ryhmittelymekanismeja sekä rajoittamalla tietomallin graafiset elementit UML:n osajoukkoon.

5.1.2 Tietomallin elementit

Tietomalli laajentaa UML - kieltä määrittäen kielen profiilin (*profile*). Laajennus on toteutettu käyttäen UML:n omia laajennusmekanismeja (UML Extensibility Mechanism - package). Profiili on yhtenäinen kokoelma stereotyyppisiä, näiden rajoitteita ja arvoja sekä joukko tietomallin tarkoituksenmukaisen käytön takaavia sääntöjä (*well-formedness rules*) ja muita kuvauksia.

Stereotyyppi on UML:n elementti, joka määrittelee arvoja, rajoitteita ja mahdollisesti graafisen esitysmuodon erikoistettavalle elementille. Stereotyyppi mahdollistaa siis uuden semanttisen merkityksen liittämisen elementille. Tietomalli määrittelee yhteensä 14 UML - elementtien stereotyyppiä. Kuvassa 9. esitetään osa UML:n metamallista. Profiilin määrittelemät stereotyyppit ovat merkitty harmaalla taustavärillä.



Kuva 9. UML laajennus stereotyypein (Luján-Mora ym., 2006)

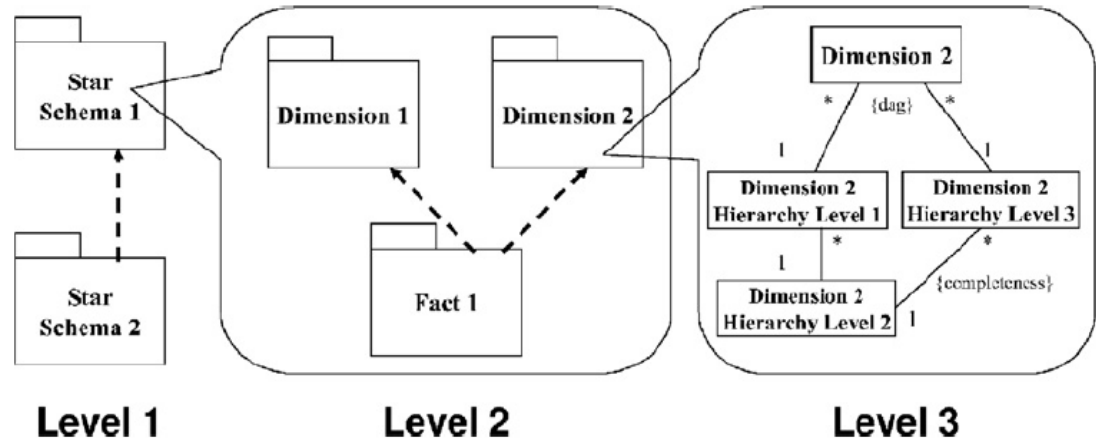
Profiilissa on määritelty kolme Package-elementin stereotyyppiä: StarPackage, DimensionPackage ja FactPackage. Näiden elementtien tarkoituksena on jakaa mallin skeema pienempiin osakokonaisuuksiin ja mahdollistaa laajojen ja kompleksisten järjestelmien mallintaminen. Skeema voidaan jakaa kolmeen tarkkuustasoon (Kuva 10):

1. Mallin määrittely (*model definition*). Tasolla esitetään joukko StarPackage-elementtejä ja näiden välisiä riippuvuuksia (*dependency*). StarPackage-elementti kuvaa käsitteellisen moniulotteisen mallin tähtimallia.¹ Riippuvuus kahden StarPackage-elementin välillä indikoi, että tähtimallit jakavat ainakin yhden dimension. Taso mahdollistaa näin vahvistettujen dimensioiden (*conformed dimensions*) tarkastelun.
2. Tähtimallin määrittely (*star schema definition*). Tasolla esitetään tähtimallin fakta (FactPackage) ja tähän liitetyt dimensiot (DimensionPackage).

¹ Tähtimalli-käsitteellä ei tässä yhteydessä tarkoiteta relaationaalista toteutusta, vaan viitataan faktaan ja tähän liitettyjen dimensioiden muodostamaan käsitteelliseen tähtimalliin.

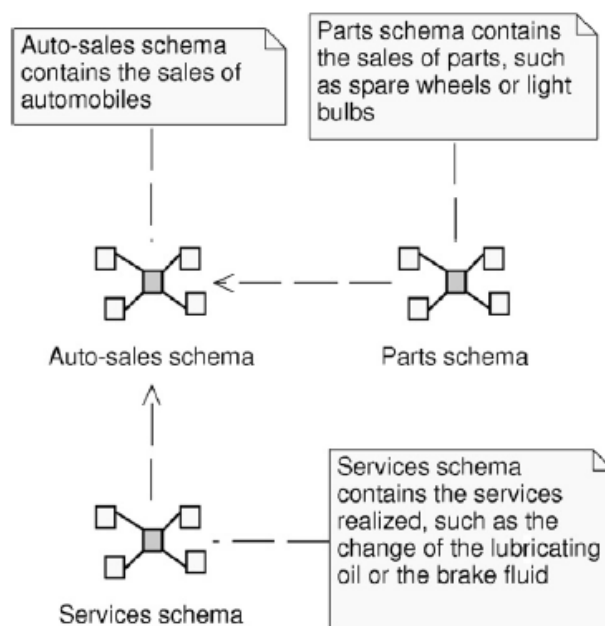
Riippuvuus kahden dimension välillä indikoi, että dimensiot jakavat ainakin yhden dimension tasohierarkian tason.

3. Dimension / faktan määrittely (*dimension/fact defnition*). Tasolla esitetään joko dimension tasohierarkia tai FactPackagen ollessa kyseessä, koko tähtimallin kuvaus.



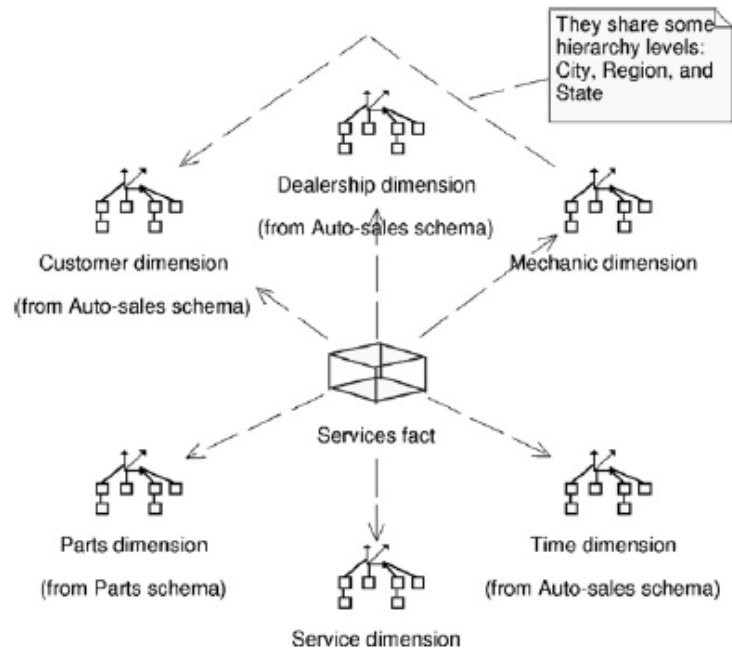
Kuva 10. Kolme moniulotteisen mallin tarkkuustasoa käyttäen Package-elementtejä. (Luján-Mora ym., 2006)

Tasolla 1 kuvataan mallin sisältämät tähtimallit ja näiden väliset riippuvuudet. Tähtimallit kuvataan StarPackage-elementein. Kuvassa 11 on esimerkki skeemasta, joka sisältää kolme tähtimallia: Auto-sales schema, Parts schema ja Services schema. Riippuvuus-assosiaatiot tähtimallien välillä kuvaavat, että Services schema ja Parts schema jakavat ainakin yhden Auto-sales schemassa määritellyn dimension.



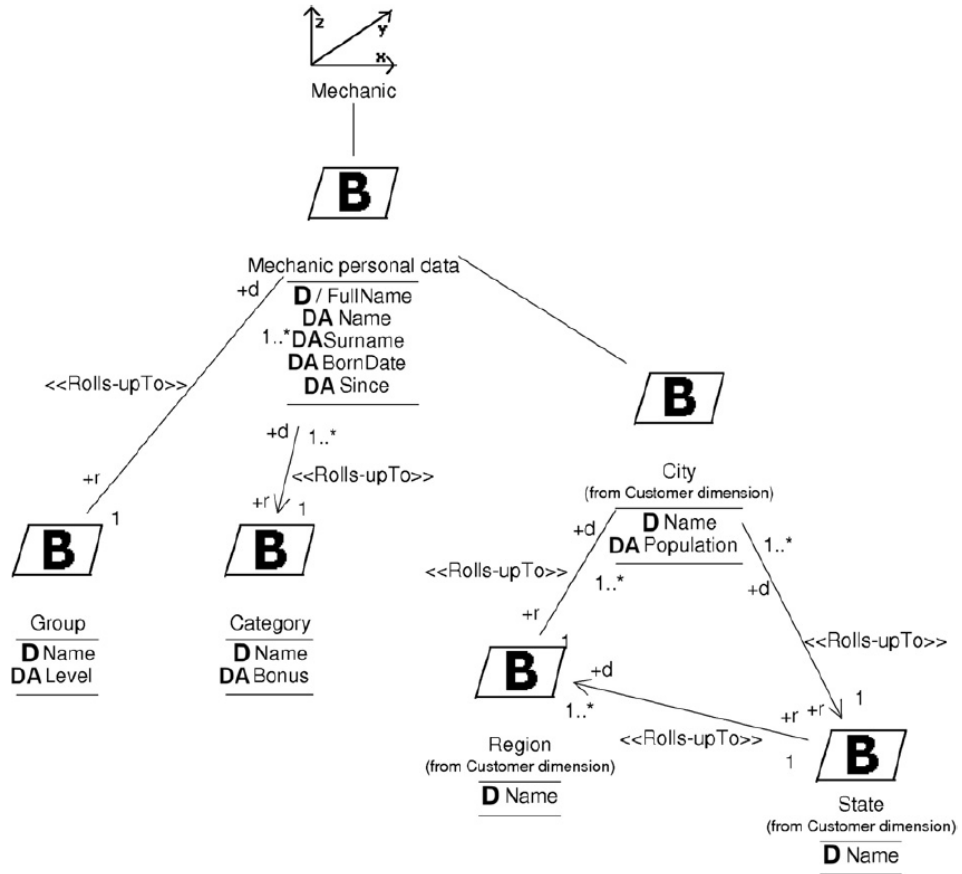
Kuva 11. Taso 1: mallin määrittely. (Luján-Mora ym., 2006)

Tasolla 2 esitetään tähtimallin fakta ja tähän liitetyt dimensiot. Fakta kuvataan skeemassa FactPackage-elementillä ja dimensiot DimensionPackage-elementein. Kuvassa 12 on esimerkki tason 2 kaaviosta, joka on tarkennettu kuvaus kuvan 11 Services schemasta. Riippuvuus-assosiaatiot dimensioiden välillä indikoivat, että nämä jakavat ainakin yhden dimension tason.

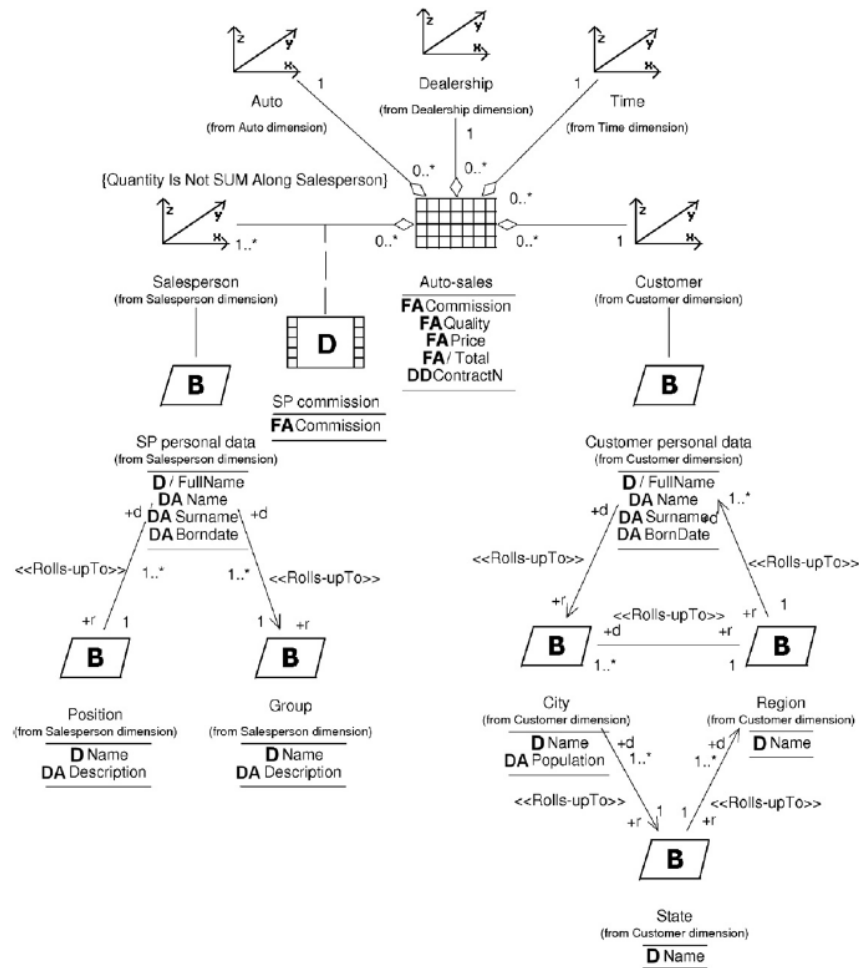


Kuva 12. Taso 2: tähtimallin määrittely. (Luján-Mora ym., 2006)

Tasolla 3 esitetään tarkennettu kuvaus dimensiosta tai faktan ollessa kyseessä, koko tähtimallista. Dimensio kuvataan tietomallissa Dimensio-luokkana (*Dimension class*) ja dimension tasot Taso-luokkina (*Base class*). Fakta kuvataan tietomallissa Fakta-luokkana (*Fact class*). Kuvassa 13 esitetään Mechanic - dimension tasohierarkia ja kuvassa 14 Auto-sales - faktan tähtimalli.



Kuva 13. Taso 3: dimension määrittely. (Luján-Mora ym., 2006)



Kuva 14. Taso 3: faktan määrittely. (Luján-Mora ym., 2006)

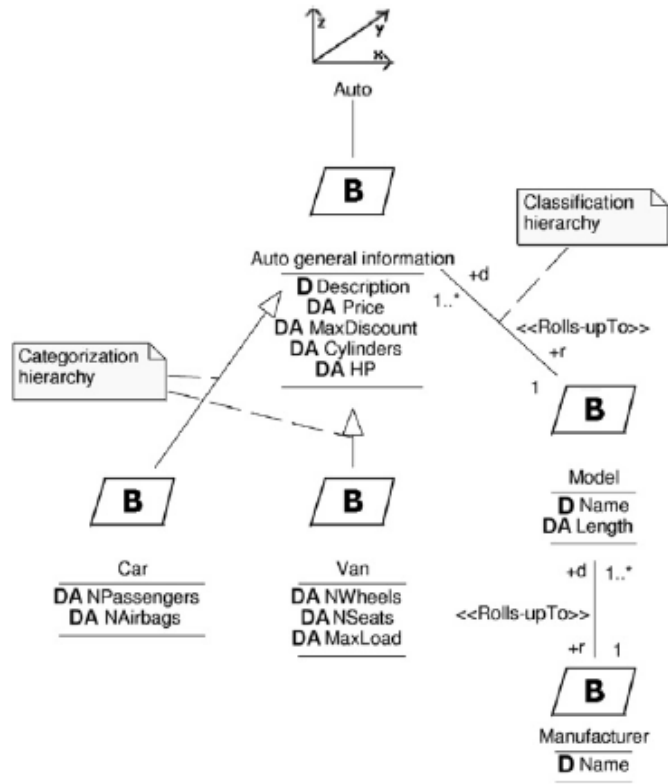
Luokkien attribuutit ovat Fakta-attribuutteja (*FactAttribute*) ja dimensio-attribuutteja (*DimensionAttribute*). Fakta-attribuutit kuvaavat faktan mittareita ja dimensio-attribuutit dimension tason attribuutteja. Tietomalli määrittelee vielä kolme attribuutin stereotyyppiä: *OID*-attribuutti, *Descriptor*-attribuutti ja *DegenerateDimension*-attribuutti. *OID*-attribuutti kuvaa luokkien ilmentymät yksiselitteisesti identifioivat attribuutit. *Descriptor*-attribuutilla määritellään taso-luokan ilmentymien selitteenä käytettävä attribuutti. *DegenerateDimension*-attribuutti ilmentää faktan attribuuttia, joka toimii luokitteluattribuuttina ilman siihen liitettyä dimensiota.

Dimension luokitteluhierarkia kuvataan skeemassa joukolla Taso-luokkia, jotka ovat liitetty toisiinsa käyttäen joko Rolls-upTo- tai Completeness-assosiaatiota. Rolls-upTo-assosiaatio kuvaa kahden tason välistä suhdetta jossa korkeamman luokan (merkitty roolilla *R*, *rolls-up*) jäsenet koostuvat alemman luokan (merkitty roolilla *D*, *drills-down*) jäsenistä. Completeness-assosiaatiolla tarkennetaan, että suhteen kaikki jäsenet ovat liitetty korkeamman luokan objektiin ja tämä objekti koostuu vain näistä jäsenistä.

Skeemassa Fakta-luokka esitetään koosteluokkana (*composite class*), joka on liitetty yhteen tai useampaan Dimensio-luokkaan jaetulla koostesuhteella (*shared aggregation relationship*). Koostesuhteessa faktan ilmentymään liittyy aina vähintään yksi dimensio-luokan ilmentymä, joten dimensio-luokan kardinaliteetti on vähintään yksi. Fakta-luokan kardinaliteetti on merkitty tähdellä (*), sillä dimensio-luokan ilmentymä voi liittyä nolnaan, yhteen tai useampaan fakta-luokan ilmentymään.

DegenerateFact elementillä kuvataan faktan ja dimension monesta-moneen suhteen ominaisuuksia. Kuvassa 14. Auto-sales faktaan liittyy yksi tai useampi Salesperson - dimension ilmentymä. Commission DegenerateFact - elementti liittää tähän suhteeseen Fakta-attribuutin, joka kuvaa kunkin kauppaan osallistuneen myyjän palkkiota.

Joskus voi olla tarkoituksenmukaista kategorisoida dimension attribuutteja. Tietomalli mahdollistaa kategorisointihierarkioiden luomisen yleistys-erikoistamis suhteen avulla. Kuvassa 15. on esitetty Auto - dimension kategorisointi kaikille autotyypeille yleisiin attribuutteihin, sekä henkilö- ja pakettiautoille relevantteihin attribuutteihin. Vain kategorisointihierarkian ylin taso voi osallistua myös luokitteluhierarkiaan.



Kuva 15. Auto-dimension kategorisointihierarkia. (Luján-Mora ym., 2006)

5.1.3 Tietomallin ominaisuuksia

Seuraavassa on luetteloitu eräitä tietomallin keskeisimpiä piirteitä:

1. Mallin jakaminen eri tarkkuustasoihin käyttäen UML:n ryhmittelymekanismia tukee laajojen ja kompleksisten tietovarastojen mallintamista.
2. Monesta-moneen suhteet faktan ja dimension välillä kuvataan jaetun koostesuhteen kardinaliteetin.
3. Johdetut mittarit (*derived measures*) esitetään tietomallissa UML metamallin määrittelemien johdettujen attribuuttien avulla (*derived attributes*). Johdetun mittarin säännöt (*derivation rules*) kuvataan käyttäen OCL kieltä.
4. Oletuksena mittarit ovat summattavia. Ei-summattavien mittareiden summattavuussäännöt esitetään rajoitteena fakta-luokan lähellä.
5. Luokitteluhierarkiat kuvataan assosiaatiolla taso-luokkien välillä. Hierarkioiden oletuspolut on esitetty suunnatuilla assosiaatioilla.
6. Luokitteluhierarkian tiukkuus ja ei-tiukkuus kuvataan assosiaation kardinaliteetin.
7. Completeness-stereotyyppi viittaa luokitteluhierarkian täydellisyyteen.
8. Dimensioita kategorisoidaan yleistys-erikoistamis-suhteella.

5.1.4 Tietomallin arviointia

Tietomalli täyttää kohdassa 3.3 käsitelmalleille esitetyt yleiset vaatimukset varsin hyvin. Tietomalli on ilmaisuvoimainen kyetessään esittämään tietovarastojen moniulotteisen mallintamisen keskeiset käsitteet käsitteellisellä tasolla. Esitetty tietomalli on kohdealueen monimutkaisuuden huomioonottaen suhteellisen yksinkertainen. Tietomallin määrittelemät käsitteet ovat minimaalisia, sillä näistä mitään ei voida esittää muiden käsitteiden koosteella. Formaalius on saavutettu käsitteiden eksaktilla määrittelyllä hyödyntäen mm. OCL - rajoitteita. OCL mahdollistaa myös formaalien kyselyiden tekemisen

UML - malliin. Tietomallin esitystapa on graafisesti täydellinen ja helppolukuinen.

Erityisesti tietovarastojen moniulotteiseen mallintamiseen tarkoitettujen käsittemallien vaatimusten osalta arviointi on vaikeampaa. Tietomallin ilmaisuvoimaa tarkasteltaessa on ensin arvioitava mitkä vaatimukset todella ovat relevantteja käsitteelliselle tietomallille. Pedersenin (2000) moniulotteisille käsitemalleille esittämät yksitoista vaatimusta tarkasteltavana oleva tietomalli täyttää osittain. Tietomalli ei kuitenkaan kykene kuvaamaan epätyypillisempiä dimensiohierarkioita kuten tasapainottamattomia ja ei-kattavia hierarkioita. Tietomalli määrittelee dimension ja faktan eksplisiittisesti eikä näiden symmetrinen käsittely ole mahdollista. Tietomalli ei myöskään tue eri tarkkuustasojen-, epätarkkuuden-, eikä muutoksen ja ajan hallintaa. Tietomalli määrittelee toisaalta muita keskeisiä käsitteitä tietovarastojen moniulotteiseen mallintamiseen, kuten Completeness - assosiaation, DegenerateFact assosiaatioluokan, DegenerateDimension-, OID- ja Descriptor- attribuutin. Tietomalli mahdollistaa myös dimensioiden kategorisoinnin ja laajojen skeemojen kuvaamisen eri tarkkuustasoilla.

Tarkasteltavana oleva tietomalli täyttää kohdassa 4.3 Abelló ym. (2001) esittämät moniulotteisten käsittemallien luokittelukriteerit välitason tietosolua lukuun ottamatta. Tietomallissa ei siten ole mahdollista ryhmitellä faktan mittareita tietosoluihin ja liittää näitä dimensioiden tasoihin. Faktan mittarit ovat siis aina keskenään samalla tarkkuustasolla.

6 YHTEENVETO

Tietovarastojen moniulotteinen mallintaminen edellyttää eri suunnittelutekniikoita kuin operatiivisten transaktiokantojen suunnittelu. Yleisesti ollaan kuitenkin yhtä mieltä siitä, että sekä transaktiokantojen että tietovarastojen suunnittelussa käsitteellinen mallintaminen on laadukkaan lopputuloksen kannalta tärkeää. Moniulotteinen käsitelmä on työkalu, jolla moniulotteisen tietovaraston keskeiset käsitteet voidaan mallintaa ottamatta kantaa varsinaiseen toteutustapaan. Tietovarastojen moniulotteiset käsitelmät voidaan luokitella kolmeen kategoriaan: ER - mallien laajennukset, UML - mallien laajennukset ja tapauskohtaiset mallit. Yksikään esitetyistä tietomalleista ei ole saanut de facto - standardin asemaa. Tässä tutkielmassa on keskitytty esitettyyn UML - mallin laajennukseen. Mallin perustuessa laajalti tunnettuun ja standardiin mallinnuskieleen, käyttäjien ei tarvitse opetella täysin uutta notaatiota tietokantojen moniulotteiseen mallintamiseen. UML:n sisäänrakennettu laajennusmekanismi tukee tietomallin käyttöä myös tietokantojen moniulotteisessa mallintamisessa.

Tietovarastojen moniulotteisille käsitelmalleille on esitetty suuri joukko erilaisia vaatimuksia. Tarkasteltavana ollut tietomalli täyttää hyvin käsitelmalleille asetetut yleiset vaatimukset. Tietomalli kattaa tietovarastojen moniulotteisen mallinnuksen keskeiset käsitteet. Tarkasteltavana ollut tietomalli ei kuitenkaan kykene kuvaamaan epätyypillisimpiä dimensiohierarkioita eikä tue eri tarkkuustasojen-, epätarkkuuden-, eikä muutoksen ja ajan hallintaa. Kaiken kaikkiaan tietomalli sopii tietovarastojen moniulotteiseen käsitelmämallintamiseen sen ollessa riittävän ilmaisuvoimainen ja formaali, sen perustuessa tunnettuun, standardiin mallinnuskieleen ja ollessaan kohdealueen monimutkaisuus huomioon ottaen varsin yksinkertainen. Tietomallin perustuessa UML:ään, voidaan myös tietovarastojen moniulotteinen käsitelmämallinnus käytännössä suorittaa käyttäen tavanomaisia mallinnusvälineitä.

LÄHDELUETTELO

- Abelló, A., Samos, J. & Saltor, F. 2001. A framework for the classification and description of multidimensional data models. Teoksessa 12th International Conference on Database and Expert Systems Applications (DEXA), volume 2113 of LNCS, Springer, pages 668-677.
- Abelló, A., Samos, J. & Saltor, F. 2002. YAM² (Yet Another Multidimensional Model): An extension of UML. Teoksessa Proceedings of the International Database Engineering & Applications Symposium, Edmonton, Canada, 172-181
- Batini, C., Ceri, S. & Navathe S. 1992. Conceptual Database Design - An Entity-Relationship Approach. The Benjamin/Cummings Publishing Company Inc..
- Chaudhuri, S. & Dayal, U. 1997. An Overview of Data Warehousing and OLAP Technology- SIGMOND Record, Vol. 26, No 1, March 1997.
- Elmasri R. & Navathe S. 2000. Fundamentals of database systems. Addison-Wesley.
- Franconi, E. & Kamble, A. 2004. A data warehouse conceptual data model. Teoksessa Proceedings of the International Conference on Statistical and Scientific Database Management, 435-436.
- Golfarelli, M., Maio, D., & Rizzi, S. 1998. The dimensional fact model: A conceptual model for data warehouses. Teoksessa International of Cooperative Information Systems Vol 7, No 2-3, 215-247.
- Holten R. 2001. Conceptual Models as Basis for Integrated Information Warehouse Development. Arbeitsberichte des Instituts für Wirtschaftsinformatik Nr. 81. Münster.

- Hüsemann, B., Lechtenbörger, J. & Vossen, G. 2000. Conceptual data warehouse design. Teoksessa Proceedings of the International Workshop on Design and Management of Data Warehouses. Stockholm, Sweden.
- Inmon, W.H. & Hackathorn, R.D.: Using the Data Warehouse. John Wiley & Sons, USA 1994.
- Inmon W. H. 1996. Building the Data Warehouse. New York: Wiley Publishing Inc.
- Luján-Mora, S., Trujillo, J. & Song, I. Y. 2002. Extending the UML for multidimensional modeling. Teoksessa Proceedings of the International Conference on the Unified Modeling Language. Dresden, Germany. 290-304
- Luján-Mora, S., Trujillo, J. & Song, I. Y. 2006. A UML profile for multidimensional modeling in data warehouses. Data & Knowledge Engineering Vol 59, Issue 3, Pages 725-769, Elsevier Science Publishers B. V.
- Nguyen, T. B., Tjoa, A. M., & Wagner, R. 2000. An object-oriented multidimensional data model for OLAP. Teoksessa Proceedings of the International Conference on Web-Age Information Management, Shanghai, China, 69-82.
- OMG, Meta Object Facility (MOF) Specification <http://www.omg.org/mof/> viitattu 11.12.2007
- Pedersen, T. B. 2000. Aspects of Data Modeling and Query Processing for Complex Multidimensional Data. Väitöskirja: Faculty of Engineering and Science, Aalborg University (Denmark)
- Rizzi, S., Abelló A., Lechtenbörger J. & Trujillo J. 2006. Research in Data Warehouse Modeling and Design: Dead or Alive? Teoksessa Proceedings

of the ACM 9th International Workshop on Data Warehousing and OLAP (DOLAP), Arlington, Virginia, USA: ACM Press, 3-10

- Rizzi, S. 2007. Conceptual Modeling Solutions for the Data Warehouse. Teoksessa Data Warehouse and OLAP: Concepts, Architectures and Solutions. IRM Press.
- Sapia, C., Blaschka, M., Hofling, G. & Dinter, B. 1998. Extending the E/R model for the multidimensional paradigm. Teoksessa Proceedings of the International Conference on Conceptual Modeling, Singapore.
- Trujillo J., Palomar M., Gomez J. & Song I. 2001. Designing Data Warehouses with OO Conceptual Models. IEEE Computer 34(12), 66-75
- Tryfona, N., Busborg, F., Borgh Christiansen, J. G. 1999. starER: A conceptual model for data warehouse design. Teoksessa Proceedings of the ACM International Workshop on Data Warehousing and OLAP. Kansas City, Kansas, 3-8.
- Tsois, A., Karayannidis, N., & Sellis, T. 2001. MAC: Conceptual data modeling for OLAP. Teoksessa Proceedings of the International Workshop on Design and Management of Data Warehouses. Interlagen, Switzerland, 5.1-5.11.