

Parantaako Data Vault -menetelmä tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta?

Tapaustutkimus, Yritys A

Jaana Sironen

Opinnäytetyöraportti

Tietojärjestelmäosaamisen koulutus-
ohjelma

2012



Tietojärjestelmäosaaminen

<p>Tekijä tai tekijät Jaana Sironen</p>	<p>Ryhmätunnus tai aloitusvuosi YTI10K</p>
<p>Raportin nimi Parantaako Data Vault -menetelmä tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta? Tapaustutkimus, Yritys A</p>	<p>Sivu- ja liitesivumäärä 100 + 4</p>
<p>Opettajat tai ohjaajat Jarmo Halenius, Martti Laiho</p>	
<p>Tässä tutkimuksessa vertailtiin kohdeyrityksen nykyistä tietovaraston mallinnustapaa uuteen Data Vault -menetelmään. Kohdeyrityksen nykyinen tietovarastoratkaisu on ollut käytössä jo melkein kymmenen vuotta ja menetelmänä sen toteutuksessa on käytetty Kimballin tähtimallia. Tietovaraston datamäärät kasvavat koko ajan, ja toiminnan muuttuessa tulee uusia tietotarpeita.</p> <p>Data Vault -menetelmä on ensimmäinen vain tietovarastointiin kehitetty menetelmä, joka johtaa skaalautuviin ja muutosjoustaviin tietovarastoratkaisuihin. Se tukee tietojen historiointia ja jäljitettävyyttä sekä parantaa tietovaraston suorituskykyä. Data Vault -menetelmä tukee myös liiketoiminnan ja tietohallinnon yhteisten tavoitteiden saavuttamista, sillä se mukautuu nopeasti liiketoiminnan muutoksiin.</p> <p>Tutkimus oli luonteeltaan kvalitatiivinen tapaustutkimus. Sen yhteydessä perehdyttiin kirjallisuuskatsauksen avulla Data Vault -menetelmästä saatavavilla olevaan teoriaan ja selvitettiin, onko menetelmä sovellettavissa kohdeyrityksen tarpeisiin. Samalla selvitettiin haastattelujen avulla nykyisen tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta sekä vertailtiin uuden Data Vault -menetelmän tuomia muutoksia nykyiseen ratkaisuun nähden. Tutkimuksen empiirinen osuus toteutettiin projektityyppisesti yhteistyössä kohdeyrityksen edustajien ja tietovarastoasiantuntijoiden kanssa, rakentamalla rajatusta liiketoiminta-alueesta konstruktio, jonka avulla havainnollistettiin ja testattiin menetelmän toimivuutta käytännössä.</p> <p>Tutkimuksen lopputulosten perusteella voidaan todeta, että Data Vault -menetelmä tuo merkittäviä etuja tietovaraston muutosjoustavuuteen ja skaalautuvuuteen, kun sitä verrataan tähtimalliin. Data Vault -menetelmän käytön aloituskustannukset ovat korkeat, mutta ylläpidon vakiinnuttua kustannushyödyt saadaan paremmin käyttöön.</p>	
<p>Asiasanat Data Vault, tietovarasto, tietomalli, mallinnusmenetelmä, skaalautuvuus, muutosjoustavuus</p>	

Information technology

<p>Authors Jaana Sironen</p>	<p>Group or year of entry YTI10K</p>
<p>The title of thesis Does the Data Vault -methodology make a data warehousing solution more scalable and change flexible? Case: Company A</p>	<p>Number of pages and appendices 100 + 4</p>
<p>Supervisor(s) Jarmo Halenius, Martti Laiho</p>	
<p>The main purpose of this thesis was to compare the Data Vault -methodology to the existing data modeling solution of the Date Warehouse in the target company. The current data warehousing solution has been used in the target company for almost ten years, and its data modeling solution has been Kimball's star schema. The quantity of data is growing constantly and new requirements appear as the organization changes.</p> <p>The Data Vault -methodology is the first methodology developed only for data warehousing, and it leads to the data warehousing solutions which are more scalable and more change flexible. The methodology supports historical data and data tracking, and makes the Data Warehouse perform better. The Data Vault -methodology supports the reaching of common goals in business and IT because it adapts quickly to the changes in business.</p> <p>The research was carried out as a qualitative case study. It focused on a literature review on available theory of the Data Vault -methodology. In addition, it clarified whether the methodology is applicable for the needs of the target company. The aim was also to investigate, through interviews, the scalability and change flexibility of the existing data warehouse solution, and what differences the Data Vault -methodology brings compared to the existing solution. The empirical part of the thesis was done as a project with the employees of the target company and data warehousing experts. The functionality of the methodology was tested by constructing an example of a limited business area.</p> <p>The thesis indicated that the Data Vault -methodology brings significant benefits in scalability and change flexibility when it is compared to the star schema. At the beginning of the use of the Data Vault -methodology the costs are high, but when the maintenance matures, the costs will reduce.</p>	
<p>Key words Data Vault, Data Warehouse, Data Model, Data Modeling, Scalability, Change flexibility</p>	

Sisällys

1	Johdanto	1
1.1	Tieto – yksi yrityksen tärkeimmistä pääomista.....	2
1.2	Tutkimuksen kohteena oleva yritys A	5
1.3	Tutkimuksen toteuttaminen ja rajaukset	5
1.4	Tutkimusongelma.....	6
1.5	Tutkimusraportin sisältö.....	7
2	Tietojen mallinnus tietovarastossa	8
2.1	Tietovarastomallinnuksen historiaa	8
2.2	EDW:n tarpeet tietovarastoinnille	10
2.3	Aiempien mallinnusmenetelmien ongelmat EDW:ssä	12
2.4	Data Vault –menetelmä.....	14
2.4.1	Hub-entiteetit.....	16
2.4.2	Link-entiteetit.....	17
2.4.3	Satellite-entiteetit	17
2.4.4	Data Vault -tietovaraston rakentaminen	18
2.4.5	Data Vault -tietokannan skaalautuvuus	21
2.4.6	Data Vault -menetelmän tuomat muutokset EDW:lle.....	22
2.4.7	Data Vault ja DW2.0™	25
2.4.8	Business Data Vault	27
3	Tutkimusperusta ja käytetyt tutkimusmenetelmät	30
3.1	Konstruktiiivinen tutkimusote.....	31
3.2	Kirjallisuuskatsaus	32
3.3	Dokumenttianalyysi	33
3.4	Teemahaastattelu	33
3.5	Vertailun suorittaminen nykytilan ja rakennetun konstruktion välillä.....	34
3.6	Tutkimuksen viitekehys ja rajaukset	35
3.6.1	Tutkimuksen kohteena oleva raportointialue	35
3.6.2	Nykyisen tietovarastoratkaisun ongelmat.....	36
3.6.3	Kohdeyrityksen tietovarastoraportoinnin kehittämishanke.....	38
4	Kehittämishankkeen vaikutukset tietovarastoraportointiin.....	40

4.1	Nykyinen tietovarastoarkkitehtuuri.....	41
4.2	Uudistettu Data Vault -tietovarastoarkkitehtuuri	41
4.3	Nykyinen tietomalli	44
4.4	Uudistettu Data Vault -tietomalli.....	45
4.5	Data Vault tietomallin vaikutus tietovarastolatauksiin	47
4.6	Viite-cheys Data Vault -mallinnetussa tietovarastossa	50
4.7	Esimerkkejä Data Vault -menetelmän vaikutuksesta raportointiin	51
4.7.1	Päättäneiden eläkekäsittelyjen raportointi nykyisessä ratkaisussa	53
4.7.2	Päättäneiden eläkekäsittelyjen raportointi uudistetussa ratkaisussa	55
4.7.3	Uusi raportoitava liiketoimintakäsite, käsittelijän eläkekäsittelytehtävä .	58
5	Empiria tutkimuksen kohteena olevasta raportointialueesta.....	66
5.1	Haastattelujen toteutus	66
5.2	Haastateltavien valinta	67
5.3	Nyky- ja tavoitetilan haastattelujen vertailu	68
5.3.1	Ylläpidettävien raporttien muutospyyntöjen ilmenemistiheys.....	69
5.3.2	Nyky- ja tavoitetilan raportoinnin yleisimmät ongelmat	71
5.3.3	Ylläpidon toteutuksen muutosjoustavuus ja helppous	72
5.3.4	Kokemuksia raportoinnin ylläpidosta.....	74
5.3.5	Millaisia raportteihin tulevat muutostarpeet ovat?	78
5.3.6	Data Vault -menetelmän tuomat vaikutukset raportointiin.....	78
6	Tutkimuksen eteneminen.....	82
6.1	Tutkimuksen tekninen ympäristö.....	84
6.2	Tutkimuksen vaiheet.....	85
7	Tutkimuksen yhteenveto.....	87
7.1	Yhteenveto tutkimustuloksista ja niiden kontribuutio käytäntöön.....	89
7.2	Tutkimuksen rajoitukset ja yleistettävyys	92
7.3	Jatkotoimenpide-ehdotukset.....	93
8	Johtopäätökset.....	95
	Lähteet.....	99
	Liitteet.....	101
	Liite 1. Eläkekäsittelyn tähtimalliin perustuva Universe BI-välineessä	101
	Liite 2. Eläkekäsittelyn Data Vault -rakenne tallennuskerroksessa.....	102

Liite 3. Käsittelijätehtävän hyödyntäminen raporteilla (luottamuksellinen)	103
Liite 4. Nyky- ja tavoitetilahaastattelujen analyysit (luottamuksellinen).....	104

Kuviot

- Kuvio 1. EDW-tietovarastoarkkitehtuuri
- Kuvio 2. Esimerkki Data Vault -menetelmästä
- Kuvio 3. Esimerkki Data Vault -arkkitehtuurilla toteutetusta EDW:stä
- Kuvio 4. Data Vault -malliin tehdyt muutokset
- Kuvio 5. DW2.0™ tietovarastosektorit
- Kuvio 6. BDV-ratkaisun sijainti EDW:ssä
- Kuvio 7. Nykyisen raportointitietokannan arkkitehtuuri
- Kuvio 8. Uuden tietovaraston arkkitehtuuri uudistuksen jälkeen
- Kuvio 9. Uuden tietovaraston eri kerrokset
- Kuvio 10. Tähtimallilla mallinnettu taulurakenne
- Kuvio 11. Esimerkki Data Vault -mallinnetusta taulurakenteesta
- Kuvio 12. Esimerkki statussatelliitista SS_Elakekasittely_Kasittelija, joka kertoo linkin voimassaolon
- Kuvio 13. Raportointiesimerkki, viimeisin versio raportointiajankohdilla 1.2. ja 1.3.2011
- Kuvio 14. Raportointiesimerkki, lukitus, kun jälkikirjaukset näytetään kirjauskuukaudella, mutta jälkipoistoja ei käsitellä lukitusta datasta (raportointiajankohdat 1.2. ja 1.3.2011)
- Kuvio 15. Raportointiesimerkki, lukitus, jälkikirjaukset ja -poistot näytetään omina mittareinaan (raportointiajankohdat 1.2. ja 1.3.2011)
- Kuvio 16. Käsittelijän tehtävien liittyminen liiketoiminnan muihin käsitteisiin
- Kuvio 17. Eläkekäsittelytehtävä-linkki (L_Elakekasittely_Tehtava_L) ja siihen liittyvät yhteydet
- Kuvio 18. Haastattelukyselyjen vastaajamäärät rooleittain nykytilan (N = 13) ja tavoitetilan (N=10) osalta
- Kuvio 19. Tavoitetilan (N = 10) vastausarvot suhteessa nykytilan (N = 13) vastausten keskiarvoon raporttimuutosten ilmenemistiheydessä
- Kuvio 20. Tavoitetilan (N = 10) vastausarvot suhteessa nykytilan (N = 13) vastausten keskiarvoon tiedon saamisesta raportille helposti ja joustavasti eri toteutustarpeiden näkökulmista katsottuna

- Kuvio 21. Tavoitetilan (N = 10) vastausarvot suhteessa nykytilan (N = 13) vastausten keskiarvoon tiedon saamisesta raportille helposti ja joustavasti
- Kuvio 22. Osittain ja täysin samaa mieltä olleiden vastaajien erot tavoitetilan (N = 10) ja nykytilan (N = 13) vastausten välillä muutos- ja kehityspyyntöihin liittyvien väittämien osalta
- Kuvio 23. Osittain ja täysin samaa mieltä olleiden vastaajien erot tavoitetilan (N = 10) ja nykytilan (N = 13) vastausten välillä ylläpidon nopeuden ja joustavuuden osalta
- Kuvio 24. Raportoinnin yleisarvosanat tavoitetilan (N = 10) ja nykytilan (N = 13) vastausten osalta
- Kuvio 25. Vastaajien roolien suhde osittain tai täysin samaa mieltä olleiden vastaajien kesken (N = 5) siitä, että perustuuko tietovarasto liiketoiminnan avainkäsitteisiin
- Kuvio 26. Uusien käsitteiden luontiominaisuuden vaikutukset raportoinnin kehittämiseen (N = 5)

Taulukot

- Taulukko 1. Esimerkkidata
- Taulukko 2. Eläkekäsittely, päätetyt eläkekäsittelyt aiemmassa raportointitietokannassa raportointihetkellä 1.2.2011
- Taulukko 3. Eläkekäsittely, päätetyt eläkekäsittelyt aiemmassa raportointitietokannassa raportointi-hetkellä 1.3.2011
- Taulukko 4. Eläkekäsittely, päätetyt käsittelyt, lukitus kauden päätteeksi uudessa tietovarastossa raportointihetkillä 1.2. ja 1.3.2011
- Taulukko 5. Käsittelijän eläkekäsittelytehtävän jakautuminen tehtävätyyppeihin
- Taulukko 6. Käsittelijän erilaiset roolit
- Taulukko 7. Tiedonkeruutehtävän määrittelyä
- Taulukko 8. Tiedonkeruutehtävään liittyvät ehdot
- Taulukko 9. Haastateltavien roolit
- Taulukko 10. Tutkimusprojektin vaiheet

Käsitteet

3NF	=	3 rd Normal Form, kolmannen normaalimuodon relaatiomalli.
BDV	=	Business Data Vault, tietovaraston mallinnusratkaisu, jota käytetään liiketoimintasäännöillä muokatun informaation hyödyntämisessä.
BI	=	Business Intelligence, liiketoimintatieto, joka on liiketoimintasäännöillä muokattua informaatiota.
Conformed Data	=	Useamman aihealueen tähtimalli.
Marts		
Data Mart	=	Datamartti, yhden aihealueen tähtimalli eli paikallisvarasto.
Data Vault	=	Tietovarastojen mallinnusarkkitehtuuri, joka sisältää sääntöjä ja standardeja ja jonka vaihtoehtoinen nimi on "Common Foundational Integration Model Architecture".
DW	=	Data Warehouse, tietovarasto, jossa tiedot jalostetaan raportointi- ja kyselykäyttöön soveltuvaan muotoon.
DW2.0™	=	Bill Inmonin kehittämä uuden polven tietovarastoarkkitehtuuri.
EDW	=	Enterprise Data Warehouse, usean liiketoiminta-alueen tiedot sisältävä keskitetty tietovarastoarkkitehtuuri.
ETL	=	Prosessi, joka muodostuu kolmesta vaiheesta Extract - Transform - Load (poiminta, muokkaus, lataus). Prosessissa luetaan tiedot operatiivisista järjestelmistä tai siirtotiedostoista, muokataan ne tietovarastokannan muotoon ja lopuksi ne kirjoitetaan tietovarastoon.
IT	=	Information Technology, informaatioteknologia.
Käsiteanalyysi / ER-mallinnus	=	Mallinnusmenetelmä operatiivisten kantojen suunnitteluun (Data Modeling). Käsiteanalyysiä kutsutaan myös ER-mallinnukseksi (Entity - Relationship Modeling), joka sopii hyvin myös tietovaraston suunnitteluun.
Snowflake Schema	=	Lumihiutalemalli, joka on yhdenmukaistettu ratkaisu tähtimallista.
Star Schema	=	Tähtimalli, menetelmä tietovaraston tietojen mallintamiseen.

1 Johdanto

Tässä tutkimuksessa perehdytään Data Vault -menetelmään, joka on suhteellisen uusi ja maailmalla suurta mielenkiintoa herättänyt tietovarastojen mallinnusmenetelmä. Se painottaa tietovaraston muutosjoustavuutta, skaalautuvuutta ja jäljitettävyyttä. Data Vault on suunniteltu myös liiketoimintamalliksi, sillä liiketoiminnan muuttuessa sitä on helppo muuttaa ja siksi sen arvo liiketoiminnalle säilyy tai jopa kasvaa ajan kuluessa. Data Vault -menetelmä on syntynyt vähitellen aiemmin käytössä olleiden tietovarastojen mallintamisessa käytettyjen menetelmien pohjalta.

Tiedon tallentamisen alkutaipaleilla luotiin relaatiomalli reaaliaikaisten tapahtumien käsittelyä varten. Tämän kolmannen normaalimuodon relaatiomallin (3rd Normal Form, 3NF) kehitti ensisijaisesti Ted Codd. Seuraavaksi relaatiomallia muokkasi tietovarastoinnin isäksi nimetty Bill Inmon ja lisäsi siihen aikaleiman avainkenttiin, jotta se sopisi paremmin tiedon varastointiin. Tämän jälkeen Ralph Kimball kehitti tähtimallin (Star Schema), josta on olemassa yhden aihealueen tähtimalli eli paikallisvarasto (Data Mart) ja useamman aihealueen kokonaisuus (Conformed Data Marts). (Martin 2008.)

Toiminnan ja tietotarpeiden muuttuessa ja kasvaessa ajan myötä, on tietovaraston oltava ketterä myös laajennuksille ja muutoksille. Ajan saatossa on selvinnyt, että relaatio- ja tähtimallissa esiintyy suorituskykyongelmia silloin, kun datamäärät kasvavat. Lisäksi yhdeksi ongelmaksi laajoissa tietovarastoissa on muodostunut muutosjäykkyys. Näiden ongelmien ratkaisemiseksi on Dan Linstedt kehittänyt Data Vault -menetelmän, joka on tarkoitettu juuri tietovarastojen mallintamiseen (Linstedt 2002).

Tutkimuksen kohteena olevassa yrityksessä tietovarasto on ollut käytössä jo yli kymmenen vuotta ja mallinnusmenetelmänä sen toteutuksessa on käytetty Kimballin tähtimallia. Tietovarasto on ratkaistu hallintomallilla, jossa tietojen hyväksikäyttö ja hallinnointi hoidetaan keskitetysti. Itse tiedot ovat tietovarastossa omina Data Martteinaan liiketoiminnoittain. Tietovaraston tietomäärät kasvavat jatkuvasti ja toiminnan muuttuessa tulee uusia tietotarpeita. Tehokkaan ja oikea-aikaisen tietopääoman ja raportoinnin ylläpitämiseksi on tärkeää, että tietovarasto on ketterä liiketoimintamuutoksille. Tästä johtuen kohdeyrityksessä on käynnissä kehittämishanke, jonka lopputuloksena syntyy

nykyisiä ja tulevia liiketoimintatarpeita paremmin huomioiva tietovarastoratkaisu, joka perustuu Data Vault -menetelmään.

Tutkimuksessa selvitetään Data Vault -ratkaisun skaalautuvuutta ja muutosjoustavuutta aiempaan tähtimallilla toteutettuun tietovarastoratkaisuun verrattuna. Lisäksi siinä verrataan Data Vault -menetelmän tuomia uusia vaihtoehtoja tähtimalliin nähden. Kohdeyrityksessä käynnissä oleva johdon raportoinnin kehittämishanke on osa tätä tutkimusta. Data Vault -menetelmän käyttöönotto kohdeyrityksessä on myös yksi vaihtoehto toiminnan parantamiseen ja tietopääoman hyödyntämiseen, sillä Martinin mukaan (2008) Data Vault on ensimmäinen juuri tietovarastoinnin mallintamiseen suunniteltu menetelmä, joka tukee ketterää tapaa tietovarastoon kohdistuvien muutos- ja laajennustarpeiden toteuttamiselle.

1.1 Tieto – yksi yrityksen tärkeimmistä pääomista

IT:n tehtävänä on tuottaa liiketoiminnalle informaatiota, jota se tarvitsee tavoitteidensa saavuttamiseksi, oikea-aikaisesti ja luotettavasti. Informaation tulee olla liiketoiminnan vaatimusten mukaista ja luotettavaa. Sen laatuun kohdistuvia kriteereitä ovat vaikuttavuus ja tehokkuus sekä turvallisuuteen kohdistuvat luottamuksellisuus, eheys ja saatavuus (GobIT 4.1).

Hovi, Hervonen & Koistinen (2009, XI) painottavatkin, että nykypäivän yritys pitää tietoa arvokkaana pääomana ja resurssina, jonka tulisi olla hyvin kuvattuna ja saatavilla palvelemaan yhä lisääntyviä liiketoimintatarpeita. Heidän mukaansa tämän päivän yrityksissä tietojärjestelmäratkaisujen tiedot ovat sirpaloituneet organisaatio-osien ja sovelusten mukaan omiin siloihinsa. Tällöin menetetään kokonaisnäkemys esimerkiksi asiakkaaseen, vaikka todellisuudessa koko organisaation näkökulmasta puhutaankin samoista asiakasta koskevista tiedoista. Tietojen hajanaisuus onkin Hovin ym. (2009, XI) mielestä johtanut tarpeeseen integroida tietoja ja valjastaa uusi informaatio liiketoiminnan päätöksenteon tueksi eli Business Intelligenceksi, BI. (Hovi, Hervonen & Koistinen 2009, XI.)

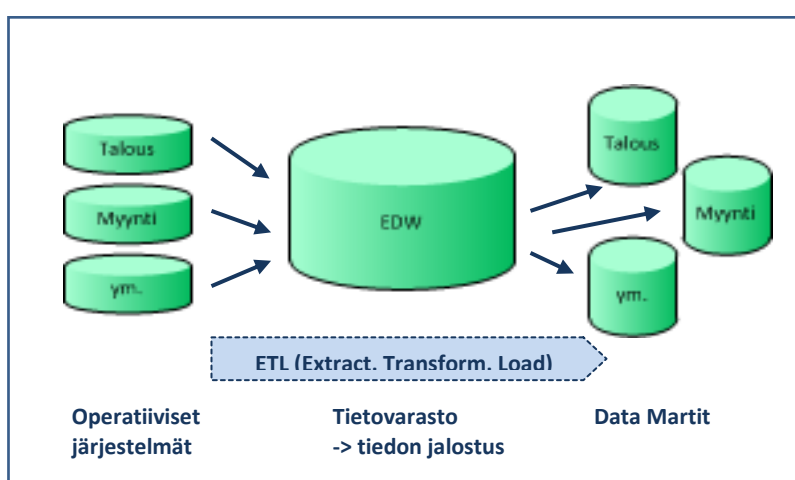
Tietoa voidaan käsitellä hallittavissa ja johdettavissa olevana asiana. Tällöin puhutaan niin sanotusta tietämyksenhallinnasta (Knowledge Management), joka Hovin ym. (2009, 190) mukaan liitetään usein myös tietojohdamiseen. He ovat sitä mieltä, että tietämyksenhallinta on joukko ajattelutapoja ja tuotteita, joiden avulla tietoyhteiskunnan työntekijä pääsee kiinni yrityksen tietoihin. Niihin luetaan mukaan myös niin sanottu hiljainen, kokemusperäinen tieto (tacit-tieto), johon kuuluvat muun muassa dokumentit, ideat, keksinnöt ja patentit. Lisäksi heidän mielestään tietovarastoinnin ja tietopääoman hallinnan näkökulmista katsottuna on hyvä erottaa toisistaan eksplisiittinen ja tacit-tieto. Valtaosa organisaation tietopääomasta on tacit-tietoa, joka on kokemusperäistä ja henkilöstön niin sanottua hiljaista tietoa. Tietovarastot taas pystyvät käsittelemään vain eksplisiittistä tietoa, joka on määriteltyä ja täsmällistä. (Hovi ym. 2009, 190; Hovi, Ylinen & Koistinen 2001, 186.)

Laamasen ja Tinnilän (2009, 105) mukaan tieto on hyvin perusteltu uskomus. He määrittelevät tietoa siten, että sillä voidaan ajatella olevan totuus- ja käyttöarvoa. Totuusarvo käsittää sen tiedon, mikä on totta, todellista ja havainnoijasta riippumatonta, jolloin se kertoo siitä, miten asiat ovat. Tällainen tieto on luonteeltaan objektiivista, josta hyvänä esimerkkinä on organisaation toimintaan liittyvät kappalemäärät, kuten montako tuotetta tai palvelua organisaatio on toimittanut sovittuna aikana. Käyttöarvo käsittää tiedon siitä, miten asioiden oletetaan olevan, ja minkä perusteella esimerkiksi kohdeyrityksen liiketoimintapäätäjät tekevät omat johtopäätöksensä. Tällöin tieto on luonteeltaan subjektiivista, josta hyvänä esimerkkinä ovat käsitykset organisaation ilmapiiristä. (Laamanen & Tinnilä 2009, 105.)

Tietojen organisointitapaa tietovarastoihin tarvitaan, sillä operatiiviset järjestelmät eivät sovi tietojen vaativaan analysointiin. Tästä johtuen eri perusjärjestelmien tiedot kannattaa yhdistellä tietovarastoon (Data Warehouse, DW), jossa ne jalostetaan raportointi- ja kyselykäyttöön soveltuvaan muotoon. Tietovarasto on juuri tietojen nopeaa ja helppoa hakua varten suunniteltu tietokanta. Tietovarasto mahdollistaa myös trendianalyysien teon, sillä tietoja ladatessa uudet tiedot menevät edellisten perään, jolloin historiatiedot säilyvät. Operatiivisten järjestelmien ja tietovaraston väliin sijoittuu ETL-prosessi, joka muodostuu kolmesta vaiheesta Extract - Transform - Load (poiminta, muokkaus, lata-

us). Prosessissa luetaan tiedot operatiivisista järjestelmistä tai siirtotiedostoista, muokataan ne tietovarastokannan muotoon ja lopuksi ne kirjoitetaan tietovarastoon (kuvio 1).

Tietovarastoa rakennettaessa sen arkkitehtuuri voi yksinkertaisimmillaan olla yksi erillinen Data Martti, joka sisältää pääasiassa yhden liiketoiminta-alueen tiedot yhteen laskettuna, tai keskitetty usean liiketoiminta-alueen tiedot sisältävä Enterprise Data Warehouse (EDW) -arkkitehtuuri. Seuraavassa kuviossa 1 tietovarastoarkkitehtuuri on kuvattu EDW:n näkökulmasta, jolloin se sisältää sekä operatiiviset järjestelmät, varsinaisen tietovaraston sekä raportointiin soveltuvat Data Martit. (Hovi ym. 2001, 29, 48.)



Kuvio 1. EDW-tietovarastoarkkitehtuuri

Yhteenvetona voidaan todeta, että jokainen organisaatio tarvitsee tietoa. Se on yksi yrityksen tärkeimmistä pääomista, jota se hyödyntää toiminnassaan. Tietoa on luonteeltaan erilaista ja siksi tärkeäksi nousee tiedon käyttöarvo, jota yrityksen päättäjät hyödyntävät päätöksenteossa sekä tiedolla johtamisessa. Myös kohdeyrityksen perustehtävä pohjautuu tietoon, jolloin sen tietopääoman tulisi olla keskeinen osa sen toiminnan johtamista ja päivittäistä päätöksentekoa.

Tietämyksenhallinta on keino, jonka avulla työntekijä pääsee kiinni yrityksen tietopääomaan. Yhtenä välineenä siihen voidaan käyttää tietovarastoa, joka pystyy käsittelemään yrityksessä olevaa eksplisiittistä tietoa. Tietovarastointi onkin osa-alue, joka tukee tiedon jakamista koko yritystä hyödyntäväksi tietopääomaksi. Jotta tietopääomaan pääsisi helposti käsiksi, on tärkeää, että tiedot on tallennettu keskitettyyn paikkaan ja mal-

linnettu siten, että ne ovat helposti liiketoiminnan hyödynnettävissä. Luvussa 2 (Tietojen mallinnus tietovarastossa) käydään läpi tietojen mallinnuksen historiaa ja nykypäivää etenkin tietovarastoinnin näkökulmasta sekä siihen liittyviä haasteita ja vaatimuksia.

1.2 Tutkimuksen kohteena oleva yritys A

Tutkimuksen kohteena on yritys A, joka on työeläkevakuutusyhtiö (myöhemmin kohdeyritys). Kohdeyritys on suomalaisen työn eläkevakuuttaja ja se huolehtii yksityisen sektorin työntekijöiden ja yrittäjien eläketurvasta. Kohdeyritys kuuluu Suomen suurimpiin työeläkeyhtiöihin ja sijoittajiin. Sen tehtävänä on kurtuttaa eläkemaksuina saatavia varoja nykyisten ja tulevien eläkkeiden maksamista varten. Hyvän työeläkeyhtiön tunnusmerkkejä ovat eläkkeiden turvaaminen, riittävä vakavaraisuus, hyvä sijoitustulos, palveluiden monipuolisuus ja hyvä laatu. Kohdeyrityksen tavoitteena on nyt ja tulevaisuudessa täyttää hyvän työeläkeyhtiön keskeiset tunnusmerkit ja olla hyvämaineinen, vakaa, osaava ja tehokas.

Kohdeyrityksen kaltaiselta työeläkeyhtiöltä vaaditaan sisäistä tehokkuutta sekä oikea-aikaista päätöksentekoa. Yksi johtamisen haasteista on tietotekniikan hyödyntämisen ja tiedolla johtamisen kompleksisuuden ymmärtäminen sekä hallinta. Tähän luo haasteita kasvavat tietomäärät, lisääntyvä integrointi, nopeusvaatimukset sekä järjestelmien monimutkaistuminen. Tähän haasteeseen vastaaminen vaatii ajan tasalla olevaa tietopääomaa ja siksi tietovaraston ylläpidettävyys ja muutosjoustavuus tulisi olla mahdollisimman tehokasta. Näillä keinoilla mahdollistetaan, että tieto on kohdeyrityksen käyttäjien saatavilla yli organisaatorajojen juuri silloin, kun sitä tarvitaan.

1.3 Tutkimuksen toteuttaminen ja rajaukset

Tämä tutkimus on Haaga-Helian ylemmän ammattikorkeakoulututkinnon tietojärjestelmäosaamisen koulutusohjelman opinnäytetyö. Tutkimuksen kohteena on yritys A ja tutkimusmenetelmänä käytetään konstruktivistista tutkimusotetta. Tutkimuksessa perehdytään uuteen Data Vault -menetelmään tietovaraston mallinnuksessa. Tarkoituksena on tutustua menetelmään liittyvään teoriaan ja pohtia sen pohjalta menetelmän soveltuvuutta kohdeyrityksen tietovarasto-ympäristöön ja sen tuomia eroja nykyiseen käyttöön verrattuna.

Data Vault -menetelmän toimivuutta testataan kohdeyrityksen tietovarastoympäristössä, rakentamalla konstruktio rajatulle raportointialueelle tietovarastossa. Lisäksi vertailaan haastattelujen avulla nyky- ja tavoitetilan eroja raportoinnin muutosjoustavuudessa sekä sitä, tuoko Data Vault -menetelmä parannusta tietovaraston muuttuvien liiketoimintatarpeiden ylläpitoon. Sen lisäksi selvitetään, onko Data Vault -menetelmä hyödynnettävissä sellaisenaan vai tuleeko sitä jalostaa kohdeyrityksen tarpeisiin.

Tutkimuksen yhteydessä saadaan käytännön tietoa myös siitä, miten menetelmän käyttöä voidaan laajentaa kohdeyrityksessä ja kuinka siihen perehtyminen voidaan toteuttaa, mikäli tutkimustulokset tukevat käytön laajentamista. Vasta tutkimuksen lopuksi voidaan varmistua siitä, kumpi mallinnusmenetelmä on tarpeeksi joustava liiketoiminnan nykyisille muutostarpeille. Lisäksi kohdeyritys saa tutkimuksen kohteena olevasta kehittämishankkeesta vertailutietoa siitä, millaisia investointeja Data Vault -menetelmän käyttöönotto nykyisillä tietovaraston informaatio- ja käyttäjävolyymeilla vaatii.

Tutkimuksen rajauksena on, että se kohdistuu vain yhteen työeläkealan yritykseen Suomen markkinoilla ja Data Vault -menetelmän hyödyntämiseen sen tietovarastoinnissa. Tutkimuksessa lähestytään yhtä kohdeyrityksessä ilmenevää ongelmaa konstruktiiivisella tutkimusotteella, jonka yhteydessä testataan Data Vault -menetelmän sopivuutta kohdeyrityksen tietovarastoon rajatulla liiketoiminta-alueella. Kohdeyrityksessä tutkimuksen aikana käynnissä on tietovarastoraportoinnin kehittämishanke ja tutkimuksessa keskitytään vain tämän käynnissä olevan kehityshankkeen tuomiin vaikutuksiin ja muutoksiin raportoinnissa. Lisäksi tutkimuksen rajauksena on, että se kohdistuu ensisijaisesti kohdeyrityksessä käytössä olevan tähtimallin ja Data Vault -menetelmän välisiin eroihin raportoinnin ylläpidossa ja muutosjoustavuudessa.

1.4 Tutkimusongelma

Tutkimusongelmana on kysymys: Parantaako Data Vault -menetelmä tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta? Tapaustutkimus, Yritys A.

Kirjallisuuskatsauksen, dokumenttianalyysin, teemahaastattelun ja toteutettavan uuden konstruktion avulla etsitään vastauksia edellä mainittuun tutkimusongelman pääkysymykseen sekä sitä tarkentaviin kysymyksiin:

- Miten Data Vault -menetelmä eroaa nykyisin käytössä olevasta tietovarastojen mallinnusmenetelmästä?
- Mitä muutoksia uusi Data Vault -menetelmä tuo tietovarastoarkkitehtuuriin ja -raportointiin nykyiseen verrattuna?

1.5 Tutkimusraportin sisältö

Ensimmäinen luku johdattaa lukijan raportin aiheeseen ja sisältöön. Raportin toisessa luvussa lähestytään tutkimusaihetta teoreettisesta näkökulmasta. Teoriakatsauksessa luodaan tutkimukselle vankka teoriaperusta lähteissä julkaistuista tiedoista sekä jäsennetään tutkimuksessa olevaa kehittämiskohdetta muun muassa kirjallisuuskatsauksen avulla. Tätä näkökulmaa hyödynnetään myös tutkimuksellisessa kehittämistyössä. Tutkimuksessa perehdytään lähdekirjallisuuteen ja muuhun materiaaliin siltä osin, kun sitä on saatavilla.

Seuraavaksi raportissa käydään läpi tutkimusperusta ja tutkimuksen viitekehys. Kolmannessa luvussa avataan lukijalle tutkimusperustan lisäksi tutkimuksessa käytettyjä tutkimusmenetelmiä sekä miten niitä on käytetty tutkimuksen yhteydessä. Samassa luvussa tuodaan esille tutkimuksen viitekehys sekä siihen liittyvät rajaukset. Neljännessä luvussa kuvataan kehittämishankkeen vaikutukset tietovarastointiin ja erot käytössä olevan tähtimallinnetun ja Data Vault -menetelmällä toteutetun tietovaraston välillä. Lisäksi aihetta havainnollistetaan raportointiesimerkkien avulla.

Tutkimuksen empiriaosuus on luvussa viisi. Se sisältää nyky- ja tavoitetilan teemahaastattelujen vertailun, joka on yhteenveto haastatteluanalyseistä. Tutkimusprojektin toteuttamiseen ja konstruktion rakentamiseen käytettyyn tekniseen kehittämissympäristöön voi perehtyä luvussa kuusi, jonka jälkeen raportin loppuun seitsemän ja kahdeksan on koottu yhteenveto tutkimuksesta ja sen pohjalta tehdyistä johtopäätöksistä ja jatkotoimenpide-ehdotuksista.

2 Tietojen mallinnus tietovarastossa

Tiedot mallinnetaan tietovarastoon mallinnusmenetelmän avulla, joihin kuuluvat myös tutkimuksen kohteena oleva Data Vault -menetelmä sekä tähtimalli. Tietokantojen mallinnuksessa käytettävä tietomalli on joko looginen tai fyysinen kuvaus tiedosta. Sen avulla pystytään kategorisoimaan, organisoimaan ja hallinnoimaan erilaisia informaatiokokonaisuuksia. Tietomalli mahdollistaa myös mekanismin, jolla pystytään toteuttamaan, muuttamaan ja lisäämään esille nousevia tietotarpeita. Toteutus tapahtuu tyypillisesti SQL-kielen (Standard Query Language) avulla, joka sisältää tietokantatyöskentelyn peruskomennot, kuten lisätä, päivittää, poistaa ja valita (insert, update, delete, select). (Linstedt, Graziano & Hultgren 2009, 26.)

Operatiivisten kantojen suunnittelussa on käytetty paljon käsiteanalyysiä eli tietojen mallinnusta (Data Modeling). Käsiteanalyysiä kutsutaan myös ER-mallinnukseksi (Entity - Relationship Modeling), joka sopii hyvin myös tietovaraston suunnitteluun. Tietovarastojen suunnittelussa erona perinteiseen käsiteanalyysiin on, että siinä rajaudutaan kuvaamaan olemassa olevien tietojärjestelmien tietosisältöä. Sen sijaan uusia operatiivisia järjestelmiä suunniteltaessa kuvataan vapaasti kohdealueen reaali maailmaa. (Hovi ym. 2009, 36.)

Muitakin eroja löytyy operatiivisen ja tietovaraston mallinnuksen väliltä. Ensinnäkin tietovarastokanta suunnitellaan rakenteeltaan selkeäksi päinvastoin kuin operatiiviset kannat. Näin ollen taulujen ja tietojen nimeämisessä pyritään selkeyteen ja ymmärrettävyyteen. Toiseksi tietovarastoon isoille massoille on tehtävä kestävät rakenteet ja siksi eri tietojen päivitystiheyttä ja määrää on mietittävä huolellisesti. Lisäksi yksi tietovarastojen suunnitteluun kuuluva asia on kannassa olevien tietojen historiointi. (Hovi ym. 2009, 36.)

2.1 Tietovarastomallinnuksen historiaa

USA:ssa Data Warehouse -konseptin eli tietovarastoinnin isänä pidetään yleisesti W. H. Inmonia. Hän on kirjoittanut aiheesta lukuisia kirjoja, joista ensimmäisen jo 1990-luvulla "Building the Data Warehouse". Terminä "Data Warehouse" esiintyi kuitenkin

ensimmäisen kerran vuonna 1988 IBM Systems Journal -lehden artikkelissa "An architecture for business an information system", jonka kirjoittivat Barry Devlin ja Paul Murphy. Yleisesti Data Warehouse suomennetaan tietovarastoksi ja sen lyhenne on DW. (Hovi 2010, 6.)

1960-luvulla Ted Codd kehitti tietovarastoinnin perustana olleen relaatiokantojen kolmannen normaalimuodon (3rd Normal Form, 3NF), joka perustuu Dr. Peter Chenin ER- mallinnukseen ja soveltuu näin myös reaaliaikaiseen tapahtumien käsittelyyn (Online Transaction Processing, OLTP) . Myöhemmin Codd ja muun muassa Boyce ovat kehittäneet näitä teoreettisia normaalimuotoja eteenpäin aina viidenteen normaalimuotoon (5NF) asti, mutta käytännön tietokannoissa katsotaan 3NF:n yleensä riittävän laadukkaaksi ratkaisuksi. Tiedon varastoinnin tarve lisääntyi 1980-luvun alkupuolella. Tällöin Inmon muokkasi kolmannen normaalimuodon tietokantatauluja ja lisäsi niiden avainkenttiin aikaleiman, jotta se soveltuisi tiedon varastointiin. (Linstedt 2002.)

1980-luvun puolivälissä Kimball kehitti tähtimallin, joka tukee reaaliaikaista tietojen analysointia (Online Analytical Processing, OLAP) ja joka sopi paremmin tiedonvarastointiin. Tähtimalli mahdollisti aggregaatiot (tietojen koostamisen, summaamisen), muutokset tietomalliin, tehokkaammat kyselyt, käyttäjälähtöisemmän lähestymistavan sekä tiedon uudelleen käytettävyyden tai jakamisen. Tähtimalli on tietovaraston yhden aihealueen mallinnustapa ja sitä kutsutaan myös nimellä paikallisvarasto (Data Mart). Ajan kuluessa tietomäärät kasvoivat ja yhden aihealueen tähtimalli ei enää vastannut liiketoiminnan tarpeita, vaan niitä ryhdyttiin yhdistämään usean aihealueen kokonaisuusiksi (Conformed Data Marts). (Linstedt 2002.)

1990-luvulla tietomäärät kasvoivat edelleen ja myös keskitettyjen tietovarastojen Enterprise Data Warehouse (EDW) käyttö lisääntyi. Tällöin havaittiin 3NF:lla ja tähtimallilla toteutettujen tietovarastojen suorituskykyongelmat, joita käsitellään tarkemmin kappaleessa 2.3 (Aiempien mallinnusmenetelmien ongelmat EDW:ssä). Data Vault -menetelmä kehitettiin poistamaan juuri näitä ongelmia. Menetelmän kehitti Daniel Linstedt vuosisadan vaihteessa ja hän sisällytti siihen 3NF:n ja tähtimallin arkkitehtuurin vahvuudet. Kuluneiden vuosien aikana Data Vault -menetelmä on otettu mielihyvin

vastaan asiantuntijoiden parissa, sillä se on kehitetty juuri tietovarastoinnin tarpeeseen. (Linstedt 2002.)

Tietojen mallinnuksen kehityksessä merkittävä askel oli Data Vault -menetelmä. Siihen on poimittu mallinnuksen peruskomponentit, joiden avulla nykyisissä tietomalleissa esille tullessiin haasteisiin pystytään vastaamaan (Linstedt ym. 2009, 26). Jokaisessa mallinnusmenetelmässä ennen Data Vault -menetelmää on rajoituksensa, etenkin silloin kun niitä hyödynnetään keskitetyissä tietovarastoissa. Tämä johtuu siitä, kun malleja on sovellettava tietovarastointiin, sillä niitä ei ole suunniteltu erityisesti juuri tätä tehtävää varten. Nämä rajoitukset vähentävät niiden käytettävyyttä sekä vaikuttavat yleiseen keskusteluun eri mallinnusmenetelmien käytettävyydestä tietovarastoinnissa. Seuraavaksi tuodaan esille, mitä tarpeita EDW luominen asettaa tietovarastoinnille.

2.2 EDW:n tarpeet tietovarastoinnille

Tietovarastoinnin isänä pidetyn Bill Inmonin suositus on keskitetty Enterprise Data Warehouse (EDW). Se pitää sisällään monen liiketoiminta-alueen yhdenmukaistetut tiedot, koska sen ideana on yhdistää ja integroida organisaation tiedot yhteen tai vain muutama isoon tietokantaan. Tietojen tarkastelunäkökulmana EDW:ssä on yritystaso, joka menee yli sovellus- ja organisaatorajojen. Hyvänä esimerkkinä tästä on eri lähteistä saadut asiakastiedot, joihin liittyvät myynti-, osto- ja muut tapahtumat on pystytty yhdistelemään ja tallentamaan pitkältä ajalta. Kokonaiskuva asiakkaasta ja kaikki asiakkaasta eri järjestelmiin syntynyt tieto on siis helposti hyödynnettävissä yhdestä paikasta tietovarastosta. (Hovi ym. 2009, 27.)

EDW:stä pystyy tekemään raportteja, muodostamaan summatauluja ja muita johdettuja Data Martti -ratkaisuja sekä kuutioita. Nämä rakenteet on suunniteltu siten, että käyttäjien on helppo BI-välineitä hyödyntäen kysellä ja analysoida tietoja. Hovi ym. (2009, 27) toteavat, että puhtaimmillaan EDW:stä ei tehdä kyselyjä juuri ollenkaan, vaan lähes kaikki raportointi ja kyselyt tehdään summaustekniikoiden ja tähtimallin avulla suunnitelluista Data Martteista tai kuutioista. EDW toimii siis operatiivisten tietojen yhdenmukaistamis- ja historiointipaikkana sekä suurten, tarkalla tasolla olevien tietomäärien varastona. (Hovi ym. 2009, 27.)

EDW:hen kohdistuu tiettyjä tarpeita, jotta siitä olisi todellista hyötyä yrityksen toiminnassa. Koska uusia tietolähteitä tulee koko ajan lisää ja liiketoimintatarpeet ja -säännöt muuttuvat, tulee EDW:n olla muutosjoustava ja helposti laajennettavissa. Siten varmistetaan, että EDW:n rinnalle ei synny erillisiä raportointiratkaisuja. (Hovi 2010, 10.)

Järjestelmien monimutkaistuminen, lisääntyvä integrointi, nopeusvaatimukset ja mitä suurimmassa määrin kasvavat tietomäärät tuovat myös lisätarpeita EDW:n ylläpidolle. Tämä vaatii suorituskykyä ja kykyä hallita isoja tietomassoja, jotka on mallinnettu tarkalle tapahtumatasolle. EDW:n tulee mahdollistaa myös kokonaisuuden hallinta yhdestä paikasta, sillä sen avulla eri lähteiden tiedot pystytään integroimaan yhteen. Näin ollen myös tietojen jäljitettävyys on tärkeä, jotta voidaan osoittaa, mistä tiedot ovat tulleet. Lisäksi tarpeita luo tietojen historiointi, vaikka nykyisin myös tietojen reaaliaikaisuustarpeet ovat kasvaneet. (Hovi 2010, 10.)

Tärkeä osa EDW:tä on metadatan hallinta. Metadata eli metatieto on "tietoa tiedosta". Metatiedossa kuvataan tietovaraston eri objektit, joita ovat muun muassa liiketoimintalue, tietokantataulut, tiedot ja näkymät. Metatiedon avulla pystytään selvittämään, onko kulloinkin tarvittava tieto tietovarastossa ja missä siellä se on. Jos metatietoa ei olisi, olisi asian selvittäminen vaikeaa. Metatietoa ajatellaan olevan kolmea erilaista:

- Liiketoimintatieto, joka on tarpeellista tietoa tietovaraston tiedoista tietovaraston käyttäjän näkökulmasta, kuten esimerkiksi johdettujen tietojen kaavoja ja määrittämiä.
- Tekninen metadata, joka sisältää tietoja muun muassa operatiivisista tietolähteistä, latausvaiheista, summataulujen tekovaiheista, eheystarkastuksista ja tietojen muutoksista.
- Prosessimetatieto, joka on tietoa esimerkiksi ETL-latausprosesseista ja tietovaraston käytöstä, jota kerätään yleensä aikaleimalla varustettuna tapahtumalokeihin.

(Hovi, ym. 2009, 43.)

Tietovarastoa toteutettaessa myös valitulla arkkitehtuurilla on merkittävä vaikutus siihen, mikä mallinnusmenetelmä kannattaa valita. Kun toteutetaan EDW-tyyppistä tietovarastoa, niin mallinnusmenetelmänä voi käyttää yleisimmin operatiivisissa tietokannoissa käytettyä ER-mallinnusta sovitettuna tietovarastointiin, koska siinä on huomioitu selkeät rakenteet, varauduttu isoihin massoihin ja historiointiin. Nykyisin vaihtoehtona on myös Data Vault -menetelmä, joka on kehitetty juuri EDW-mallinnukseen ja perustuu Bill Inmonin DW2.0™-arkkitehtuuriin. (Hovi, ym. 2009, 43.)

2.3 Aiempien mallinnusmenetelmien ongelmat EDW:ssä

3NF-tietomallin rajoituksena tietovarastojen suunnittelussa on, että se suunniteltiin alun perin operatiivisten järjestelmien reaaliaikaisen ja taktisen tiedon keruuseen eikä tiedon integrointiin ja historiointiin. Tästä huolimatta 3NF osoittautui aikanaan erittäin skaalautuvaksi, kun kyse oli määrältään suurien tapahtumamäärien käsittelystä. 3NF tukee nopeaa ja suorituskykyistä taktisten kyselyjen indeksointistrategiaa OLTP-järjestelmissä ja juuri siksi se vaikutti aikanaan myös EDW:n toteutukseen. Mutta ajan kuluessa ilmeni, että mallia oli vaikea sopeuttaa käsittelemään tietoja pidemmällä aikajaksolla. Siksi Linstedt ym. (2009, 26) ovatkin sitä mieltä, että 3NF-tietomallia tai mitään sen sovellutusta, ei pitäisi käyttää EDW:n tietojen mallintamiseen. (Linstedt ym. 2009, 26.)

Tähtimallin ongelmana taas on sen rajoittunut ilmaisuvoima yleismallinnuksessa ja EDW:ssä. Tähtimalli, jonka rakennetta kuvataan tarkemmin kappaleessa 5.3 (Nykyinen tietomalli), ei salli dimensio- ja fakta-taulujen välisiä yhteyksiä eikä lapsitauluja. Mallin niin sanotussa yhdenmukaistetuissa ratkaisussa tulee hankaluuksia hallita dimensioita varsinkin silloin, kun ne on yhdenmukaistettu useamman tähden kesken. Toinen näistä yhdenmukaistetuista ratkaisuista on lumihiutalemalli (Snowflake Schema), jossa sekä fakta- että dimensio-taulut on normalisoitu. Tavallisesti tähtimallin dimensiot sisältävät toisteisuutta eivätkä näin ollen ole 3NF-muodossa. Lumihiutalemallissa myös dimensiot ovat normalisoituja ja siksi sen rakenne muistuttaa lumihiutaletta. Toinen vaihtoehto dimensioden yhdenmukaistamiseen on jo aiemmin mainittu usean aihealueen ratkaisu (Conformed Data Marts), jossa useamman tähden dimensiotaulut ovat yhteisiä, joista esimerkkinä mainittakoon asiakas- ja tuotedimensiot. Ongelmana kuitenkin näiden yh-

denmukaistettujen ratkaisujen historioinnissa on tilanne, jossa sarakkeita on paljon ja vain osa niistä muuttuu. Tällöin joudutaan toistamaan turhaan monia tietoja (aina kokorivin tiedot). Tästä seuraa, että EDW:n muutosjoustavuus heikkenee laajenemisen myötä. (Hovi 2010, 7, 9, 23; Hovi ym. 2009, 28; Hovi, Huotari & Lahdenmäki 2005, 139.)

Tähtimallia hyödynnetään myös tietovarastoissa, jotka on rakennettu yhdenmukaistettujen, Data Martteihin perustuvan Warehouse Bus -arkkitehtuurin mukaisesti. Tämä Kimballin suosittama arkkitehtuuriratkaisu sisältää useita Data Martteja, joihin on mallinnettu tiettyjen liiketoiminta-alueiden tiedot (myynti, tuotteet jne.). Data Martit sisältävät pääasiassa dimensio- ja faktatauluja. Faktataulut voivat sisältää joko yksittäisiä tietoja tai yhteen laskettua dataa. Dimensioista esimerkkinä voidaan mainita tuote- ja asiakasdimensiot. Warehouse Bus -arkkitehtuuri ei kuitenkaan ole sama asia kuin keskittetty tietovarasto, vaan siinä Data Martit on rakennettu yhteismitallisiksi (conformed) useamman Data Martin kesken. Tällöin dimensiot, jotka on jaettu erityisellä tavalla faktataulujen tai Data Marttien kesken, ovat yhteisiä. Kuten edellä on jo tullut ilmi, niin myös Warehouse Bus -arkkitehtuurilla rakennetussa tietovarastossa tulee esiin tähtimallissa esiintyvät ongelmat. Etenkin rakenteen joustamattomuus, jolloin tietovaraston laajennettavuus ei pysty vastaamaan kasvaviin liiketoimintatarpeisiin. (Hovi 2009, 28.)

Lisäksi tähtimalli ei alun perin sisältänyt historiointia, sillä se oli tarkoitettu vastaamaan vain tiettyjen aihealueiden "nyt"-kysymyksiin sekä tilannekohtaisiin (snapshot) kysymyksiin kyselyhetken tiedosta. Myöhemmin tähtimallin käytön laajentuessa siihen kuitenkin lisättiin historioinnin kolme ensimmäistä astetta, jotka ovat: type 1 (muutokset kirjoitetaan päälle), type 2 (historioidaan koko dimensiotietue) ja type 3 (historioidaan vain osa dimensiotietueen sarakkeista). Malliin tehdyistä muutoksista huolimatta se ei ratkaissut kaikkia ongelmia. Linstedt ym. (2009, 26) vertaavatkin tähtimallia pyöreään tappiin neliskulmaisessa reiässä silloin, kun sitä käytetään EDW:n mallinnuksessa. (Linstedt ym. 2009, 26.)

Linstedt ym. (2009, 26) ovat sitä mieltä, että 3NF:n ja tähtimallin sovittamisesta tietovarastointiin seuraa sivuvaikutuksia, jotka ovat:

- Skaalautuvuuden puute, jolloin tietovarasto ei ole helposti laajennettavissa.

- IT-ketteryyden puute, jolloin puuttuu kykyä vastata uusiin liiketoimintatarpeisiin oikea-aikaisesti.
- Joustamattomuus, joka seuraa tietojen yhdenmukaistamisperiaatteesta.

(Linstedt ym. 2009, 26.)

Myös Hovi (2010, 12) painottaa, että mikäli laajennettavuus ja muutosjoustavuus ovat hitaita, niin EDW:n rinnalle alkaa syntyä erillisiä Data Martteja tai muita raportointiratkaisuja. Siksi EDW:n ylläpitämiseen tarvitaan ketteryyttä, jonka Data Vault -menetelmä mahdollistaa. Se on eräänlainen hybridi kummastakin mallista ja siinä hyödynnetään molempien mallien parhaimmat puolet, jotta voitettaisiin edellä mainitut puutteet EDW:n saavuttamiseksi. (Hovi 2010, 12; Linstedt ym. 2009, 26.)

2.4 Data Vault –menetelmä

Tässä luvussa käydään tarkemmin läpi Data Vault -menetelmän osia ja rakennetta. Lisäksi luvun lopussa tuodaan esille, miten Data Vault liittyy tietovarastoinnin uuden polven DW2.0™-arkkitehtuuriin sekä mitä hyötyjä saavutetaan Business Data Vault -ratkaisun (BDV) käytöstä liiketoimintasäännöillä muokatun informaation hyödyntämisessä.

Data vault - termi on moniselitteinen. Termin suora käänös tarkoittaa tietoholvia.

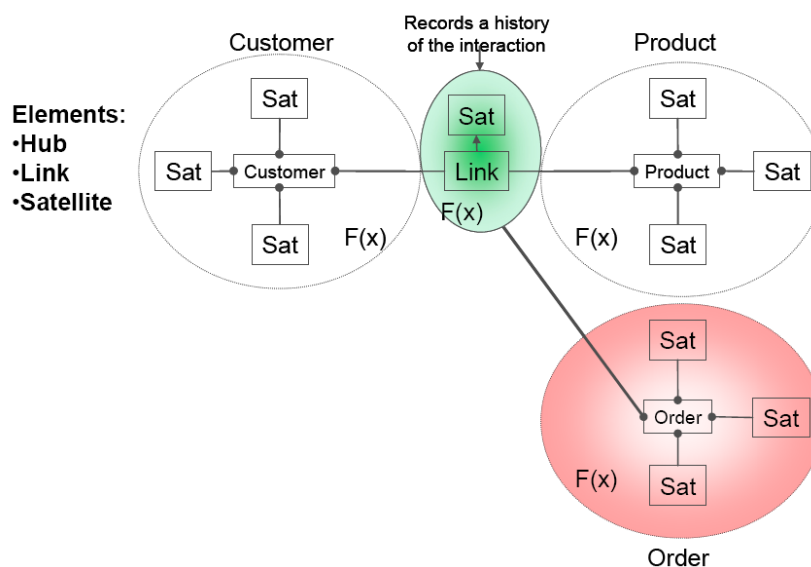
Tässä tutkimustyössä käytän termiä siten kuin Linstedt sen määrittelee:

The Data Vault is a detail oriented, historical tracking and uniquely linked set of normalized tables that support one or more functional areas of business. It is hybrid approach encompassing the best of breed between 3rd normal form (3NF) and star schema. The design is flexible, scalable, consistent and adaptable to the needs of the enterprise. It is a data model that is architected specifically to meet the needs of enterprise data warehouses. (Linstedt 2002.)

Data Vault -menetelmän kehittäjä Daniel Linstedt (2011a) selventää kotisivullaan olevassa artikkelissa, että Data Vault on vapaasti käytettävissä oleva mallinnusarkkitehtuuri - ei siis erillinen tuote. Se on avoin ja käytettävissä julkisesti. Menetelmän vaihtoehtoi-

nen nimi on "Common Foundational Integration Model Architecture" ja kuten 3NF ja tähtimalli, niin myös Data Vault sisältää sääntöjä ja standardeja. Hän painottaakin, että Data Vault -mallinnustekniikassa on kyse menetelmästä - lähestymistavasta, joka on mallistandardi, kun rakennetaan tai kuvataan fyysistä tietomallia EDW:lle. (Linstedt 2011b.)

Data Vault -menetelmä sisältää muutaman pääkomponentin: hubit (Hub), linkit (Link) ja satelliitit (Satellite) (Linstedt 2002). Menetelmää voi yksikertaisimmillaan kuvata näkymänä hermoverkkoon hermoineen, tuojahaarakkeineen ja synapseineen, jolloin sen osista hubit ja hubien satelliitit ovat hermoja, linkit niiden merkityksestä riippuen tuojahaarakkeita ja synapseja (Wikipedia). Kuviossa 2 havainnollistetaan komponenttien välisiä yhteyksiä.



Kuvio 2. Esimerkki Data Vault -menetelmästä (Linstedt, Graziano & Hultgren 2009, 3)

Data Vault -menetelmän etuna on, että se seuraa kaikkia Bill Inmonin asettamia määrittäjätyksiä tietovarastoinnille paitsi yhtä; Data Vault pohjautuu toimintoihin eikä aihealueisiin. Tämä tarkoittaa, että siinä käytetään luonnollisia liiketoiminta-avaimia, jotka mahdollistavat näkyvyyden tietoon yli liiketoimintatarajojen. Mallissa hubit ja linkit ovat runko, johon sisältö eli satelliitit asetetaan. (Linstedt 2011c, 39.)

Menetelmä keskittyy siis yritykselle tärkeiden liiketoiminta-alueiden ympärille siten, että hubit edustavat rakenteen yksilöiviä pääavaimia (luonnollisia perusavaimia), jotka ku-

vaavat toiminnan todellisia kohteita, kuten asiakas, tuote ja tilaustapahtuma. Pääavaimet ovat linkki liiketoiminnan prosessien ja tietojen välillä. Linkit edustavat hubien välisiä yksittäisiä yhteyksiä. Satelliitit ovat hubien kohteiden sisältöä (attribuutteja), kuten nimi, osoite, tuotekoodi ja tapahtuman perustiedot. Satelliitit sisältävät lisäksi sisältötiedon historioinnin. (Linstedt ym. 2009, 3, 34.)

Data Vault -komponenteilla on yhteisiä tietokenttiä. Tällaisia tietokenttiä ovat surrogaatit eli keinotekoiset avaimet, joilla korvataan luonnolliset perusavaimet. Surrogaatit voivat olla hubin, linkin ja mahdollisesti myös satelliitin perusavaimina. Myös latauspäivämäärä on yhteinen ja löytyy hubista, linkistä ja satelliitista (osa pääavainta). Lisäksi kaikilla komponenteilla on jäljitettävyyden mahdollistamiseksi tieto lähteestä, jotta saadaan selville, mistä järjestelmästä tieto on tullut. Loppupäivämäärä on pakollinen ainoastaan satelliiteissa, jotta sen sisältämät attribuuttitiedot voidaan versioda. Seuraavaksi käydään tarkemmin läpi eri Data Vault -komponentteja sekä sitä, mistä tietokentistä ne koostuvat. (Hovi 2010, 15.)

2.4.1 Hub-entiteetit

Hub-entiteetit eli hubit sisältävät minimissään yksilöivän listauksen liiketoiminta-avaimista. Näitä avaimia liiketoiminta käyttää toiminnassaan päivittäin, joista esimerkiksi voi mainita laskunumeron, työntekijänumeron ja asiakasnumeron. Mikäli liiketoiminta ei pysty käsittelemään liiketoiminta-avaimia, niin samalla se kadottaa yhteyden avaimen liittyvään asiasisältöön ja sitä ympäröivään tietoon. Hubiin liittyvät muut attribuutit ovat seuraavat:

- Surrogaattiavain, joka ei ole välttämätön. Voi sisältää lisäavaimen tai juoksevan numeron.
- Latausajankohta, joka kertoo, milloin liiketoiminta-avain luotiin tietovarastoon.
- Lähdejärjestelmä, josta selviää, mistä järjestelmästä tieto on lähtöisin. Liittyy tiedon jäljitettävyyteen.

(Linstedt 2002.)

2.4.2 Link-entiteetit

Link-entiteetit eli linkit kuvaavat fyysisen monen suhde moneen 3NF-yhteyden, jolloin ne sisältävät yhden tai useamman liiketoiminta-avaimen yhteyden toisiinsa tai niiden välisen tapahtuman. Linkit koostuvat seuraavista attribuuteista:

- Surrogaattiavain, joka ei ole välttämätön. Voi sisältää lisäavaimen tai juoksevan numeron. Numero tarvitaan vain silloin, jos linkki yhdistää useamman kuin yhden hubin tai jos yhdistetty pääavain voi aiheuttaa suorituskykyongelmia.
- Yhden tai useamman hubin avaintiedon. Linkki yhdistää pääsääntöisesti hubeja toisiinsa.
- Latausajankohta, joka kertoo, milloin yhteys tai tapahtuma luotiin ensimmäisen keran tietovarastoon.
- Lähdejärjestelmä, josta selviää, mistä järjestelmästä tieto on lähtöisin. Liittyy tiedon jäljitettävyyteen.

(Linstedt 2002.)

2.4.3 Satellite-entiteetit

Satellite-entiteetit eli satelliitit sisältävät hub-avaimen kuuluvan sisällön liiketoiminnan kohteesta. Kaikki siinä oleva tieto voi muuttua ajan kuluessa, joten rakenteen tulee olla sellainen, joka mahdollistaa muuttuneen tiedon tallennuksen tarvittavalla rakeisuudella eli tietojen päivityssyklillä. Satelliitit koostuvat seuraavista attribuuteista:

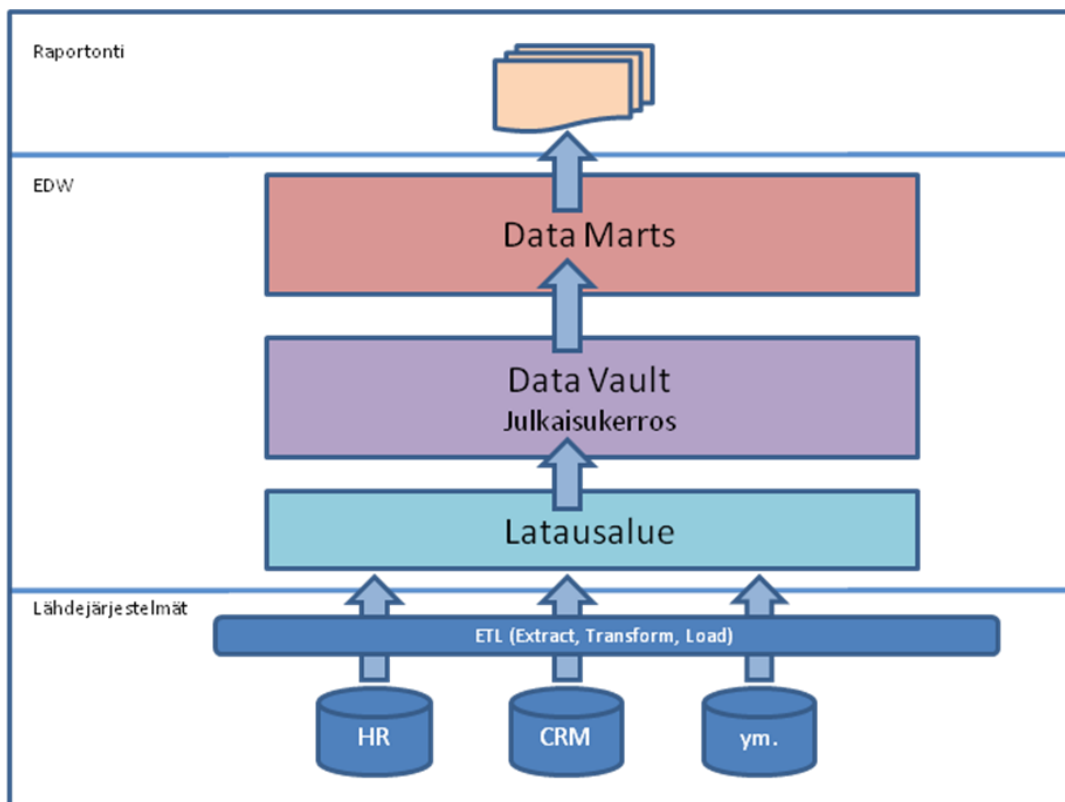
- Osa satelliitin pääavainta on hubin tai linkin pääavain, joka yhdistää satelliitin hubiin tai linkkiin.
- Osa satelliitin pääavainta on myös latausajankohta, joka kertoo, milloin kohteen sisältöön liittyvä tieto on siirretty tietovarastoon (tietokantaan lisätään aina uusi tietorivi).
- Satelliitin vaihtoehtoinen pääavain on surrogaattiavain juoksevalla numeroinnilla, joka mahdollistaa satelliitin, jossa on useampia arvoja ja näiden arvojen pitämisen järjestyksessä ja ryhmittelemisen sekvenssinumerolla.

- Lähdejärjestelmä, josta selviää, mistä järjestelmästä tieto on lähtöisin. Liittyy tiedon jäljitettävyyteen.
- Yksittäiset tietokentät sisältötiedoista.

(Linstedt 2002.)

2.4.4 Data Vault -tietovaraston rakentaminen

Data Vault -menetelmän etuna on sen liiketoimintaa tukeva selkeä modulaarinen rakenne. Lisäksi se siirtää yhteistyön IT:n ja liiketoiminnan välillä lähemmäksi liiketoimintaa, jolloin muutostarpeiden toteuttaminen nopeutuu. Data Vault -menetelmä mahdollistaa skaalautuvan ja iteratiivisen tietovarastokehityksen. Sen avulla myös tietojen lataus ja hallinnointi helpottuu, jonka mahdollistaa erillinen julkaisukerros, joka on nähtävissä kuviossa 3. Lisäksi konversio nykyisestä joko 3NF-relaatiokannasta tai tähtimallilla toteutetusta kannasta Data Vault -menetelmään on hyvin helppoa (Linstedt).



Kuvio 3. Esimerkki Data Vault -arkkitehtuurilla toteutetusta EDW:stä

Edellä läpikäytyjen peruskomponenttien lisäksi Data Vault -mallinnettu tietokanta sisältää myös muita tietokannan hallintaan ja suorituskykyyn liittyviä tauluja, kuten Stand-alone -, Point in Time - ja Bridge-tiloja. Stand-alone -tilat ovat yhteiskäyttöisiä tiloja, joilla ei ole viittauksia muihin tiloihin, vaan ne sisältävät tietoa esimerkiksi kalenteriin, aikaan ja koodeihin liittyen. Tietokannan suorituskyvyn tehostamiseksi tarvitaan Point in Time - ja Bridge-tilat, joita käytetään vain tietokantakyselyjen nopeuttamiseen. Tutkimuksen kannalta tärkeimpiä ovat kuitenkin Data Vault -rakenteen perustaulut hubi-, satelliitti- ja linkkitaulut, jotka muodostavat perustavaa laatua olevan muutoksen verrattuna aiempiin tietokantojen mallinnusmenetelmiin. (Hovi 2010, 36, 37; Linstedt 2002.)

Tietovarastoa mallinnettaessa Data Vault -menetelmällä tietokannan rakentamisessa tulee edetä tietyssä järjestyksessä. Etenemisen eri vaiheet ovat seuraavat:

1. Kuvataan liiketoiminnan käsitteet, jotka ovat tulevia hubeja ja määritellään niiden avaintiedot. Tämä vaihe vaatii liiketoimintaymmärrystä.
2. Kuvataan tietojen väliset suhteet, joista syntyvät linkit. Ne yhdistävät toisiinsa hubit eli liiketoiminta-avaimet.
3. Kuvataan hubien attribuutit, joista muodostetaan satelliitit. Nämä antavat liiketoiminta-avaimille ja niitä yhdistävien linkkien tapahtumille sisällön ja hoitavat tietojen historioinnin.
4. Ryhmitellään tietosisältö eli satelliittien tiedot päivityssykkien mukaan omiin tauluihinsa.
5. Kuvataan mahdolliset Stand-alone ja Point in Time -tilat.

(Hovi 2010, 36; Linstedt 2002.)

Ladattaessa tietoja Data Vault -mallinnettuun tietokantaan ei latausajojen järjestyksellä ole yhtä suurta merkitystä kuin esimerkiksi tähtimallisessa tietokannassa. Tavoitteena on, että Data Vault -lataukset pidetään yksinkertaisina ja tehokkaina, jolloin ne sisältävät ainoastaan tietojen lisäyksiä. Päivityksiä tehdään vain loppupäivämääriin. Lataukset ovat yleensä nopeita, rinnakkaisia, uudelleenaloitettavia, skaalautuvia ja partioituja (osittettuja). Latausajoissa edetään siten, että kaikki hubit ladataan ensimmäisinä. Seuraavak-

si ladataan hubien satelliitit, jotka voidaan halutessa ladata vaikka rinnakkain, sillä niillä ei ole viite-ehyksiä keskenään. Linkit, jotka ladataan viimeisenä, voidaan ladata vasta, kun kaikki niiden tarvitsema pohjatieto on jo olemassa. Lisäksi Data Vault -menetelmä hallitsee myös kantaan eri tahdissa ladatut tiedot esimerkiksi, kun laskutietoja ladataan ilman asiakkaita tai tilausrivejä ilman tuotteita. Tätä ominaisuutta käydään läpi tarkemmin kappaleessa 4.6 (Viite-ehyys Data Vault -mallinnetussa tietovarastossa). (Hovi 2010, 39.)

Rakennettaessa tietovarastoa Data Vault -menetelmällä tulee muistaa, että tiedot muunnetaan liiketoiminnan ymmärtämään muotoon vasta julkaisukerroksessa, jota Data Martti -raportointikerros hyödyntää. Data Vault -latauskerros, joka sisältää raakadatan, ei huomioi, miten tietoja käytetään tai analysoidaan eikä ota kantaa siihen, mitkä tiedot ovat oikein tai mitkä väärin. Tietovarastoon ladattua raakadataa ei siis ole laatu-tarkastettu eikä puhdistettu eikä se sovellu loppukäyttäjän kyselyjen kohteeksi. Vaikka tiedot olisivat virheellisiä, niitä ei muuteta, jotta säilytetään jäljitettävyyttä. Data Vault -raakadakerros sisältää paljon liitoksia, mikä edellyttää huolellista Data Marttien suunnittelua, sillä laaduntarkastusvastuu siirtyy Data Martti -raportointikerrokseen. (Hovi 2010, 30, 41.)

Linstedt (2002) suosittelee että, mikäli halutaan varmistaa, että Data Vault -mallinnettu tietovarasto on helposti ylläpidettävissä ja joustavasti muutettavissa, on noudatettava tiettyjä mallinnussääntöjä. Nämä säännöt ovat seuraavat:

- Yhteydet hubien välillä ovat sallittuja vain linkkien kautta.
- Hubien perusavain (Primary Key, PK) on oma surrogaattinsa.
- Hubeissa ei ole viiteavaimia (Foreign Key, FK) eikä mitään kuvaavaa tietoa (attributteja).
- Hubien liiketoiminta tai pääavain ei koskaan muutu.
- Latausten yksinkertaistamista tukee se, että linkkirakenteessa pitää olla vähintään kaksi viiteavainta.
- Linkkien perusavain on yleensä surrogaatti.
- Linkki voi yhdistää useita hubeja.

- Linkkirakenteella hoidetaan rekursiiviset (itseensä viittaava) yhteydet.
- Yhteys kahden linkin välillä on mahdollinen, mutta sitä ei suositella.
- Linkkien voimassaolon toteutukseen suositellaan satelliittia.
- Satelliiteilla pitää aina olla yksi isätaulu (hubi tai linkki).
- Satelliitin yksilöivä pääavain (PK) sisältää aina isätaulun pääavaimen ja latauspäivämäärän.
- Satelliitilla ei voi olla muuta viiteavainta kuin pääavaimen osana oleva isätauluun viit-taava viiteavain.
- Yhteen hubiin voi liittyä useita satelliitteja, jotka on jaettu osiin esimerkiksi muutos-tiheyden tai tietolähteen mukaan.
- Yleensä tapahtumatyypisistä tauluista tulee linkki.
- Hubit, satelliitit ja linkit sisältävät aina latauspäivän ja tietolähteen (source).

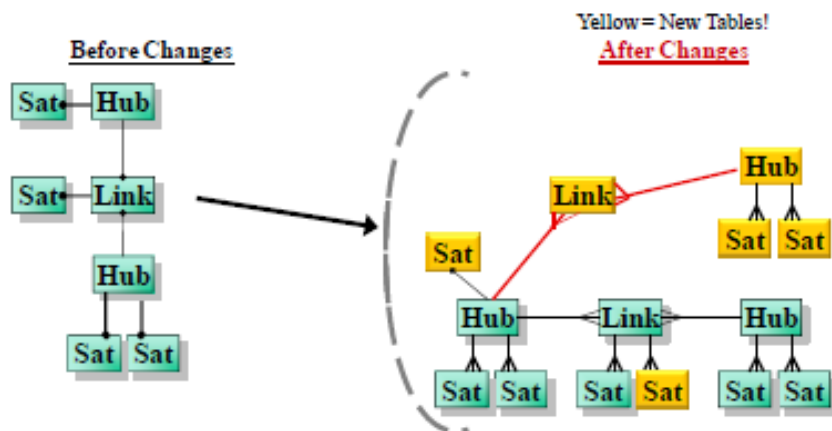
(Hovi 2010, 18; Linstedt 2002.)

2.4.5 Data Vault -tietokannan skaalautuvuus

Data Vault -menetelmän etuja on sen skaalautuvuus eli laajennettavuus. Uusien kokonaisuuksien, kuten liiketoimintayksiköiden ja erilaisten datalähteiden lisääminen malliin on helppoa. Data Vault -menetelmällä toteutettu tietokanta skaalautuu inkrementaalisesti, jolla tarkoitetaan sitä, että uuden alueen lisääminen ei aiheuta muutoksia olemassa olevaan tietokantarakenteeseen (Hovi 2010, 12).

Data Vault -menetelmällä mallinnettua tietokantaa pystytään siis muuttamaan ilman, että mihinkään olemassa olevaan latausrutiiniin jouduttaisiin tekemään uudelleensuunnittelua (Re-engineering). Alla olevasta kuvioista 4 näkyy, miten keltaisella merkityt mallin komponentit on lisätty uuden liiketoiminta-alueen avaimiksi, attribuuteiksi ja yhteyksiksi. Jotta tämä uusi data saadaan osaksi olemassa olevia rakenteita, aiheutuu siitä uutta koodia, mutta ei kuitenkaan vanhan koodin uudelleensuunnittelua tai olemassa olevien latausprosessien uudelleen asennusta. Näin toimimalla varmistetaan, että muutokset liiketoimintasääntöihin eivät vaikuta Data Vault -mallinnettuun tietokantaan li-

säämällä monimutkaisuutta sen latausprosesseissa. Juuri tämä tekee Data Vault -mallista, mitä ketterimmän ja joustavimman laajennuksille. (Linstedt ym. 2009, 60.)



Kuvio 4. Data Vault -malliin tehdyt muutokset (Linstedt ym. 2009, 60)

Oleennaista on, että mitä enemmän muutokset liiketoimintasäännöissä vaikuttavat EDW:hen, niin sitä todennäköisempää on, että uudelleensuunnittelun myötä IT-asiiantuntijat joutuvat vastaamaan muutoksiin, joista seuraa lisäkuluja ja pidempiä toimitusaikoja. Tämän seurauksena liiketoiminta joutuu tilanteeseen, jossa he mieluummin ohittavat IT:n ja alkavat luoda omia ratkaisujaan (suorat tietokantayhteydet ja Excel-taulukot). Lisäksi laajat muutokset liiketoimintasääntöihin ovat haasteellisia, kun kyseessä on isot massat ja saantiajat (latenssi). Data Vault -menetelmä tuo vastauksen tähän tilanteeseen joustavan laajennettavuuden kautta, jolloin se mukautuu nopeasti yrityksen liiketoiminnan muutoksiin. Tämä lisää osaltaan yhteistyötä liiketoiminnan ja IT-asiiantuntijoiden välillä ja mahdollistaa yhteisten tavoitteiden asettamisen EDW:n käytölle. (Linstedt ym. 2009, 18.)

2.4.6 Data Vault -menetelmän tuomat muutokset EDW:lle

Liiketoiminnan monimuotoisuus tuo haasteita EDW:n ylläpitoon. Kompleksisuudet ladattavissa tiedoissa ja lähdejärjestelmien rutiineissa lisäävät kustannuksia ja hidastavat ylläpidon toteutusta. Kun lähdetään selvittämään, mistä nämä ongelmat johtuvat - varsinkin jo muutaman vuoden käytössä olleesta EDW:stä, niin syy on hyvin yksinkertainen - huonosti toteutettu tietoarkkitehtuuri ja puuttuvat standardit. Linstedtin ym.

(2009,1) mukaan tämä onkin ensisijainen syy EDW:n ongelmiin tänä päivänä. (Linstedt ym. 2009, 1.)

EDW:n suunnitteluun tuo helpotusta Data Vault -menetelmä, joka perustuu Inmonin ja Kimballin malleihin ja on koottu niiden parhaista paloista. Se on ensimmäinen ja vain tietovarastointiin kehitetty menetelmä, joka johtaa skaalautuviin ja muutosjoustaviin tietovarastoratkaisuihin. Se tukee hyvin tietojen historiointia ja parantaa tietovaraston suorituskykyä. Menetelmässä on panostettu erityisesti tietojen jäljittämiseen ja auditointiin. Data Vault -menetelmä tukee myös liiketoiminnan ja tietohallinnon yhteisten tavoitteiden saavuttamista, sillä se mukautuu nopeasti liiketoiminnan muutoksiin. Sen avulla tietovaraston muutoshallinta helpottuu ja näin ollen voidaan löytää mahdollisia ongelmakohтия liiketoiminnasta, joita aiemmin ei ole havaittu. Lisäksi uusien liiketoiminta-alueiden ja erilaisten tietolähteiden lisääminen Data Vault -menetelmällä mallinnettuun kantaan on helppoa, sillä menetelmässä osa muutoksen kohteista on siirretty menetelmän sisältöön menetelmän rakenteen sijaan. (Martin 2008.)

Data Vault -menetelmä kytkeytyy myös yrityksen toimintaprosesseihin, sillä se on menetelmä tietokannan suunnitteluun, joka mahdollistaa useasta lähdejärjestelmästä kootavan tiedon varastoinnin siten, että kaikki tietovarastoon tallennetut tiedot voidaan jäljittää lähdejärjestelmään. Lisäksi menetelmä on suunniteltu joustavaksi liiketoimintamuutoksille. Tämä on mahdollista, koska menetelmässä tietomallin suunnittelu aloitetaan liiketoiminta-alueista olettaen, että ne muuttuvat harvemmin kuin liiketoimintaa tukevat operatiiviset järjestelmät (Wikipedia).

Linstedt ym. (2009, 2) ovatkin sitä mieltä, että Data Vault -menetelmä on käyttöönottosuositus EDW:n ja integroinnin haasteille, sillä menetelmän juuret ovat monissa hyvin tunnetuissa toiminnan parantamisen menetelmissä, kuten CMMI taso 5 (Capability Maturity Model Integration) ja Six Sigma. Heidän mielestään jokaisen hyvän menetelmän taustalla on tiedonmallintamisen arkkitehtuuri. Data Vault -menetelmä tarjoaa yksinkertaisen vastauksen helppoon ja selkeään toimialapohjaiseen ratkaisuun, jonka lopputuloksena on skaalautuva, joustava, toistettava ja automatisoitu EDW-ratkaisu. (Linstedt ym. 2009, 2.)

Etenkin keskusteltaessa liiketoiminnan kanssa raporttivaatimuksista on Data Vault -menetelmä osoittautunut ketteräksi tavaksi vaatimusten kokoamiseen. Koska menetelmässä liiketoimintasäännöt viedään osaksi EDW-ratkaisua, mahdollistaa se niin sanottuun raakadataintegraation, jolloin liiketoiminnan on helpompi nähdä, mitä tietoja heidän järjestelmissään todella on. Tietovarastoasiantuntijat pystyvät nopeasti luomaan raakadatasta "mitä-on" -tähtimalleja, joista liiketoiminnan on helpompi hahmottaa käytettävissä olevat tiedot. Tähtimalleja käytetäänkin, jotta liiketoiminta ymmärtäisi paremmin, mistä on kyse. Vasta kun tarpeet on selvitetty ja lähdetään yhdessä työstämään lopullista tähtimallia, on liiketoiminnan paljon helpompi kertoa, mitkä tiedot ovat väärin tai turhia ja mitkä tiedot puuttuvat. Tällainen lähestymistapa tehostaa IT:n kykyä vastata liiketoimintatarpeisiin ja antaa tietovarastoasiantuntijoille mahdollisuuden toimittaa nopeammin tähtimalleja, jotka ovat todenmukaisempia ja liiketoimintalähtöisempiä. Näin aikaa jää enemmän erojen analysoimiseen ja lopullisen tähtimallin seuraavien toteutusiteraatioiden tekemiseen. (Linstedt ym. 2009, 48.)

Menetelmään liittyy kuitenkin myös joitakin ongelmia ja se vaatii rakennusvaiheessa enemmän työtä. Lisäksi Data Vault ei tue OLAP-käsittelyä, joten sillä mallinnettuun tietokantaan on hankala tehdä suoria kantakyselyjä. Liiketoiminnan käyttöön tarkoitettuja Data Martteja rakennettaessa onkin otettava huomioon, että Data Vault -tietokanta sisältää paljon liitoksia. Varsinkin loppukäyttäjän näkökulmasta se edellyttää hyvin suunniteltuja Data Martteja, jotta he pystyvät hyödyntämään tietovarastoa tehokkaasti. Hovi (2010, 42) korostaakin, että Data Vault -menetelmässä on siirretty monia ETL-latauksessa aiemmin tehtyjä asioita Data Marttien lataukseen. (Hovi 2010, 42.)

Myös Martin (2008) toteaa, että Data Vault -mallilla toteutettu tietovarasto ei sellaisenaan sovellu käyttäjien kyselyalustaksi, sillä sen rakenne on normalisoitu, josta seuraa, että kyselyt muodostuvat monimutkaisiksi. Mallin käyttö edellyttää paikallisvarastoja eli Data Martteja. Sen sijaan tiedon louhinnan (Data Mining) alustaksi Data Vault -tietovarasto sopii mainiosti. (Martin 2008.)

2.4.7 Data Vault ja DW2.0™

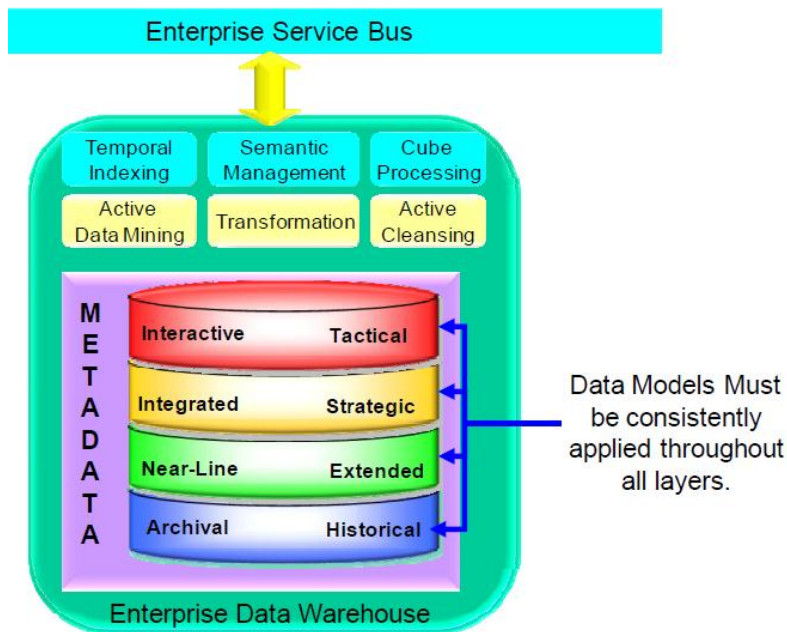
Nykyisin EDW:tä suunnitellessa tulisi ottaa huomioon, että yleensä tiedot, jotka sinne tuodaan, säilytetään myös siellä. Tietokannan koko vain siis kasvaa ja vanhimpia tapahtumatietoja on jossain vaiheessa poistettava, siksi myös tietovaraston hallinta on hyvä suunnitella huolellisesti etukäteen. Tässä kappaleessa käydään läpi Inmonin kehittämää uuden polven DW2.0™-arkkitehtuuria, joka on yksi vaihtoehto EDW:n toteutukseen.

Aiemmin on tullut jo ilmi, että Data Vault on yksi tiedon mallinnuksen menetelmistä, joita yleensä käytetään tänä päivänä ja jotka auttavat kohtaamaan nykyisiä haasteita tietovarastoinnissa. Data Vault -menetelmä on tarkoitettu ja suunniteltu juuri EDW-mallinnukseen perustuen Bill Inmonin DW2.0™-arkkitehtuuriin. Data Vault -menetelmää kehitettiin kymmenisen vuotta R&D-työnä (Research and Development) vuodesta 1990 vuoteen 2000. Se on ollut vapaasti hyödynnettävissä jo vuodesta 2000 lähtien ja on nyt myös "tietovarastoinnin isän" ja DW2.0™-arkkitehtuurin luojaan hyväksymä. Linstedt ym. (2009, 48) mainitsevatkin teoksessaan *The New Business Supermodel* Inmonin itse sanoneen kesäkuussa 2007, että Data Vault -menetelmä on optimaalinen valinta, kun EDW:tä mallinnetaan DW2.0™ arkkitehtuuriviitekehukseen. (Linstedt ym. 2009, 48.)

Inmonin kehittämä uuden polven DW2.0™-arkkitehtuuri tuo valtaviin tietomassojen hallintaan sektorijajattelun. Siinä tietovarastokanta jaetaan neljään eri kerrokseen eli sektoriin (kuvio 5):

1. Interaktiivinen sektori, joka on lähes ajan tasalla (Interactive).
2. Integroitu sektori, joka sisältää alle 3 vuoden tiedot päivästä kuukauteen -tasolla (Integrated).
3. Lähes suorakäyttöinen sektori, jossa on 3 - 4 vuotta vanhat tiedot (Near-Line).
4. Arkistosektori, jossa on yli viisi vuotta vanhat tiedot (Archival).

(Hovi 2009, 30.)



Kuvio 5. DW2.0™ tietovarastosektorit (Linstedt ym. 2009, 70)

Interaktiiviseen sektoriin, joka tukee taktista päätöksentekoa, voidaan tuoda osa operatiivisten järjestelmien tiedoista lähes sellaisenaan. Tiedot ovat sovelluskohtaisia, sillä aikaa ei ole integroimiseen. Hyvä esimerkki on myyntitapahtuma, joka siirretään interaktiiviselle sektorille heti, kun se on tallennettu operatiiviseen järjestelmään. Näin se on heti hyödynnettävissä operatiiviseen raportointiin. (Hovi ym. 2009, 30.)

Perustietovarastoa vastaa integroitu sektori, jonne ETL-prosessilla viedään ja integroidaan operatiivisten järjestelmien tietoja. Integroitu sektori tukee ensisijaisesti strategista päätöksentekoa. ETL-prosessilla tuodaan ja integroidaan tiedot myös interaktiiviselle sektorille. (Hovi ym. 2009, 30.)

Tiedot, jotka eivät enää ole niin ajankohtaisia tai joita ei oletettavasti kysellä enää kovin usein, viedään lähes suoraan käyttöiselle sektorille. Tämä mahdollistaa tietojen tallentamisen halvemmalle medialle kuin integroidun sektorin tiedot. Lähes suoraan käyttöisen sektorin tiedot siirretään integroidulle sektorille vasta, kun niitä tarvitaan. Näin harvemmin tarvittavat vanhemmat tiedot ovat saatavilla, mutta halvemmalla vasteajalla, ja integroitu sektori pysyy pienempänä ja suorituskykyisempänä. Myös arkistointisektori tehostaa DW2.0™ suorituskykyä, sillä sinne tallennetaan tiedot, joita todennäköisesti tarvitaan

todella harvoin. Kyseessä on vanhempi tieto, josta liiketoiminta ei ole enää kiinnostunut ja niitä tarvitaan vain lainsäädännöllisistä syistä. (Hovi ym. 2009, 31.)

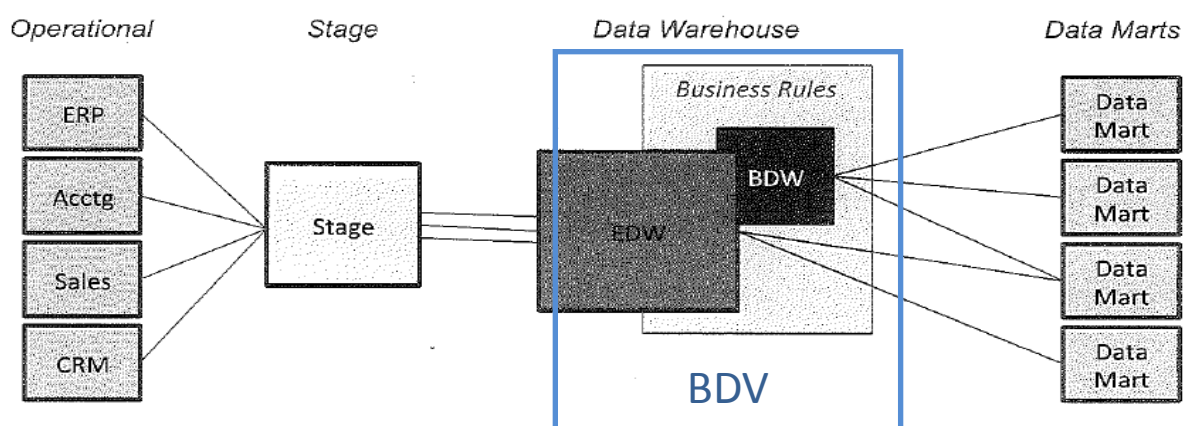
Ilmeistä on, että tietovarastoarkkitehtuurissa myös sektoriajattelu tehostaa keskitetyn tietovaraston toimintaa, sillä näin saadaan tietovaraston aktiivinen osa pidettyä pienenä ja tehokkaana. Nykyisin tietojen määrä lisääntyy koko ajan, joka kasvattaa tietovaraston kokoa ja hidastaa sen käyttöä. Lisäksi osa tiedosta vanhenee ja menettää merkityksensä. Tietojen kerrostaminen sektoreihin eri käyttötarpeen mukaan mahdollistaa myös vanhempien tietojen käytettävyyden tehokkaasti, kun niitä tarvitaan. Hovi ym. (2009, 31) kuitenkin muistuttavat, että tietovarastoarkkitehtuuria suunnitellessa tulee ottaa huomioon, että DW2.0TM mukaisten sektoreiden toteuttaminen myös monimutkaistaa tietovarastoympäristöä.

2.4.8 Business Data Vault

Yksi vaihtoehtoinen tapa EDW:n hyötyjen lisäämiseksi on Business Data Vault (BDV). Se on Linstedin suosittelema ja viime vuosina keskustelua herättänyt tapa, jossa hyödynnetään Data Vault -menetelmää myös liiketoimintasäännöillä muokattuun informaatioon. Olennaista BDV:ssa on, että myös tietojen liiketoimintakerros on toteutettu Data Vault -menetelmää hyödyntäen ja koostuu menetelmän peruskomponenteista: hubeista, satelliiteista ja linkeistä. Näin ollen BDV:ssa sijaitseva informaatio on liiketoimintasäännöillä muokattua dataa raakadatakerroksesta, jolloin se on yhteen laskettua, auditoitavaa ja soveltuu hyvin EDW:n vaatimuksiin. (Linstedt 2011a.)

BDV:ssa tieto on laatutarkastettua, laaturpuhdistettua, jalostettua, yhtenäistä, yhteen laskettua ja yhdisteltyä sekä lisäksi liitettävissä metadata ontologioihin. Periaatteessa BDV sisältää liiketoiminnalle tärkeän jälkikäsitellyn informaation ja datasetit, jotka sisältävät merkinnät sen hetken datavirheistä. BDV-ratkaisussa noudatetaan melkein kaikkia samoja sääntöjä ja standardeja kuin raakadatakerroksessa. Tämä tarkoittaa sitä, että siinä esitetään liiketoiminta-avaimet hubeissa, yhteydet linkeissä ja kuvaava data satelliiteissa. Data eroaa vain siltä osin, että sitä on muokattu tämän päivän liiketoimintasäännöillä. (Linstedt 2011, 2.)

Alla olevassa kuviossa 6 on BDV-ratkaisu kuvattu osana EDW-arkkitehtuuria. BDV:tä hyödynnetään liiketoimintasäännöillä muokatussa liiketoimintakerroksessa Business Data Warehouse (BDW). Informaatiovirta luetaan sisään latausalueen (Stage) kautta raakadatakerrokseen (EDW), jonka jälkeen data muokataan liiketoimintasäännöillä samalla, kun se siirretään liiketoimintakerrokseen. Huomioitavaa on, että BDV:tä hyödynnettäessä metadatalta on merkittävä rooli. Sitä tarvitaan, jotta ymmärrettäisiin BDV:n sisältö. Linstedtin (2011a, 2) mielestä ilman metadataa BDV ei tuota siitä saatavaa täyttä arvoa liiketoiminnalle tietopääoman hyödyntämisessä. (Linstedt 2011a, 2.)



Kuvio 6. BDV-ratkaisun sijainti EDW:ssä (Genesee Academy, LLC, 2011)

Liiketoiminnalle lisäarvon tuottamisen lisäksi BDV:n käytöstä seuraa muitakin hyötyjä. Näitä lisähyötyjä ovat:

- Uusien Data Marttien nopea toteutusmahdollisuus.
- Yksi yhdistetty ETL-latausrutiini liiketoimintasäännöille (tietojen jalostaminen yhdestä paikasta organisaatiossa).
- Mahdollisuus perustaa liiketoiminta-aiheisiin perustuva liiketoiminnan metatietontologia.
- Mahdollistaa yhtenäisten perustietojen eli Master Data -hallinnan käyttöönoton.

(Linstedt 2011a, 3.)

BDV tuo myös merkittävän mahdollisuuden lisätä liiketoiminnan ketteryyttä mahdollistamalla niin sanotut virtuaaliset Data Martit eli tietokantanäkymät, jotka pystytään luomaan BDV:n liiketoimintakerroksen päälle, koska ne eivät ole fyysisiä. Näkymien avulla liiketoimintakäyttäjät pääsevät suoraan käsiksi valmiisiin tietokokonaisuuksiin, jolloin he voivat yhdenmukaistaa tietotarpeitaan ja niiden tarkoituksenmukaisuutta. Esimerkiksi liiketoiminnan ei enää tarvitse tyytyä huonoon tiedon laatuun, vaan he voivat tarkastaa syyt tiedossa ilmenneisiin virheisiin virtuaalisista virhe Data Marteista, jotka eivät sisällä pysyvää dataa, vaan niitä hyödynnetään vain tarvittaessa. (Linstedt 2011a, 3.)

3 Tutkimusperusta ja käytetyt tutkimusmenetelmät

Tutkimuksen tavoitteena on tutkia Data Vault -menetelmän soveltuvuutta ja sen tuomia vaihtoehtoja eläkepalvelujen tietovaraston mallinnukseen kohdeyrityksen nykyiseen ratkaisuun verrattuna. Lisäksi tavoitteena on havainnollistaa ja testata menetelmän toimivuutta kohdeyrityksen tietovarastoympäristössä rakennetun konstruktion avulla ja selvittää, tuoko se muutosta tietovaraston muutosjoustavuuteen ja skaalautuvuuteen. Tutkimus toteutetaan projektina, jossa tutkitaan Data Vault -menetelmän soveltuvuutta kohdeyrityksen käyttöön sekä laaditaan raportti tutkimuksen vaiheista ja lopputuloksista.

Tutkimus kohdistuu yhteen eläkealan yritykseen, Yritys A ja se toteutetaan konstruktivisella tutkimusotteella. Tällöin tutkimuksessa korostuu toteuttajien ja tutkimuksen hyödyntäjien välinen viestintä sekä samalla kohdeyritys saa puolueettoman ja teoreettisen ratkaisun ongelmaan (Ojasalo, Moilanen & Ritalahti 2009, 66). Lisäksi konstruktivisen tutkimusotteen etuna on tutkimusvaihe, jossa toteutetaan ratkaisu ja testataan sen toimivuutta käytännössä (Lukka 2009). Tutkimusvaiheen tavoitteena on saada käytännön ongelmaan teoreettisesti perusteltu ja uudenlainen ratkaisu, joka tuo kohdeyrityksen liiketoiminnalle sekä mahdollisesti myös tiedeyhteisölle uutta tietoa (Ojasalo ym. 2009, 65).

Tutkimuksen yhteydessä muina tutkimusmenetelminä käytetään kirjallisuuskatsausta, dokumenttianalyysejä ja puolistrukturoitua teemahaastattelua. Kirjallisuuskatsauksen ja dokumenttianalyysin avulla kerätään laaja teoriaperusta ja viitekehys tutkimuskysymykseen liittyen, jonka pohjalta tarkastellaan kohdeyrityksen käyttöön sovellettua ratkaisua Data Vault -menetelmästä. Haastattelua hyödynnetään nykytilan ja siihen liittyvien ongelmatilanteiden kartoituksessa sekä lopputulosten tulkitsemisessä. Toteutetun konstruktion avulla pyritään tuottamaan syvällistä ja yksityiskohtaista tietoa Data Vault -menetelmän käyttämisestä tietovaraston mallinnuksessa kohdeyrityksessä.

Tutkimusmenetelmälähteinä tutkimuksessa käytetään Ojasalon, Moilasen ja Ritalahden teoksen Kehittämistyön menetelmät: Uudenlaista osaamista liiketoimintaan ja Lukan julkaisemaa artikkelia Konstruktivinen tutkimusote. Lähteissä kuvataan asiantuntevasti

ja selventävästi tutkimusmenetelmiä, joita hyödynnetään tutkimuksen toteuttamisessa. Tutkittavaa tapausta lähestytään konstruktiiivisella tutkimusotteella. Tutkimuksessa keskitytään ensisijaisesti laadulliseen haastattelumenetelmään. Niistä käytetään puolistrukturoitua haastattelua eli teemahaastattelua, jonka avulla testataan teorian paikkaansa pitävyyttä ja selvitetään tietoa lähtötilanteesta sekä analysoidaan kehittämistyössä saatuja tuloksia.

3.1 Konstruktiiivinen tutkimusote

Konstruktiiivinen tutkimusote on yksi tapa suorittaa tapaustutkimusta ja sitä voi soveltaa useilla eri alueilla, kuten tietojärjestelmätieteissä. Sen avulla parannetaan nykyysteemien suorituskykyä tai luodaan tietämystä suunnittelua ja toteutusta varten. Se on innovatiivisia konstruktioita tuottava metodologia, joka tuo panosta sille tieteenalalle, jossa sitä sovelletaan. Sen avulla luodaan uutta todellisuutta ja vaikutetaan tosielämään, kehittämällä konstruktio, joka poikkeaa kaikesta jo olemassa olevasta. Lisäksi se on luonteeltaan kokeellista, jolloin uutta konstruktioita voidaan tarkastella välineenä, jolla yritetään havainnollistaa, testata tai jalostaa aikaisempaa teoriaa, tai luoda kokonaan uusi teoria. (Lukka 2009.)

Konstruktiiivinen tutkimusote keskittyy tosielämän ongelmiin, jotka koetaan tarpeelliseksi ratkaista käytännössä. Alkuperäisen tosielämän ongelman ratkaisemiseksi tuotetaan konstruktio, joka sisältää kehitetyn konstruktion toteuttamisyrityksen, käytäntöön soveltuvuuden testaamiseksi. Konstruktiiivisen tutkimusotteen yhteydessä odotetaan tapahtuvan kokemuksellista oppimista, joka merkitsee tutkijan ja käytännön edustajien hyvin läheistä tiimimäistä yhteistyötä. Lisäksi siinä on oleellista sitoa käytännön ongelma ja sen ratkaisu teoreettiseen tietoon ja kiinnittää erityistä huomiota empiiristen löydösten reflektointiin takaisin teoriaan. (Lukka 2009.)

Tässä tutkimuksessa lähestyttiin yhtä kohdeyrityksessä ilmennyttä ongelmaa konstruktiiivisella tutkimusotteella. Näin ollen tutkimuksen avulla pyrittiin luomaan uusi konstruktio, jolla havainnollistettiin ja testattiin Data Vault -menetelmän sopivuutta kohdeyrityksen tietovarastoratkaisuun rajatulla liiketoiminta-alueella. Testauksen yhteydessä havainnollistettiin menetelmän toimivuutta ja rakenteen tuomia vaihtoehtoja tietovarast-

ton muutosjoustavuuteen ja skaalautuvuuteen. Samalla pystyttiin sitomaan käytäntö teoreettiseen tietoon ja toimimaan yhteistyössä käytännön edustajien kanssa. Konstruktiivinen lähestymistapa mahdollisti myös sen, että uuden ratkaisun kehittämisen ja testaamisen yhteydessä luodaan mahdollisesti uusi kohdeyrityksen käyttöön soveltuva ratkaisu, joka pystyy paremmin vastaamaan liiketoiminnan muutostarpeisiin.

3.2 Kirjallisuuskatsaus

Kirjallisuuskatsaus on yksi tutkimusmenetelmä ja osa jokaista tutkimustyötä. Sen avulla luodaan teoreettinen viitekehys tutkimukselle ja kiinnitetään tutkimus aikaisemmin tehtyihin tutkimuksiin. Kirjallisuuskatsauksen yhteydessä pyritään tutkimaan lähteiden validiteetti ja reliabiliteetti. Se on synteesi sisältölähteistä, joista selvitetään relevanssi ja keskeiset käsitteet tutkimuksen kannalta.

Tutkimusmenetelmänä kirjallisuuskatsaus tuki hyvin tutkimuskysymykseen vastaamista, sillä kyseessä on uusi asia, josta ei ole paljon aikaisempaa tutkimusta. Kirjallisuuskatsauksen avulla saatiin kerättyä teoriaperusta kokeellisen konstruktion tarkastelua varten. Lisäksi teoriaperustan kautta saatiin jo osittain vastauksia tutkimuskysymyksiin. Tutkimuksen teoria-aineisto koostui aiheeseen liittyvistä kirjallisista ja internet -julkaisuista. Teoriaosuus sisälsi myös omakohtaista tietoa ja kokemuksia tietovarastojen mallinnuksesta.

Tutkimuksen kontekstiin liittyvissä lähteissä on vain kaksi kirjaa, sillä Data Vault -menetelmä on suhteellisen uusi eikä siitä ole vielä saatavilla muuta kirjallisuutta. Dan Linstedt, menetelmän luoja on julkaissut ja osallistunut kirjojen Supermodel: The Business of Data Vault modeling ja Super Change Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault toteuttamiseen. Koska Linstedt on Data Vault -menetelmän kehittäjä, on hän näin ollen sen paras asiantuntija. Kirjoissa perehdytään menetelmän perusteisiin ja sen olemassa oloon sekä tuodaan esille, että Data Vault ei ole vain keskitetty tietovarasto, vaan se kytkeytyy myös liiketoimintaprosessien ja organisaation toiminnan kehittämiseen ja seurantaan. Muina lähteinä käytetään Linstedtin julkaisemaa materiaalia internetissä sekä muiden asiantuntijoiden aihee-

seen liittyviä artikkeleita ja tietovarastomallinnukseen liittyvää materiaalia. Aiheeseen liittyvää muuta tutkimusmateriaalia ei tutkimuksen tekohetkellä ollut saatavilla.

3.3 Dokumenttianalyysi

Tutkimusmenetelmänä dokumenttianalyysissä pyritään tekemään päätelmiä kirjalliseen muotoon saatetusta erityisesti symbolisesta, verbaalisesta tai viestinnällisestä aineistosta. Nämä käsittävät esimerkiksi tekstiksi muutetut haastattelut, www-sivut, vuosikertomukset, markkinointimateriaalit, ideointipalaverien muistiot, puheet, keskustelut, raportit, lehtiartikkelit ja muut kirjalliset materiaalit. Dokumenttianalyysin päävaiheita kuvaa hyvin laadullisen tutkimuksen yleinen malli, jonka vaiheita ovat aineiston kerääminen ja valmistelu esimerkiksi haastattelujen litterointi, aineiston pelkistäminen, aineistossa toistuvien rakenteiden tunnistaminen ja tulkinta sekä kaikkiin edellisiin vaiheisiin liittyvä kriittinen tarkastelu. Menetelmän vahvuutena voidaan pitää sen herkkyyttä asiayhteydelle, mikä tarkoittaa sitä, millaisena kehittämisen kohteena oleva ilmiö esiintyy luonnollisessa ympäristössään. (Ojasalo ym. 2009, 121, 122.)

Tutkimuksen empiriaosuuteen liittyen käytetään tutkimuksen kontekstiin liittyvää kirjalliseen muotoon saatettua materiaalia kohdeyrityksestä, sen tietovarastoinnin nykytilasta sekä tutkimuksen kohteeksi valitusta liiketoiminnan raportointialueesta. Dokumenttianalyysi tehdään sisältölähtöisesti, jolloin aineistoa tiivistetään ja selkeytetään pelkistämisen avulla. Näin aineistosta saadaan rajattua ja tunnistettua pieni määrä näkökulmia. Analyysin avulla aineiston informaatioarvo kasvaa ja sitä voidaan hyödyntää nyky- ja tavoitetilan vertailussa sekä päätelmien ja johtopäätösten tekemisessä.

3.4 Teemahaastattelu

Haastattelu on luottamusta vaativa menetelmä, sillä siinä on aina mukana vuorovaikutusta. Se on yksi käytetyimmistä tiedonkeruumenetelmistä tutkimus- ja kehittämistyössä. Haastattelu mahdollistaa syvällisen tiedon keräämisen kehittämisen kohteesta ja sen avulla on mahdollista saada kerätyksi uusia näkökulmia avaavaa aineistoa, varsinkin jos kehittämiskohdetta on tutkittu vain vähän, kuten tämän tutkimuksen kohteena olevaa aihetta. Haastattelumenetelmiä on erilaisia, joista tämän tutkimuksen yhteydessä käytettiin puolistrukturoitua eli teemahaastattelumenetelmää. (Ojasalo ym. 2009, 95, 97.)

Teemahaastattelun avulla kartoitetaan nykytilannetta ja kerätään aineistoa tutkimuksen aikana toteutetusta konstruktiosta. Haastatteluissa noudatetaan soveltuvin osin teemahaastattelun periaatteita. Haastattelujen kautta kerätty aineisto analysoidaan ja selvitetään mahdollisten jatkohaastattelujen tarve. Haastattelu sopii hyvin tämän konstruktivisen tutkimuksen tiedonkeruumenetelmäksi, koska kyseessä on kvalitatiivinen eli laadullinen tutkimus. Teemahaastattelu mahdollistaa etukäteen määritellyt teemat ja strukturoidut vastaukset, jolloin vastaukset pystytään rajaamaan halutulle aihealueelle. Haastattelun avulla kerätään tietoa tietovarastomallinnuksen nykytilasta kohdeyrityksessä sekä kokemukseräistä tietoa uuden konstruktion rakentamisesta. Analysoitua haastattelumateriaalia hyödynnetään myös loppupäätelmien tekemisessä.

3.5 Vertailun suorittaminen nykytilan ja rakennetun konstruktion välillä

Tutkimuksen yhteydessä toteutetaan kokeellinen konstruktio Data Vault -menetelmän käytöstä yhdellä kohdeyrityksen liiketoiminta-alueista. Konstruktion avulla pyritään tuottamaan ratkaisuehdotus kohdeyrityksen tietovaraston toimivuuden parantamiseksi, jotta se pystyy paremmin vastaamaan liiketoiminnan muutostarpeisiin. Samalla selvitetään teoriaperustan avulla, miten menetelmä on sovellettavissa kohdeyrityksen kaltaiseen organisaatioon ja tuoko se parannusta aiempaan tapaan toimia tai ketteryyttä tietovaraston muutoshallintaan.

Vertailu nykytilan ja rakennetun konstruktion välillä suoritetaan haastatteluista saadun materiaalin avulla. Haastattelujen analysoinnissa luokitellaan litteroitu aineisto ja pyritään löytämään siitä yhteyksiä käytettyyn teoriaan. Koska kyseessä on teemahaastattelu, niin haastatteluaineisto puretaan myös teema-alueittain. Lisäksi haastatteluaineistoista pyritään laskemaan tiettyjä ilmiöitä ja niiden esiintymistiheyttä. Vertailun tulos, joka esitetään ilmiöstä riippuen yhteisessä taulukossa, diagrammissa tai prosentteina, on luetavissa tämän raportin luvussa 5 (Empiria tutkimuksen kohteena olevasta raportointialueesta).

3.6 Tutkimuksen viitekehys ja rajaukset

Seuraavaksi käydään läpi tutkimuksen viitekehys sekä siihen liittyvät rajaukset. Viitekehystenä tälle tutkimukselle toimivat lukuun 2 (Tietojen mallinnus tietovarastossa) koottu teoriaosuus sekä kohdeyrityksen eläkepalvelujen raportoinnin osa-alue tietovarastossa, johon kohdistui muutoksia operatiiviseen järjestelmään tehtävistä uudistuksista johtuen. Eläkepalvelujen raportointiin tehtävät muutokset toteutettiin kehittämishankkeessa, jossa käyttöön otettiin Data Vault -menetelmää hyödyntävä uudistettu tietovarastoratkaisu. Tutkimuksessa keskitytään vain tutkimuksen aikana kehityshankkeen tuomiin vaikutuksiin ja muutoksiin raportoinnissa. Kohdeyrityksen kehittämishankkeeseen liittyvät dokumentit ja tuotokset ovat organisaation sisäisiä. Tässä tutkimusraportissa viitataan projektin sisältöön tarvittavilta osin. Lisäksi tuodaan esille nykyisen tietovarastoraportointiratkaisun ongelmia.

3.6.1 Tutkimuksen kohteena oleva raportointialue

Kohdeyrityksessä on ollut käytössä tietovaraston keskitetty hallintomalli kymmenisen vuotta, jonka avulla liiketoimintatiedon hyödyntäminen ja hallinnointi on hoidettu keskitetysti tietovarastossa. Kyseessä ei kuitenkaan ole puhdas EDW-ratkaisu, sillä tietovarasto ei sisällä yrityksen kaikkia tietoja, vaan vain osan niistä. Tietovaraston tietoarkkitehtuurin toteutuksessa mallinnusmenetelmänä on käytetty Kimballin tähtimallia. Tietovarasto on rakennettu useammasta erillisestä Data Martista, jotka perustuvat tähtimalliin. Tähtimallin etuna on ollut nopea toteutus, jolloin paikallisille käyttäjille on saatu toimitettua helposti raportteja. Ongelmana tässä arkkitehtuurissa on kuitenkin ollut, että Data Marteista on tullut erillisiä kunkin raportointialueen tarpeiden mukaisesti. (Hovi 2009, 26.)

Tutkimuksen kohteena on eläkepalvelujen tietovarastopohjainen raportointi, joka on yksi tietovaraston raportoinnin osa-alueista. Raportointi keskittyy työmäärien ja eläkeasioiden seurantaan ja on näin ollen osa johtamisjärjestelmää. Eläkepalvelujen raportoinnin ylläpito- ja kehitystyötä tekee BI-ratkaisun toimittaja, joka on ollut kehittämässä kohdeyrityksen raportointia vuodesta 2004 lähtien. Toimittaja vastaa myös Data Vault -arkkitehtuurin käyttöönotosta eläkepalvelujen raportoinnissa.

Kohdeyritys käyttää eläkevakuutusten käsittelyyn tällä hetkellä eläkekäsittelyjärjestelmää (K), josta raportoitavat tiedot siirretään eläkepalvelujen tietovarastoraportointiin. Näitä raportteja käyttävät useat käyttäjäryhmät kohdeyrityksen sisällä ja ulkopuolella. Talon sisällä käyttäjiä on yhteensä noin 200. Asiakkaiden ja esimiesten lisäksi myös eläkekäsittelijät hyödyntävät raportteja omien tilastotietojensa seurantaan. Eläkekäsittelyraportteja oli tutkimuksen käynnistyessä noin 140 kappaletta ja ne käsittelevät seuraavia kokonaisuuksia:

- eläkeläisten perustiedot ja vanhuuseläkkeiden tiedot
- yritysasiakkaille menevät listaukset (osoite- ja kuolinilmoituslistauksia)
- eläkkeensaajalistaukset
- tilastointiraportit.

Eläkekäsittelyraportoinnin ensisijaisena lähteenä on eläkekäsittelyjärjestelmä (K), joka on eläkepalveluiden nykyinen eläkekäsittelyjärjestelmä. Kuten aiemmin on jo mainittu, niin kohdeyrityksessä on käynnissä eläkekäsittelyn operatiivisen järjestelmän uusimishanke, jonka aikana eläkekäsittelyjärjestelmä (K) tullaan korvaamaan uudella.

3.6.2 Nykyisen tietovarastoratkaisun ongelmat

Nykyisessä tietovarastoratkaisussa on ilmennyt ongelmia eläkepalvelujen raportoinnissa. Ongelmia selvitettiin eläkekäsittelyjärjestelmän uudistamisen kokonaishankkeen käynnistysvaiheessa tehdyn kartoituksen yhteydessä sekä tämän tutkimuksen nykytilan haastattelututkimuksen yhteydessä. Seuraavaksi käydään läpi uudistamishankkeen kartoituksen yhteydessä esille tulleita ongelmia, joita verrataan haastattelututkimuksen tuloksiin.

Nykyisen tietovaraston rakenne ja tietojen mallinnus perustuivat vain vanhan eläkekäsittelyjärjestelmän (K) tietorakenteisiin eikä käytettävissä ollut ulkopuolisia tietolähteitä. Lisäksi tietovarastoraportoinnin ohella perusjärjestelmän tietokantaan tehtiin suorita Ad hoc -kyselyjä, jotka ovat yleensä tietokantaan nopeasti tehtäviä, ennalta määrittelemättömiä kyselyjä ja raportteja. Tietovaraston rakenne aiheutti myös sen, että oikean tiedon löytäminen oli vaikeaa, jolloin esimerkiksi eläkekäsittelyn ja -maksamisen tiedot meni-

vät helposti sekaisin. Tämä ilmeni myös haastattelututkimuksen yhteydessä, sillä joka neljäs vastaaja oli sitä mieltä, että on vaikea tietää, mitä tietoja raportille voi poimia. Tästä seurasi myös, että tietovaraston hyödyntäminen muussa toiminnassa jäi hyvin vähäiseksi.

Nykyisessä tietovarastossa olleiden tietojen oikeellisuus ja käytettävyys toivat myös ongelmia. Ongelmat johtuivat etenkin siitä, että eläkekäsittelyn ja -maksamisen tiedot oli tuotettu eri säännöillä, jolloin ne eivät olleet yhteismitallisia ja niiden vertailu oli vaikeaa. Lisäksi haastattelututkimukseen osallistuneista joka viides koki, että raportille ei löytynyt tarvittavia tietoja tietovarastosta tai raporttia ei saanut halutuilla poimintaehdoilla.

Uudistamishankkeen kartoituksen yhteydessä ilmeni, että raporttien vertailtavuus yli ajan oli myös vaikeaa, sillä tietovarastossa historiatiedot eivät pysyneet samoina. Tiedoissa oli sisällöllisiä puutteita, sillä uusia tietoja päivitettiin vanhojen päälle, tarpeellista tietoa puuttui kannasta ja näiden tietojen korjaaminen oli vaikeaa tai jopa mahdotonta. Lisäksi ennustetietoa ei ollut saatavilla.

Raporttien sisältöön liittyviä ongelmia lisäsivät muun muassa asiakasraportit, jota tuotettiin asiakkaiden toiveiden mukaisesti ja ilman erilliskustannuksia. Haastattelututkimuksesta käy ilmi, että näitä Ad hoc -kyselyjä tehdään suhteessa yhtä paljon kuin tehdään uusia tai muutetaan ja kehitetään jo olemassa olevia raportteja. Ongelmana on, että raporttien sisällöissä on päällekkäisyyttä ja sisällöllisiä virheitä. Tästä seuraa turhia raportteja sekä vaikeuksia raporttien tietosisältöjen vertailussa. Haastatelluista myös yksi kymmenestä oli sitä mieltä, että nykyraporttien tietosisältö on epäselvä

Kartoituksen mukaan uusien raporttien, analyysien ja ennusteiden tuottaminen oli vaikeaa tai jopa mahdotonta, jonka seurauksena raporttitarpeet eivät täyttyneet ja raportoinnin rinnalle kehittyi erillistä manuaaliseurantaa. Nykyisten raporttien tiedot ovat myös liian tarkalla tasolla, jolloin raporttien käyttökelpoisuus on kyseenalainen. Lisäksi raporttien tarpeellisuutta ei ole tutkittu ja kaikkia aiemmin tehtyjä raportteja käytetään edelleen. Eräs haastatelluista ottikin esille sen, että nykyisin raportteja on liian paljon tarpeeseen nähden.

Lisäksi ylläpito-prosessissa ei oteta kantaa raportointitarpeiden priorisointiin, joka on olennainen osa nopeaa ja joustavaa ylläpitoa. Selvää on, että muutosten priorisoinnista hyötyvät sekä raporttien käyttäjät että tietovaraston kehittäjät, koska sen avulla saadaan totuuden mukainen näkemys raportoinnin tilasta. Lisäksi se tehostaa toimintaa ja antaa liiketoiminnalle mahdollisuuden tietää, missä vaiheessa heidän tilaamansa työt ovat ja valmistuvatko ne sovitussa ajassa. (Koistinen 2002, 93, 147.)

Ongelmallisuutta lisäsi nykyisen raportoinnin muutoksenhallinta, koska siihen liittyvä ylläpito-prosessi koettiin kartoituksen mukaan raskaaksi ja aikaa vieväksi. Myös haastattelututkimuksen perusteella vain puolet vastaajista koki raportoinnin ylläpidon riittävän nopeaksi ja helpoksi. Kuitenkin haastatelluista kahden kolmasosan mielestä ylläpito on joustavaa, mutta vain yhden kolmasosan mielestä tarpeeksi nopeaa.

Osaan edellä mainituista ongelmista, kuten historiointiin ja raporttien oikeellisuuteen pystytään vastaamaan Data Vault -menetelmään perustuvalla tietovarastoratkaisulla. Se mahdollistaa myös liiketoimintaprosessien sekä organisaation toiminnan kehittämisen ja seurannan parantamisen tehokkaamman raportoinnin ja ylläpidon myötä. Seuraavassa kappaleessa tuodaan tarkemmin esille kehittämishankkeen vaikutuksia eläkepalvelujen tietovarastoraportointiin sekä konkreettisia esimerkkejä raportointimuutoksista.

3.6.3 Kohdeyrityksen tietovarastoraportoinnin kehittämishanke

Kohdeyrityksessä on käynnissä eläkekäsittelyn operatiiviseen järjestelmään liittyvä kokonaishanke, jolla mahdollistetaan eläketapausten käsittelyn automatisointi. Lisäksi hankkeen aikana on tarkoitus yhtenäistää perusjärjestelmien tietorakenteiden erilaisuutta. Kokonaishanke tukee myös yritystasoista tietovarastointia, jolloin eri liiketoiminta-alueiden yhteiset tiedot saadaan koko organisaation käyttöön, josta hyvänä esimerkkinä ovat asiakassegmenttiedot. Siinä otetaan kantaa myös tietoarkkitehtuuriin sekä yhteisiin tiedonhallinnan periaatteisiin tietovaraston kehittämisessä ja ylläpidossa.

Osa tätä kokonaishanketta on raportoinnin kehittämishanke, jossa tarkoituksena on uudistaa eläkepalvelujen raportoinnin tietovarasto ja raportit. Kehittämishanke jakaantuu useampaan projektiin. Ensimmäisen vaiheen projektissa tehdään uusien raporttien

määrittely, suunnitellaan tietovaraston uudistaminen sekä etenemissuunnitelma seuraavalle projektille. Toisen vaiheen projektin tarkoituksena on eläkepalvelujen raportoinnin nykytoiminnallisuuden laajentaminen tietovarastoksi sekä kokonaishankkeen raportointitarpeiden tarkan määrittelyn aloittaminen. Lisäksi uusista raporteista toteutetaan ensimmäisen vaiheen raportit valmiiksi tehtyjen määrittelyjen perusteella. Näihin raportteihin liittyvät tietovarastomuutokset toimivat tämän tutkimuksen vertailumateriaalina.

Raportoinnin kehittämishankkeen ensisijaisena tavoitteena on uuden, historiointia ja useampaa päällekkäistä lähdejärjestelmää tukevan tietovaraston toteutus. Hankkeessa määritellään myös uudistetun eläkepalvelujen raportoinnin mittarit ja raportit sekä toteutetaan osa raporteista valmistuneiden määrittelyjen mukaisesti. Lisäksi tavoitteena on uuden ja vanhan eläkekäsittelyjärjestelmän tietojen yhdistäminen tietovarastossa yhteisraportoinnin mahdollistamiseksi. Data Vault -mallinnustavan käyttöönotolla on tarkoitus myös helpottaa uusien raportointialueiden mukaanottoa tulevaisuuden tietovarastoraportoinnissa. Kehittämishankkeen lopputuloksena käynnistetään myös eläkepalvelujen raportoinnin ylläpito, jolla varmistetaan raportointijärjestelmän toimivuus sekä se, että muutokset ja kehitettävät uudet raportit vastaavat liiketoiminnan johtamisen ja asiakasyritysten tarpeita lyhyelläkin varotusajalla.

4 Kehittämishankkeen vaikutukset tietovarastoraportointiin

Tutkimuksen toteutuksen aikana kohdeyrityksessä oli käynnissä eläkekäsittelyjärjestelmän uudistamisen kokonaishanke. Tämän uuden operatiivisen järjestelmän käyttöönoton myötä myös eläkekäsittelyjärjestelmästä tehtyyn tietovarastoraportointiin kohdistui uudistamistarpeita sekä lähteiden että raportoinnin sisällön suhteen. Tietovarastoraportoinnin uudistaminen päätettiin toteuttaa siten, että nykyinen raportointitietokanta päivitettiin tietovarastoksi. Liiketoiminnan tärkeimmät tavoitteet uudistetulle tietovarastoraportoinnille olivat seuraavat:

- Mukana ovat vakioraportit, jotka kattavat 80 % ulkoisten asiakkaiden tarpeista.
- Eläkekäsittelyjen aikaansaaminen on suhteissa resursseihin.
- Sisäinen raportointi on keskitetty ja sisältää myös keskeiset vakioraportit.
- Uuden tietovaraston tietomalli on mietitty kokonaan uudestaan siten, että se vastaa liiketoiminnan tavoitteita sekä on luotettava, selkeä, ymmärrettävä ja looginen.

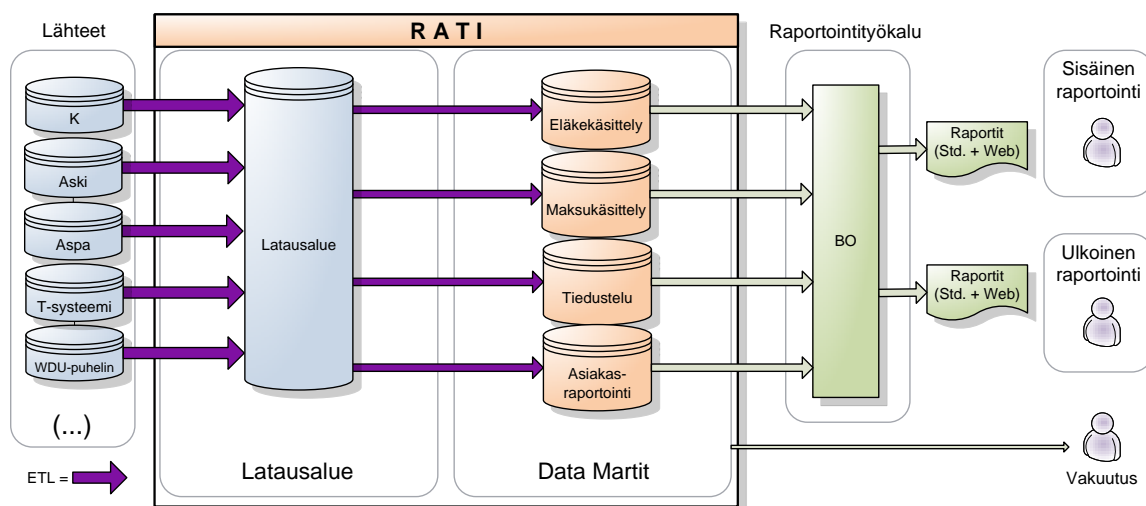
Alkuperäisestä tutkimuksen suunnitelmasta poiketen, uuden operatiivisen eläkekäsittelyjärjestelmän uudistamishankkeen aikataulu muuttui tutkimuksen aikana, jonka seurauksena myös uuden tietovarastoraportoinnin kehittämishankkeen näkökulma muuttui vuoden 2011 alussa. Sen johdosta päätettiin, että uudet raportit toteutetaan ja olemassa olevat uudistetaan vielä vanhan järjestelmän käytön aikana siten, että ne toteutetaan aluksi myös vanhasta eläkekäsittelyjärjestelmästä (K). Nämä muutokset heijastuivat myös tietovarastoraportointiin vuoden 2012 aikana.

Aikataulun muuttumisesta johtuen uutta eläkekäsittelyjärjestelmää ei ehditty ottaa käyttöön tutkimuksen aikana. Näin ollen tutkimuksen kohteena olevia muutoksia ei analysoitu uuden järjestelmän käyttöönotosta johtuvista muutoksista, vaan muutoksista, jotka toteutettiin vanhaan eläkekäsittelyjärjestelmään (K). Tämä muutos ei kuitenkaan estänyt Data Vault -menetelmän muutosjoustavuuden ja skaalautuvuuden tutkimista, vaan tutkimuksen tulokset perustuvat muutoksiin, jotka tehtiin vanhan eläkekäsittelyjärjestelmän (K) tietovarastoraportointiin.

4.1 Nykyinen tietovarastoarkkitehtuuri

Nykyisen raportointitietokannan (Rati) arkkitehtuuri koostuu latausalueesta ja Data Marteista (kuvio 7). Tiedot latausalueelle haetaan lähdejärjestelmistä ETL-välineen latausajoilla. Nykyisen raportointitietokannan latausalueelle tallennetaan lähdedatasta vain uusin versio, jolloin historiatietoja ei ole saatavilla.

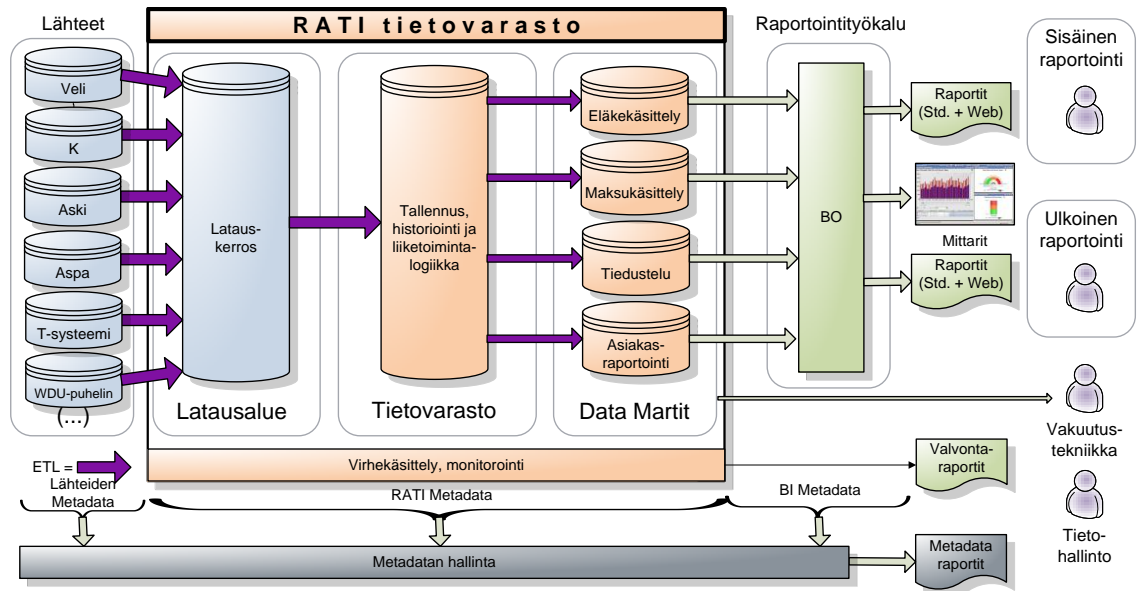
Latausalue toimii lähteenä Data Marttien latauksille. Latausalueelta tiedot siirretään Data Martteihin, joista ne luetaan liiketoimintaraportteille hyödyntäen raportointivälinettä. Latausaluetta käytetään muun muassa lähdetietojen muuntamiseen raportointiin sopivaksi. Liiketoiminnan käytössä olevien Data Marttien aihealueet ovat seuraavat: eläkekäsittely, maksukäsittely, tiedustelut ja asiakasraportointi.



Kuvio 7. Nykyisen raportointitietokannan arkkitehtuuri

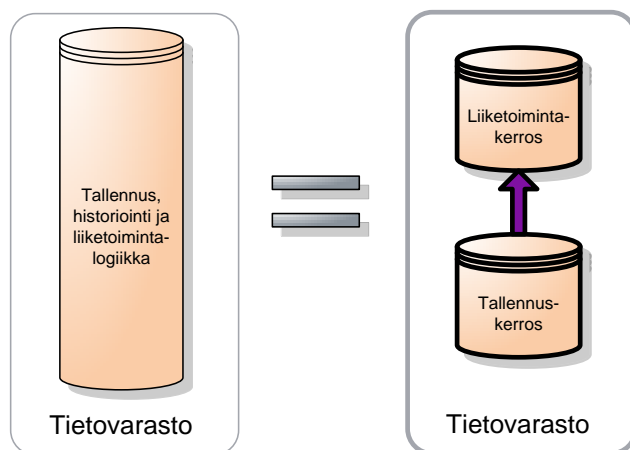
4.2 Uudistettu Data Vault -tietovarastoarkkitehtuuri

Uudistuksen myötä nykyiseen raportointitietokantaan lisätään tietovarastointikerros, joka mahdollistaa tietojen tallennuksen, historioinnin ja useamman lähdejärjestelmän tietojen yhtenäistämisen. Uusi tietovarastoarkkitehtuuri on kuvattu tarkemmin kuviossa 8. Arkkitehtuuri sisältää myös metadatan hallinnan, virhekäsittelyn ja monitoroinnin.



Kuvio 8. Uuden tietovaraston arkkitehtuuri uudistuksen jälkeen

Uudessa arkkitehtuurissa latauskerroksen ja Data Marttien väliin lisättiin tietovarasto, joka koostuu loogisesti kahdesta eri osasta, liiketoiminta- ja tallennuskerroksesta, joilla on eri funktiot (kuvio 9). Liiketoimintakerros mahdollistaa erimuotoisten tietojen yhdistämisen ja esittämisen määriteltyjen liiketoimintasääntöjen puitteissa. Liiketoimintakerroksessa on hyödynnetty kappaleessa 2.4.8 (Business Data Vault) esille tuodun BDV:n etuja ja se on toteutettu Data Vault -menetelmällä. Tallennuskerros hoitaa tietojen tallennuksen ja historiointin kaikista eri tietolähteistä ja sisältää raakadatan. Tämä rakenne mahdollistaa, että yhdistelysääntöjä voidaan tarvittaessa joustavasti muuttaa. Tärkeää on, että yhdistelysäännöt on määritelty ja kuvattu, jotta yhdistely voidaan toteuttaa ja säännöt ovat ylläpidettävissä.



Kuvio 9. Uuden tietovaraston eri kerrokset

Uuden tietovaraston tallennus- ja liiketoimintakerros on toteutettu Data Vault -menetelmällä. Näin ollen se koostuu menetelmän peruskomponenteista: hubeista, satelliteista ja linkeistä. Tämä ratkaisu tekee uudesta tietovarastosta erityisen muutosjoustavan, sillä liiketoimintakerroksessa oleva tietojen yhdistelylogiikka voidaan helposti korvata kokonaan uusilla säännöillä. Koska tallennuskerros sisältää raakadatan, niin siitä johdettu liiketoimintakerros on generoitavissa uudelleen, vaikka yhdistelysäännöt purettaisiinkin, sillä kaikki tarvittavat tiedot lähdejärjestelmistä ovat edelleen tallennettuna tallennuskerroksessa. Selvää on, että tarvetta tällaiseen uudelleen generointiin tulee, koska liiketoimintasäännöt ovat monimutkaisia ja liiketoimintakerroksessa yhdistelyihin ja yhteenlaskettuihin tietoihin kohdistuu jatkuvasti muutostarpeita.

Data Vault -menetelmän käyttöönotto uudessa tietovarastossa ei kuitenkaan poista tarvetta tähtimallille. Näin ollen myös uudessa arkkitehtuurissa Data Martti -raportointikerros toteutetaan edelleen tähtimallilla, jotta liiketoiminta ymmärtäisi paremmin, mitä tietoja he voivat hyödyntää raportoinnissa. Raportointiasiantuntijoiden käyttämä Universe, joka on Data Martti -raportointikerroksesta eläkekäsittelytietojen raportointia varten luotu raportointivälineen tietomalli (liite 1). Uudet raporttitarpeet ovat ohjanneet raportoinnin kehittämistä, jonka seurauksen myös raportointivälineen tietomallin rakennetta on laajennettu muun muassa kausi- ja käsittelijätehtävätiedoilla.

Yhtenä olennaisena erona aiempaan arkkitehtuuriin on, että tietovarasto tallentaa myös lähdedatan historian. Historiatiedon tallentaminen mahdollistaa seuraavat asiat:

- Uudet näkökulmat raportointiin, kuten esimerkiksi tietojen ”lukitsemisen” tiettyjen kausien päätteeksi sekä näiden lukittujen tietojen käytön kausiraportoinnissa.
- Datassa/tiedossa ajan kuluessa tapahtuvista muutoksista pystytään niin sanotusti irrottamaan tapahtumiin myös erilaisia näkökulmia.

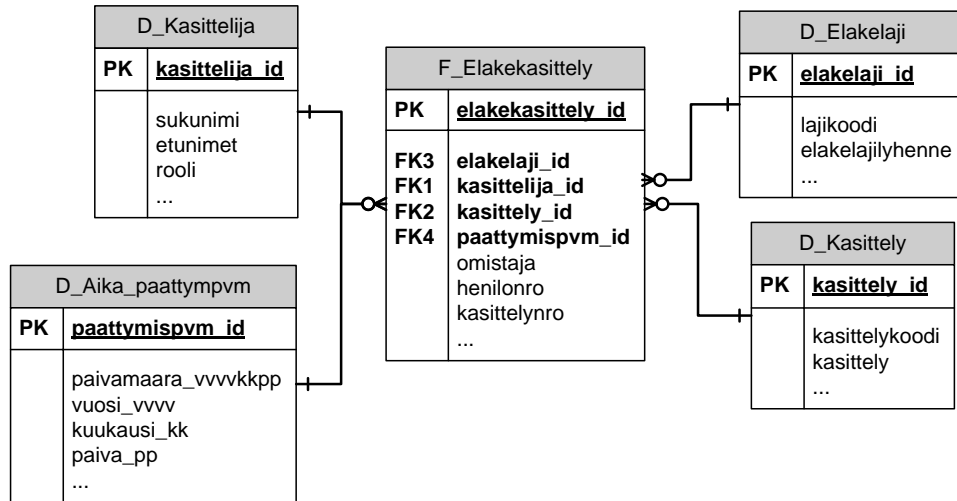
Näitä raportointimahdollisuuksia käsitellään tarkemmin tämän tutkimusraportin kohdassa 4.7 (Esimerkkejä Data Vault -menetelmän vaikutuksesta raportointiin), jossa kuvataan esimerkkien avulla muun muassa eläkekäsittelytehtävän lukitusta ja sen vaikutusta raportointiin.

4.3 Nykyinen tietomalli

Nykyisen raportointitietokannan toteutuksessa käytettiin ainoastaan Data Martteja, joihin tallennettiin vain viimeisin versio lähdedatasta. Tiedot Data Martteihin oli mallinnettu tähtimallilla, joka on Kimballin suositus Data Marttien suunnitteluun. Tähtimalli tukee hyvin moniulotteista ja numeerista tapahtumatietoa. Siksi sille hyvin tyypillisiä alueita ovat esimerkiksi myynnin seuranta ja taloushallinto kuin myös tutkimukseen liittyvä eläkekäsittely. Tähtimallin nimi tulee siitä, että sen rakenne muistuttaa tähteä. Sen avulla tavallaan simuloidaan moniulotteisuutta, vaikka käytetään perinteistä relatiokantaa. Lisäksi tähtimallia pidetään helppona ja selkeänä rakenteena myös käyttäjien ymmärtää. (Hovi ym. 2009, 36).

Tähtimalli on Linstedtin (2010, 14) mukaan yksinkertaisin tapa mallintaa tietovaraston tietokanta. Se sisältää faktataulun (mahdollisesti vain yhden, jotta malli olisi tähden mallinen) sekä siihen liittyvät dimensiotaulut. Luonteeltaan faktataulun tiedot ovat tapahtumatyyppisiä esimerkiksi myynti- tai kirjapitotapahtumia. Faktataulun perusavain on yhdistelmä dimensiotaulujen perusavaimista. Yleensä faktataulun numeroarvoja lasketaan yhteen erilaisilla dimensioista saaduilla tasoilla esimerkiksi, kun lasketaan yhteen faktataulun tietyn kuukauden myyntirivit myymälöittäin. Faktataulu on kooltaan tiedon rakeisuudesta riippuen yleensä suhteellisen iso (miljoonia rivejä). Dimensiotaulu taas sisältää staattista tietoa, josta saadaan faktataulun rivejä yhteen laskeville kyselyille erilaisia ryhmittelyjä ja hakukenttiä. Kooltaan dimensiotaulut ovat pieniä (sadoista tuhansiin riveihin) faktatauluihin verrattuna. (Hovi ym. 2005, 135.)

Seuraavassa kuviossa 10 on tähtimallin mukainen taulurakenne, joka sisältää yhden faktataulun F_Elakekasittely ja neljä dimensiotaulua, kuten D_Kasittelija. Taulujen perusavaimet on alleviivattu. Faktataulu on normalisoitu, joten siinä olevia tietoja ei toisteta, kun taas dimensiotauluissa tietoja toistetaan. Esimerkiksi käsittely ja käsittelykoodi toistuvat.



Kuvio 10. Tähtimallilla mallinnettu taulurakenne

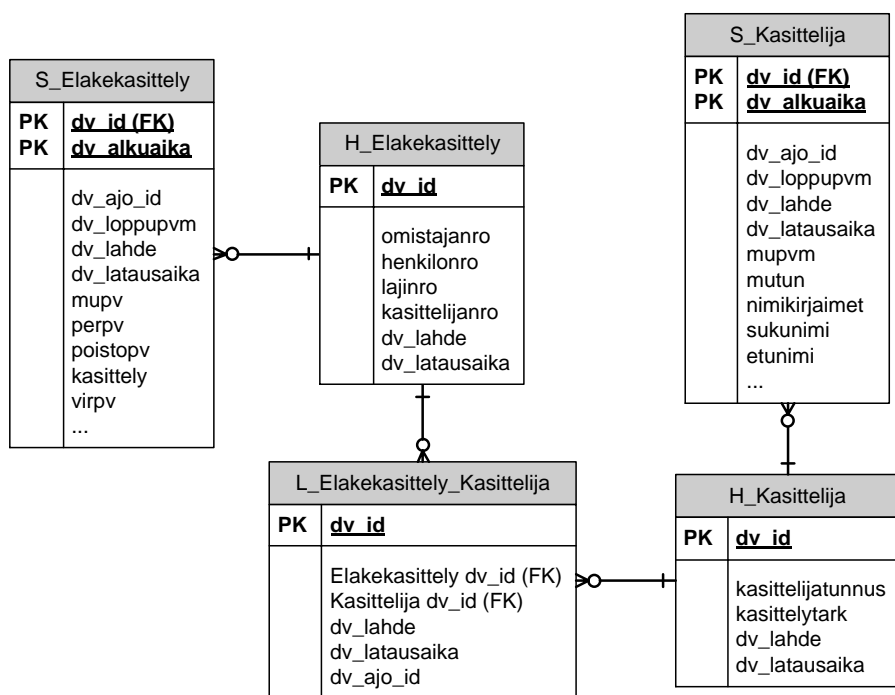
Usein tähtimallilla mallinnetussa tietovarastossa dimensioiden luonnolliset perusavaimet (PK), kuten eläkelaji tai käsittelijä korvataan niin sanotuilla surrogaateilla eli keino-tekoisilla avaimilla. Näin muutokset perusavaimiin, eivät aiheuta jälkeinpäin muutoksia fakta- ja dimensiotauluihin. Surrogaatit ovat yleensä juoksevia numeroita (esimerkiksi 1, 2, 3...). (Hovi ym. 2009, 37.)

Tietovaraston tietoja halutaan myös laskea yhteen tietyillä hierarkiatasoilla. Hierarkiat sisältyvät dimensioiden tietoihin, kuten kuvan D_Aika_paattymispvm-taulun kentät vuosi_vvvv, kuukausi_kk ja paiva_pp tai esimerkiksi siihen, että tuotteet kuuluvat tiettyihin tuoteryhmiin. Eri hierarkiatasot mahdollistavat faktatietoja summaavien liitoskyselyjen tekemisen, joista esimerkkinä voi mainita vuosittaiset tai kuukausittaiset eläkekäsittelyjen määrät. (Hovi ym. 2009, 37.)

4.4 Uudistettu Data Vault -tietomalli

Uudessa tietovarastossa hyödynnetään Data Vault -menetelmään perustuvaa tietomallia, joka koostuu jo aiemmin kerrotuista hubeista, satelliiteista ja linkeistä. Mielestäni Hovi (2010, 13) tiivistää hyvin Data Vault -rakenteen määritelmän yhteen lauseeseen: "Hubit ovat liiketoiminnan avaimia, linkeillä hoidetaan kaikki riippuvuudet (yhteydet) ja satelliitit pitävät sisällään kaiken tietosisällön sekä muutosten historioinnin."

Hubit ovat luonnollisia liiketoiminta-avaimia, kuten esimerkiksi omistajanumero ja henkilönnumero ja ovat näin ollen yksikäsitteisiä. Avaimet voivat koostua myös useammasta osasta. Hubien tietoja ei historioida. Satelliitti voi liittyä joko hubiin tai linkkiin ja samalla hubilla voi olla useita satelliitteja muun muassa silloin, kun satelliiteilla on eri päivitystiheys tai tietolähde. Satelliitti on päivämääräulottuvuudella varustettu taulu, joka sisältää hubiin liittyvät yksittäiset tietokentät eli attribuutit. Päivämäärät mahdollistavat tietojen historioinnin alku- ja loppupäivämäärän avulla. Linkit taas ovat yksikäsitteisiä yhteyksiä hubien välillä, jotka voivat olla myös tapahtumaperusteisia. (Hovi 2010, 13.)



Kuvio 11. Esimerkki Data Vault -mallinnetusta taulurakenteesta

Edellä olevaan kuvioon 11 on otettu vain eläkekäsittely- ja käsittelijätaulut edellä kuvattun tähtimalliesimerkin tauluista. Data Vault -mallinnuksessa eläkekäsittelystä on tehty H_Elakekasittely-hubi, joka liittyy H_Kasitteliija-hubiin L_Elakekasittely_Kasitteliija-linkin kautta. Kummankin hubin attribuutit sijaitsevat satelliittitauluissa, jotka ovat S_Elakekasittely ja S_Kasitteliija. Koska Data Vault -tietokannan rakenne on modulaarisempi, niin sisältää se huomattavasti enemmän tauluja kuin tähtimallilla toteutettu tietokanta.

4.5 Data Vault tietomallin vaikutus tietovarastolatauksiin

Tutkimuksen kohteena on Data Vault -mallinnuksen tuomat muutokset lähdejärjestelmän (K) K-ELAKETAPAHTUMA-tauluun, joka sisältää eläketapahtuman perustiedot. Eläketapahtuman eli eläkekäsittelyn tiedoista käy ilmi henkilön eläkeoikeus ja eläkeoikeudessa tapahtuneet muutokset. Kohdeyrityksen Data Vault -mallinnuksen toteutuksessa tietovarastoon noudatetaan standardin mukaisia suosituksia, jotta toteutusta pystytään automatisoimaan mahdollisimman paljon.

Tutkimuksen kohteena olevassa esimerkissä, jossa havainnollistetaan uuden tietomallin vaikutuksia tietovarastoon, puretaan K-ELAKETAPAHTUMA-taulun tiedot Data Vault -muotoon. Esimerkki kuvaa vain kohdeyrityksessä toteutettuun ratkaisuun liittyviä mallinnusperiaatteita eikä siinä oteta kantaa muihin mahdollisiin vaihtoehtoihin. Lopullinen Data Vault -taulurakenne ja taulujen väliset suhteet tallennuskerroksessa on kuvattu liitteessä 2. Tallennuskerroksen Data Vault -peruskomponentit luodaan seuraavassa järjestyksessä:

Vaihe 1 Luodaan hubit eli liiketoiminta-avaimet. Hubit eivät sisällä historiointia. Eläkekäsittelyn hubi on H_Eläkekäsittely ja sen liiketoiminta-avain koostuu neljästä tiedosta, jotka ovat omistaja (KET_OMISTAJA), asiakas (KSG_KET_HENKNR), vakuutuslaji (KET_LAJI) sekä käsittelynumero (KET_KASNO). Lisäksi hubi sisältää surrogaattiavaimen, tiedon lähdejärjestelmästä ja luontipäivän.

Vaihe 2 Luodaan satelliitit, joita hubilla voi olla useampia. Mutta eläkekäsittelyyn liittyen satelliitteja on vain yksi (TK_KET_S), joka sisältää melkein kaikki attribuutit lähdejärjestelmän K-ELAKETAPAHTUMA-taulusta. Satelliitin attribuutteihin ei sisälly tietoja, jotka löytyvät linkkien avulla muista hubeista, kuten käsittelijätunnus, joka löytyy H_Käsittelijä-hubista.

Satelliitin avain koostuu hubin surrogaattiavaimesta ja satelliitin olemassa olon alkupäivämäärästä. Data Vault -menetelmän yleinen periaate on, että aikaisemmin K-ELAKETAPAHTUMA-taulussa olleet viiteavaimet sijait-

sevat nykyisin linkeissä. Satelliitin attribuutteihin ei siis sisälly tietoja, jotka sijaitsevat linkeissä. Koska satelliitti ei sisällä muita viiteavaimia kuin vain omaan hubiinsa, niin linkkeihin liittyvät viiteavaimet tietävät, mihin hubeihin ne kuuluvat. Satelliittitaulu on aina lapsitaulu eikä koskaan isätaulu.

Lisäksi satelliitissa on loppupäivämäärä, joka ei sisälly avaimen. Loppupäivämäärä on lisätty helpottamaan historioinnin toteutusta tietokantamootorien suhteen. Myös alkupäivämäärä tarvitaan tietojen historiointia varten. Kun satelliitti luodaan, niin ensimmäisellä kerralla alkupäivämäärä saa arvon 1.1.1900 ja loppupäivämäärä 31.12.9999. Data Vault mallinnuksessa loppupäivämäärä ei ole pakollinen.

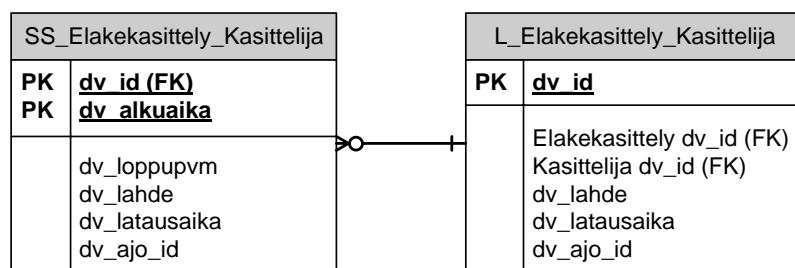
Tietojen historiointi tapahtuu seuraavasti. Kun satelliitissa jokin tieto muutuu, niin luodaan uusi ilmentymä (rivi) tauluun, jonka alkupäivämääräksi tulee uuden ilmentymän luontipäivä ja loppupäivämääränä säilyy edelleen 31.12.9999. Aiempi ilmentymä saa loppupäivämääräksi uuden ilmentymän luontipäivän.

Vaihe 3 Luodaan linkit, jotka ovat viiteavaimia hubien välillä. Linkki sisältää siihen liittyvien hubien surrogaattiavaimet ja linkillä on myös oma surrogaattiavain. Esimerkki yhdestä eläkekäsittelyyn liittyvästä linkistä on eläkekäsittelyn ratkaisijaan liittyvät tiedot (TK_KET_CKAS_RATKAISIJA_L). Linkin avulla saadaan selville käsittelijään liittyvät attribuutit H_Käsittelijä-hubista ja siihen liittyvistä satelliiteista.

Eläkekäsittelyyn luotavat linkit ovat Data Vault -mallinnuksen hierarkkisia linkkejä. Tämä tarkoittaa sitä, että linkin (lapsitaulu) elinkaari noudattaa H_Eläkekäsittely-hubin (isätaulu) elinkaarta.

Lisäksi eläkekäsittelyn mallinnuksessa on hyödynnetty statussatelliitteja, jotka ovat eräs muoto linkeistä. Statussatelliitissa on kyse yhteydestä, joka on olemassa vain tietyllä aikavälillä ja sen avulla saadaan selville, milloin yksi

käsittelijä on käsitelty kyseessä olevaa tapausta. Statussatelliittitaulu mahdollistaa käsittelyhistorian lukemisen, jolloin käsittelylle saadaan oikea ajankohde. Kuviossa 12 on kuvattu esimerkki eläkeasiaa hoitavan käsittelijän tunnukseen liittyvästä statussatelliittitaulusta.



Kuvio 12. Esimerkki statussatelliitista SS_Elakekasittely_Kasittelija, joka kertoo linkin voimassaolon

Statussatelliitti SS_Elakekasittely_Kasittelija ei sisällä omaa surrogaattiavainta eikä sillä ole muita viiteavaimia kuin vain linkkiin, jonka lapsitaulu se on. Tämä antaa enemmän vapautta ja mahdollisuuksia toimia eri tavoilla. Statussatelliittitaulun rivit ovat siis olemassa vain kun niitä tarvitaan. Ainoastaan, jos niiden avaintieto muuttuu, niin silloin niille muodostuu uusi ilmentymä hubiin ja niiden tilatieto päivittyy. Eläkekäsittelyyn liittyvässä ratkaisussa ei kaikilla linkeillä ole statussatelliittitaulua, sillä osa linkeistä on voimassa aina, jolloin niiden olemassaolo ei päätty.

Tässä ratkaisussa eläkekäsittelyyn liittyen kaikilla linkeillä ei ole statussatelliitteja, koska lähdejärjestelmän (K) taulujen pääavainten käsittely ohjaa toteutusta siten, että linkkiin liittyvien isätaulujen, kuten eläkekäsittely tai kuntoutus, luonnolliset eli liiketoiminta-avaimet ovat samanlaiset ja sisältyvät lapsitaulun avaimen. Siksi uudessa tietovarastossa pääperiaate on, että avaimet sisältävät samoja tietoja, mutta eivät kuitenkaan välttämättä kaikkia toisen hubin avainkenttiä. Näin varmistetaan eläkekäsittely- ja kuntoutustaulun tietojen muuttuminen samassa tahdissa.

Eläkekäsittelyyn liittyviä muita liiketoiminta-avaimia ovat esimerkiksi eläkelaji, asiakas, päätökset ja kuntoutus. Eläkekäsittelyyn kuuluvien käsitteiden linkkejä on kahdeksan kappaletta ja ne ovat: KEM - eläkkeenmaksutaulu, KES - pidätystiedot, KKA - ansio-toiminnot, KKK - laskun käsittely, KPS - eläkepäätösuhteet, KRM - ratkaisu ja kuntoutus muistiot, KTM - muistiot ja KUN - kuntoutus.

4.6 Viite-eheys Data Vault -mallinnetussa tietovarastossa

Tietoja ladattaessa tietovarastoon voivat ne tulla sinne myös eri tahdissa. Tällaisia eri aikaan tulevia tietoja voivat olla esimerkiksi eläkekäsittelytehtävät ilman käsittelijää. Näissä tilanteissa Data Vault -mallinnetussa tietokannassa viite-eheydestä huolehditaan siten, että latauksen yhteydessä puuttuvat tiedot on mahdollista luoda myös itsenäisesti. Näin tietokantaan ei siis synny ollenkaan niin sanottuja orpoja tietoja ilman vanhempia. Itsenäisesti luotavat tiedot mahdollistavat myös latausten riippumattomuuden toisistaan. Esimerkiksi jos lähdejärjestelmän käsittelijätaulusta ei löydy käsittelijätunnusta SAAA, niin se luodaan ensin H_Käsittelijä-hubiin, jotta linkki voidaan muodostaa. Näin varmistetaan, että tietojen eheys noudattaa, Data Vault -mallinnuksen suosituksia.

Kyse on siis tilanteesta, jolloin H_Käsittelijä-hubiin luodaan liiketoiminta-avain niillä tiedoilla, jotka linkin luonnin yhteydessä on käytettävissä. Tieto tulee siis muualta kuin ensisijaisesta lähteestä. Kyseessä ei ole tilapäinen arvo eikä osittainen ratkaisu, vaan jos lähdejärjestelmä ei tältä osin ole eheä, niin näin varmistetaan, että Data Vault -ratkaisu on. Esimerkiksi eläkekäsittelyä hoitavan lääkärin tunnuksia ei välttämättä ole käsittelijätaulussa, joten sille luodaan oma rivi H_Käsittelijä-hubiin. Huomioitavaa on, että kun tieto ei tule ensisijaisen lähteen kautta, niin tieto jää ilman kuvaavaa dataa (attribuutteja). Toinen esimerkki tällaisesta tilanteesta ovat jo poistetut käsittelijätiedot, jolloin kyseessä on tilanne, jossa lähdejärjestelmä sallii poiston ilman tietojen viite-eheyden tutkimista tai tietojen putsaamista. Lisäksi myös latausjärjestys voi aiheuttaa tilanteen, jossa tieto jää puuttumaan ja liiketoiminta-avain tulee vasta myöhemmässä latausajossa.

Data Vault -kannan latausajojen järjestyksellä ei ole yhtä suurta merkitystä kuin aiemmassa ratkaisussa käytetyssä tähtimallissa. Kannan rakenne mahdollistaa sen, että hubien lataamisen jälkeen satelliitit voidaan halutessa ladata samanaikaisesti, sillä niillä ei ole

viite-ehyksiä toisiinsa. Linkkitauluilla taas saattaa olla viite-ehyksiä ja siksi ne vaativat pohjatietoa, jonka tulee olla jo ladattuna tietovarastoon ennen linkkien luomista. Uudessa tietovarastoarkkitehtuurissa ajetaan kahdesta neljään latausajoa rinnakkain.

4.7 Esimerkkejä Data Vault -menetelmän vaikutuksesta raportointiin

Tässä tutkimuksessa perehdytään esimerkkien avulla uuden Data Vault -tietovarastoarkkitehtuurin tuomiin mahdollisuuksiin eläkekäsittelyn tietovarastoraportoinnissa. Esimerkeissä kuvataan, miten uusi Data Vault -menetelmällä mallinnettu liiketoimintakerros lisää raporttien luotettavuutta ja mahdollistaa asiasisällöltään uudenlaisten raporttien toteuttamisen. Sen avulla pystytään myös luomaan uusia liiketoimintakäsitteitä, joita voidaan hyödyntää raportoinnissa ja toiminnan johtamisessa.

Lisäksi esimerkkien kautta perehdytään raporttien sisällön oikeellisuuden paranemiseen historioinnin ja tietojen lukituksen avulla, jotka takaavat sen, että raportit sisältävät ajantasaisempaa tietoa eivätkä aikaisempien kausien tiedot muutu raporteilla. Tutkimuksessa tarkastellaan myös Data Vault -liiketoimintakerroksessa luodun uuden käsitteen, käsittelijätehtävän tuomia uusia mahdollisuuksia raportointiin.

Nykyisestä raportointitietokannasta raporteilla näytetään vain lähdedatan uusin versio. Erona uudessa tietovarastossa on, että sinne tallennetaan myös lähdedatan historia. Tämä mahdollistaa raportointiin uusia näkökulmia, kuten esimerkiksi tietojen ”lukitsemisen” tiettyjen kausien päätteeksi ja näiden lukittujen tietojen käytön kausiraportoinnissa. Lisäksi datassa ajan kuluessa tapahtuvista muutoksista pystytään myös tehtäviin ”irrottamaan” erilaisia näkökulmia.

Tarkastelun kohteena olevat tiedot ladataan uuteen tietovarastoon eläkekäsittelyjärjestelmästä (K). Tutkimuksessa käytetyt esimerkit sisältävät käsittelyn päättämisen muutoshistorian muutaman perustiedon osalta. Oheisessa taulukossa 1 on kuvattu esimerkkidata, jota käytetään kaikissa seuraavissa esimerkkitapauksissa.

Taulukko 1. Esimerkkidata

Käsittely	Tekopäivä	Käsittelijä	Vireilletulo.pvm	Päät.anto.pvm	Päättämispvm	Poistopvm
EK001	20.12.2010	AAA	20.12.2010			
EK001	22.12.2010	AAA	20.12.2010	22.12.2010		
EK001	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK002	20.12.2010	AAA	20.12.2010			
EK002	22.12.2010	AAA	20.12.2010	22.12.2010		
EK002	4.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK003	20.12.2010	AAA	20.12.2010			
EK003	22.12.2010	AAA	20.12.2010	22.12.2010		
EK003	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK003	8.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	8.2.2011
EK004	20.12.2010	AAA	20.12.2010			
EK004	22.12.2010	AAA	20.12.2010	22.12.2010		
EK004	3.1.2011	BBB	20.12.2010	3.1.2011		
EK004	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	
EK004	3.2.2011	AAA	20.12.2010	2.2.2011		
EK004	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	
EK005	20.12.2010	AAA	20.12.2010			
EK005	22.12.2010	AAA	20.12.2010	22.12.2010		
EK005	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	
EK005	29.1.2011	AAA	20.12.2010	24.1.2011		
EK005	7.2.2011	BBB	20.12.2010	4.2.2011	4.2.2011	
EK006	20.12.2010	AAA	20.12.2010			
EK006	22.12.2010	AAA	20.12.2010	22.12.2010		
EK006	3.2.2011	AAA	20.12.2010	2.2.2011		
EK006	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	

Seuraavissa kappaleissa käydään läpi havainnollistavat raportointiesimerkit. Ensimmäiseksi kuvataan tilanteita nykyraportoinnissa, joissa esimerkissä käytettyjen päättyneiden eläkekäsittelyjen raporttien virheet syntyivät. Tästä seurasi, ettei kuukausitasoinen raportointi aiemmin ollut luotettavaa. Toisessa esimerkissä havainnollistetaan, miten muun muassa lukituksen ja historioinnin avulla uudesta tietovarastosta tuotetut raportit eivät enää sisällä edellä mainittuja virhetilanteita. Viimeisessä esimerkissä otetaan esille, miten uuden Data Vault -arkkitehtuurin myötä raportointiin on mahdollisuus luoda uusia käsitteitä, joita lähdejärjestelmässä ei vielä ole. Esimerkissä käydään läpi käsitteli-

jän eläkekäsittelytehtävä, joka tallennetaan liiketoimintakerrokseen ja jonka avulla mahdollistetaan toiminnan seurata eri tehtävien näkökulmasta.

4.7.1 Päättyneiden eläkekäsittelyjen raportointi nykyisessä ratkaisussa

Kuten edellä on jo tullut esille, niin nykyisessä raportointitietokannassa lähdejärjestelmän (K) datasta käytössä on vain uusin tilanne. Tämä mahdollistaa ainoastaan viimeimmän kantaan tallennetun tilanteen raportoinnin ja sen seurauksena saattavat jo aiemmille kausille raportoidut tiedot muuttua. Seuraavassa esimerkissä havainnollistetaan tilanteita, joissa raportoinnin virheet syntyivät. Tilanteet perustuvat todellisiin käsittelytapahtumiin.

Esimerkin näkökulmana on eläkekäsittelyyn liittyvien päättyneiden käsittelyjen lukumäärä, jota tarkastellaan raportoinnin lopputuloksen kautta. Taulukossa XX on punaisella korostettu esimerkkidatasta 1.2.2011 raportointihetken tieto. Esimerkissä määritellään eläkekäsittelyn päättäminen seuraavasti:

- Eläkekäsittely on päätetty, kun sille on asetettu päättämispäivämäärä ja sitä ei ole poistettu.
- Päättäminen kohdistuu samalle kaudelle kuin asetettu päättämispäivä.

Taulukko 2. Eläkekäsittely, päätetyt eläkekäsittelyt aiemmassa raportointitietokannassa raportointihetkellä 1.2.2011

Käsittely	Tekopäivä	Käsittelijä	Vireilletulo.pvm	Päät.anto.pvm	Päättämispvm	Poistopvm
EK001	20.12.2010	AAA	20.12.2010			
EK001	22.12.2010	AAA	20.12.2010	22.12.2010		
EK001	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK002	20.12.2010	AAA	20.12.2010			
EK002	22.12.2010	AAA	20.12.2010	22.12.2010		
EK002	4.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK003	20.12.2010	AAA	20.12.2010			
EK003	22.12.2010	AAA	20.12.2010	22.12.2010		
EK003	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK003	8.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	8.2.2011

EK004	20.12.2010	AAA	20.12.2010			
EK004	22.12.2010	AAA	20.12.2010	22.12.2010		
EK004	3.1.2011	BBB	20.12.2010	3.1.2011		
EK004	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	
EK004	3.2.2011	AAA	20.12.2010	2.2.2011		
EK004	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	
EK005	20.12.2010	AAA	20.12.2010			
EK005	22.12.2010	AAA	20.12.2010	22.12.2010		
EK005	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	
EK005	29.1.2011	AAA	20.12.2010	24.1.2011		
EK005	7.2.2011	BBB	20.12.2010	4.2.2011	4.2.2011	
EK006	20.12.2010	AAA	20.12.2010			
EK006	22.12.2010	AAA	20.12.2010	22.12.2010		
EK006	3.2.2011	AAA	20.12.2010	2.2.2011		
EK006	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	

Seuraavasta taulukosta 3 ilmenee muutos tammikuun raportin lopputulokseen, kun samasta datasta tehdään raportti 1.3.2011 helmikuun tilanteesta. Koska datasta on käytössä vain eläkekäsittelyn uusin tieto, niin raportille poimitaan 1.3.2011 tilanteesta vain taulukossa vihreällä korostettu data, joka muuttaa jo raportoituja tietoja. Muutos tammikuun aiemmin raportoituun tilanteeseen on korostettu punaisella.

Taulukko 3. Eläkekäsittely, päätetyt eläkekäsittelyt aiemmassa raportointitietokannassa raportointihetkellä 1.3.2011

Käsittely	Tekopäivä	Käsittelijä	Vireilletulo.pvm	Päät.anto.pvm	Päättämispvm	Poistopvm
EK001	20.12.2010	AAA	20.12.2010			
EK001	22.12.2010	AAA	20.12.2010	22.12.2010		
EK001	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK002	20.12.2010	AAA	20.12.2010			
EK002	22.12.2010	AAA	20.12.2010	22.12.2010		
EK002	4.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK003	20.12.2010	AAA	20.12.2010			
EK003	22.12.2010	AAA	20.12.2010	22.12.2010		
EK003	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011	
EK003	8.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	8.2.2011
EK004	20.12.2010	AAA	20.12.2010			
EK004	22.12.2010	AAA	20.12.2010	22.12.2010		

EK004	3.1.2011	BBB	20.12.2010	3.1.2011		
EK004	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	3
EK004	3.2.2011	AAA	20.12.2010	2.2.2011		
EK004	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	
EK005	20.12.2010	AAA	20.12.2010			
EK005	22.12.2010	AAA	20.12.2010	22.12.2010		
EK005	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	
EK005	29.1.2011	AAA	20.12.2010	24.1.2011		
EK005	7.2.2011	BBB	20.12.2010	4.2.2011	4.2.2011	
EK006	20.12.2010	AAA	20.12.2010			
EK006	22.12.2010	AAA	20.12.2010	22.12.2010		
EK006	3.2.2011	AAA	20.12.2010	2.2.2011		
EK006	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	

Taulukon 3 pohjalta selviää myös, miten helmikuun alussa (1.2.) raportoitujen päätettyjen eläkekäsittelyjen lukumäärä muuttuu maaliskuun alussa ajettuun raporttiin verrattuna. Taulukon kohdassa 1 jälkikirjaus lisää tammikuulle päättämisen, kun taas kohdassa 2 poistaminen mitätöi yhden tammikuulle osuvan päättämisen. Kohdassa 3 aiemmin tammikuulle osunut päättäminen siirtyy helmikuulle. Raportille tulevien tietojen sisällöt on kuvattu seuraavassa kuviossa 13, josta ilmenee muutos tammikuun tiedoissa.

Ajopäivä 1.2.2011

Kuukausi	Päätetyt käsittelyt
2011-01	3

Ajopäivä 1.3.2011

Kuukausi	Päätetyt käsittelyt
2011-01	2
2011-02	3

Kuvio 13. Raportointiesimerkki, viimeisin versio raportointiajankohdilla 1.2. ja 1.3.2011

4.7.2 Päättyneiden eläkekäsittelyjen raportointi uudistetussa ratkaisussa

Uudessa tietovarastossa toteutettu tietojen historiointi mahdollistaa jonkin tietyn tilanteen ”lukituksen” tietylle ajanhetkelle ja raportoinnin muodostamiseen tämän lukitun ajanhetken tietoihin. Lukituksessa periaatteena on, että raportoidut kaudet (kuukaudet /vuodet) eivät muutu jälkeenpäin. Ominaisuutta hyödynnetään raporteissa, joissa kyse on käsittelymäärien seurannasta ja lukitus päätetään raporttikohtaisesti. Seuraavassa esimerkissä havainnollistetaan lukituksen vaikutusta päätettyjen eläkekäsittelyjen luku-

määrien raportointiin. Siinä keskitytään yhteen lukittavaan tietoon, eläkekäsittelyn päättämiseen ja päättämislukumäärien raportointiin kuukausitasolla. Esimerkissä määritellään eläkekäsittelyn päättäminen siten, että eläkekäsittely on päätetty tietyllä kaudella, jos

- sitä ei ole päätetty aikaisemmilla kausilla
- sen päättämispäivämäärä on kauden lopussa asetettu johonkin tiettyyn päivään.

Taulukko 4. Eläkekäsittely, päätetyt käsittelyt, lukitus kauden päätteeksi uudessa tietovarastossa raportointihetkillä 1.2. ja 1.3.2011

Käsittely	Tekopäivä	Käsittelijä	Vireilletulo.pvm	Päät.anto.pvm	Päättämispvm	Poistopvm	Lukituspvm
EK001	20.12.2010	AAA	20.12.2010				
EK001	22.12.2010	AAA	20.12.2010	22.12.2010			
EK001	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011		1.2.2011
EK002	20.12.2010	AAA	20.12.2010				
EK002	22.12.2010	AAA	20.12.2010	22.12.2010		1	1.2.2011
EK002	4.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011		1.3.2011
EK003	20.12.2010	AAA	20.12.2010				
EK003	22.12.2010	AAA	20.12.2010	22.12.2010			
EK003	26.1.2011	BBB	20.12.2010	24.1.2011	26.1.2011		1.2.2011
EK003	8.2.2011	BBB	20.12.2010	24.1.2011	26.1.2011	8.2.2011	1.3.2011
						2	
EK004	20.12.2010	AAA	20.12.2010				
EK004	22.12.2010	AAA	20.12.2010	22.12.2010			1.1.2011
EK004	3.1.2011	BBB	20.12.2010	3.1.2011			
EK004	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011		1.2.2011
EK004	3.2.2011	AAA	20.12.2010	2.2.2011			
EK004	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	3	1.3.2011
EK005	20.12.2010	AAA	20.12.2010				
EK005	22.12.2010	AAA	20.12.2010	22.12.2010			
EK005	26.1.2011	CCC	20.12.2010	24.1.2011	26.1.2011	4	
EK005	29.1.2011	AAA	20.12.2010	24.1.2011			1.2.2011
EK005	7.2.2011	BBB	20.12.2010	4.2.2011	4.2.2011	5	1.3.2011
EK006	20.12.2010	AAA	20.12.2010				
EK006	22.12.2010	AAA	20.12.2010	22.12.2010			1.2.2011
EK006	3.2.2011	AAA	20.12.2010	2.2.2011			
EK006	5.2.2011	AAA	20.12.2010	2.2.2011	4.2.2011	6	1.3.2011

Lukituksen vaikutukset raportoitavaan dataan on kuvattu tarkemmin edellä olevassa taulukossa 4. Tammikuun (1.2.) lukitustilanne on korostettu punaisella. Helmikuun jakson lukitus raportointia varten tapahtuu 1.3.2011 ja se on korostettu vihreällä. Lukituksen myötä raportoitava data eroaa aiemmasta seuraavasti:

- Kohdassa 1 jälkikirjaus lisää tammikuulle päättämisen. Lukituksen johdosta se näkyy vasta helmikuun raportilla.
- Kohdassa 2 poistaminen mitätöi yhden tammikuulle osuvan päättämisen. Poiston raportointitapa voidaan määrittellä halutunlaiseksi.
- Kohdassa 3 päättäminen on tapahtunut ja raportoitu jo tammikuulle, joten sitä ei raportoida enää helmikuulle.
- Kohdassa 4 tammikuun lopussa käsittelyllä ei ole asetettua päättymispäivää, joten sitä ei ole päätetty.
- Kohdat 5 ja 6 tammikuun lopussa käsittely ei ole ollut päätetty tilassa, joten se kirjataan helmikuulle.

Tietojen raportointiin uudesta tietovarastosta voidaan ottaa kaksi eri näkökulmaa. Kummassakin tilanteessa käytetään lukittua näkymää, joka mahdollistaa sen, että aiemmin lukitut ja raportoidut luvut säilyvät ennallaan. Ensimmäisessä näkökulmassa edellisille kausille jo lukittujen eläkekäsittelyjen poistoja ei vähennetä poistokauden päättämislukumäärästä. Tästä kuitenkin seuraa, että eläkekäsittelyjen yhteenlaskettu määrä on enemmän kuin kannan viimeinen tilanne todellisuudessa näyttää. Raporttien sisällöt on kuvattu kuviossa 14.

Ajopäivä 1.2.2011

Kuukausi	Päätetyt käsittelyt
2011-01	3

Ajopäivä 1.3.2011

Kuukausi	Päätetyt käsittelyt
2011-01	3
2011-02	3

Kuvio 14. Raportointiesimerkki, lukitus, kun jälkikirjaukset näytetään kirjauskuukaudella, mutta jälkipoistoja ei käsitellä lukitusta datasta (raportointiajankohdat 1.2. ja 1.3.2011)

Toisessa näkökulmassa edellisille kausille kohdistuvat eläkekäsittelyjen lisäykset ja poistot raportoidaan omina tietoinaan. Raportoinnin tietosisältö tarkentuu, mutta samalla monimutkaistuu. Kuten seuraavasta raporttiesimerkistä (kuvio 15) käy ilmi, niin helmi-kuun yhteensä lukumäärä on eri (2) kuin edellisessä esimerkissä (kuvio 14), sillä nyt poisto huomioidaan laskennassa. Näin raportoitavat tiedot vastaavat paremmin todellista tilannetta ja raportin sisältö on informatiivisempi.

Ajopäivä 1.2.2011

Kuukausi	Kauden aikana kaudelle päätetyt käsittelyt (kpl)	Aikaisempien jaksosten lisäykset	Aikaisempien jaksosten poistot	Yhteensä
2011-01	3	0	0	3

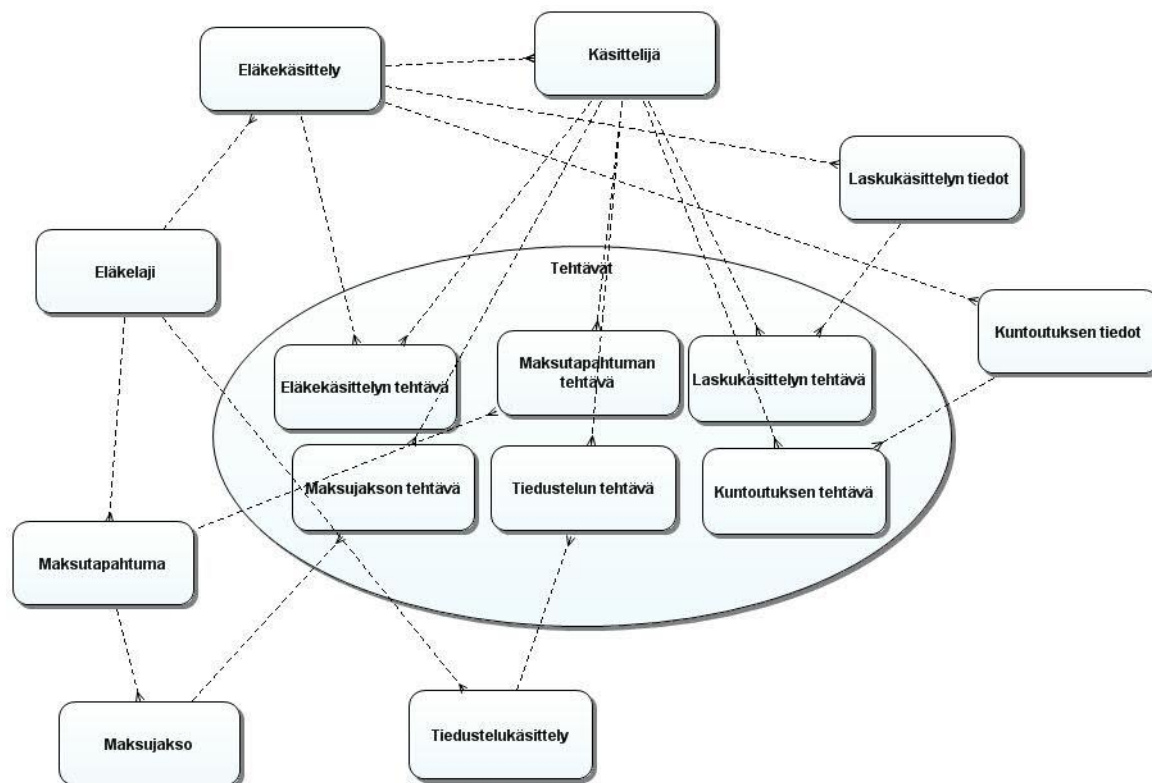
Ajopäivä 1.3.2011

Kuukausi	Kauden aikana kaudelle päätetyt käsittelyt (kpl)	Aikaisempien jaksosten lisäykset	Aikaisempien jaksosten poistot	Yhteensä
2011-01	3	0	0	3
2011-02	2	1	-1	2

Kuvio 15. Raportointiesimerkki, lukitus, jälkikirjaukset ja -poistot näytetään omina mittareinaan (raportointiajankohdat 1.2. ja 1.3.2011)

4.7.3 Uusi raportoitava liiketoimintakäsite, käsittelijän eläkekäsittelytehtävä

Uudessa tietovarastossa on Data Vault -arkkitehtuurin myötä mahdollisuus luoda myös uusia käsitteitä, joita lähdejärjestelmässä ei vielä ole. Yksi esimerkki tästä on käsittelijän eläkekäsittelyyn liittyvä tehtävät, jotka koostuvat päivittäin tallennettavista käsittelijän "tekemisistä" eli tehtävistä. Käsittelijän tehtävät voivat eläkekäsittelyn lisäksi liittyä myös tiedustelu-, kuntoutukseen maksu- ja laskukäsittelyyn sekä muutoksenhakuun ja ulkoa tuleviin asioihin. Käsittelijöiden tehtävien yhteys liiketoiminnan muihin käsitteisiin on kuvattu alla olevassa käsitelmissä (kuvio 16).



Kuvio 16. Käsittelijän tehtävien liittyminen liiketoiminnan muihin käsitteisiin

Tarve uuden käsitteen luomiselle on lähtenyt liiketoimintajohdon halusta seurata toimintaa eri tehtävien näkökulmasta. Käsittelyyn liittyvistä tehtävistä halutaan tietää, kuka on käsitellyt kulloistakin tehtävää ja minä ajankohtana. Uuden käsittelijän tehtävän myötä voidaan tietää esimerkiksi, kuka osallistui eläkekäsittelyyn tässä kuussa ja kuka seuraavassa. Kun tiedot on tallennettu tietovarastoon, niin ne pysyvät aina ajan tasalla eivätkä raportit vääristy.

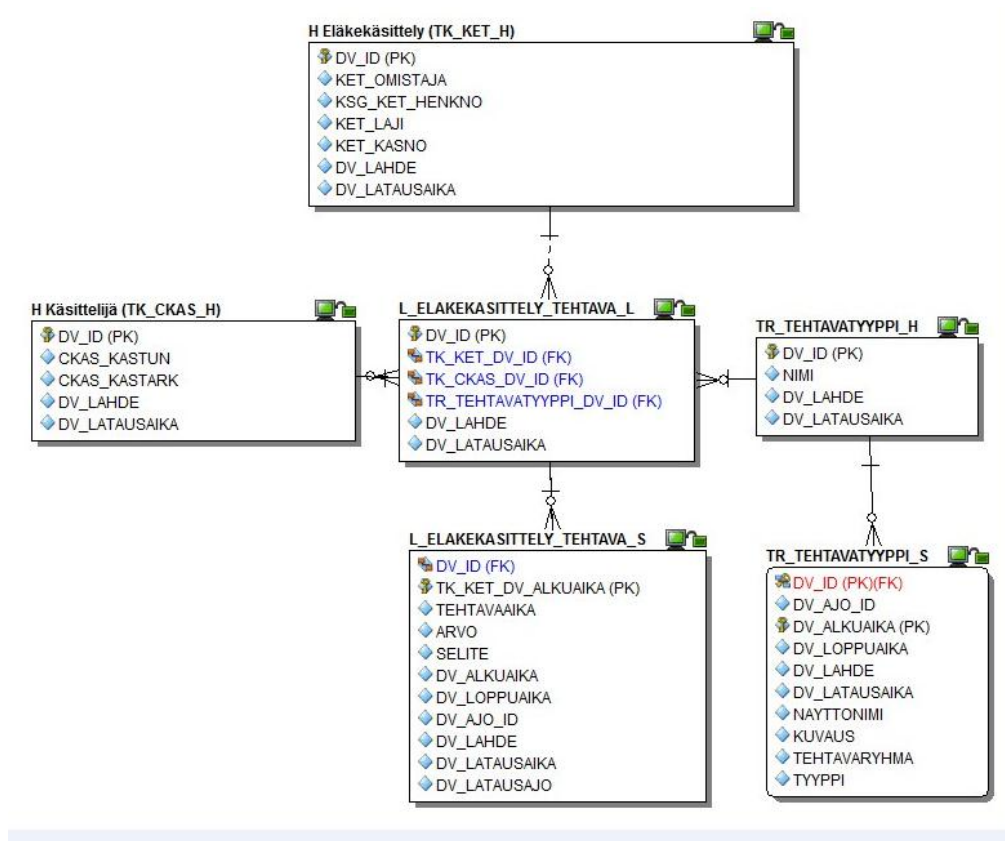
Käsittelijän tehtävät -käsitteen luominen lähti liikkeelle operatiivisen eläkekäsittelyjärjestelmän uusimishankkeessa, jossa hyödynnetään prosessiajattelua järjestelmän toteuttamisessa. Uusimishankkeen edetessä havaittiin, että käsittelijän tehtävät ajattelumalli kannatti siirtää osaksi myös nykyisin käytössä olevan eläkekäsittelyjärjestelmän (K) tietovarastoon tuotavia tietoja. Jotta ajattelumalli saatiin koskemaan myös nykyistä eläkekäsittelyjärjestelmää (K), lähdettiin asiaa lähestymään siten, että järjestelmässä tehdyistä toimenpiteistä irrotetaan tehtäviä, jotka liittyvät eläkeprosessin eri vaiheisiin (perustaminen, ratkaisu, päättäminen, asiakkaalle lähettäminen). Periaatteena on, että käsittelijän tehtävän mukainen tehtävä (esimerkiksi perustamistehtävä) muodostuu tietovarastoon vasta silloin, kun tietyt siihen liittyvät ehdot täyttyvät.

Tehtävä syntyy, kun käyttäjä on vinyt manuaalisesti tai järjestelmä on automaattisesti päivittänyt päivämäärätiedon eläkekäsittelyjärjestelmän (K) keskusteluun. Joissakin tehtävissä edellytetään myös muiden järjestelmästä löytyvien ehtojen täyttymistä ennen kuin siitä muodostuu käyttäjälle tehtävä. Tehtävät tallennetaan päivittäin tietovarastoon. Nämä uudet käsittelijätehtävät ovat aina tiettyä tehtävätyyppiä, joka on ryhmitelty tehtäväryhmään tehtävän vaiheen mukaisesti. Eläkekäsittelyyn liittyviä vaiheita ovat muun muassa käsittelyn perustaminen, tiedonkeruu, ratkaisu, päätöksen tekeminen sekä käsittelyn päättäminen. Käsittelijän eläkekäsittelytehtävän jakautumista tehtäväryhmiin ja -tyyppeihin on havainnollistettu taulukossa 5. Tehtävätyypit tallennetaan tiedostoon, josta ne ladataan tietovarastoon.

Taulukko 5. Käsittelijän eläkekäsittelytehtävän jakautuminen tehtävätyyppeihin

Tehtäväryhmä	Tehtävätyyppi	Kuvaus
Alkukäsittely	Tiedonkeruu	Tiedonkeruu jonolla olevia uusia hakemuksia, jotka odottavat virittämistä lähdejärjestelmään (K)
Päätöksen anto	Päätöksenanto, myöntö	Kaikki päätökset yhteensä
Päätöksen anto	Päätöksenanto, hylky	Kaikki hylyt yhteensä
Käsittelyn päättäminen	Eläkekäsittelyn päättäminen	Tilastojaksolla löytyy käsittelyn päättymispäivä lähdejärjestelmästä (K)
...

Tämä uusi käsittelijän eläkekäsittelytehtävä luo yhteyden eläkekäsittelyn, käsittelijän ja tehtävätyypin välille. Käsittelijä voi muuttua yhden käsittelyn sisällä ja käsittelystä voi muodostua useampia tehtäviä, jotka voivat olla samaa tyyppiä. Kyseessä on tapahtumätiedon tallennus tietyllä ajanhetkellä. Seuraavassa kuviossa 17 on kuvattu eläkekäsittelytehtävä-linkin rakenne Data Vault -liiketoimintakerroksessa sekä siihen liittyvät yhteydet.



Kuvio 17. Eläkekäsittelytehtävä-linkki (L_Elakekasittely_Tehtava_L) ja siihen liittyvät yhteydet

Kuten kuvio 17 käy ilmi, niin hubien H_Kasittelija, H_Elakekasittely ja TR_Tehtavatyypit_H välille on luotu yhteys käsittelijätehtävä-linkkitaulun avulla L_Elakekasittely_Tehtava_L. Linkkitaulun eläkekäsittelytehtävään liittyvät tarkemmat attribuutit sijaitsevat satelliitissa L_Elakekasittely_Tehtava_S.

Lisäksi eläkekäsittelyn ja käsittelijän välillä on olemassa yhteys (linkki), joka on tietyn roolin mukainen. Käsittelijällä voi olla neljää erilaista roolia, jotka on kuvattu taulukossa 6. Lähdejärjestelmässä (K) kerätään eläketapahtumatauluun kaikki tiedot eri käsittelijöistä ja siitä, mitä he ovat tehneet käsittelytehtävän aikana. Tauluun tallennetaan myös roolitieto. Rooleista kaksi liittyy käsittelyn hoitajaan, joka on se henkilö, joka saattaa käsittelyn loppuun. Lisäksi käsittelijä voi olla perustajan tai ratkaisijan roolissa. Kohdeyhteyksessä käsittelijät hoitavat sekä kohdeyhteyksen että erään toisen yrityksen eläkekäsittelytapauksia eri tietokannoissa. Tästä johtuen käsittelyn hoitajarooleja tarvitaan kaksi erilaista, jotka kuitenkin ovat keskenään samanarvoisia. Tietovarastosta raporteille poimitaan vain kohdeyhteyksen käsittelijöiden tekemät tehtävät ja ne vastaavat muun

muassa kysymyksiin siitä, kuinka monta eläkekäsittelyä tietty henkilö on perustanut tai kuinka monta eläkekäsittelyä tietty henkilö on päättänyt?

Roolilinkkien kytkeytyminen muihin tietokannan tauluihin on kuvattu tarkemmin liitteessä 2, josta selviää myös Data Vault -tallennuskerroksen rakenne.

Taulukko 6. Käsittelijän erilaiset roolit

Rooli (linkki)	Statussatelliitti	Kuvaus
TK_KET_CKAS_ KASTUN_HOIT_L	TK_KET_CKAS_ KASTUN_HOIT_SS	Eläkeasiaa hoitavan käsittelijän tunnus. Tieto ei ole aina sama kuin viimeisin muutoksentehtävä (KET-MUTUN).
TK_KET_CKAS_ KASTUN_EPS_L	TK_KET_CKAS_ KASTUN_EPS_SS	Eläkeasiaa hoitavan käsittelijän tunnus. Tieto ei ole aina sama kuin viimeisin muutoksentehtävä (KET-MUTUN) tai eläkeasiaa hoitavan käsittelijän tunnus (KET-KASTUN-HOIT). Tieto on käytössä vain, jos omistaja saa tietyt arvot.
TK_KET_CKAS_ PERTUN_L	TK_KET_CKAS_ PERTUN_SS	Käsittelijän tai ohjelman tunnus, joka on perustanut tiedot.
TK_KET_CKAS_ RATKAISIJA_L	TK_KET_CKAS_ RATKAISIJA_SS	Eläkeasian ratkaisijan tunnus.

Data Vault -ratkaisussa roolien yhteys on luotu käsittelijä- ja eläkekäsittely-hubin välille. Eläkekäsittelytehtävä-linkki on sen sijaan kytketty suoraan käsittelijä-hubiin, jota kautta roolitieto on raportoitavissa tehtävittäin. Roolien toteutus on ratkaistu näin siksi, että avaimet pysyvät yksiselitteisimpänä eikä ratkaisu rajaa liikaa eläkekäsittelytehtävälinkin muodostamista siihen kytkeytyvien linkkien kautta. Mikäli yhteys vaihtoehtoisesti olisi muodostettu kahden linkkitaulun välille (eläkekäsittelytehtävä ja rooli), olisi linkkejä muodostunut huomattavasti enemmän ja ne olisivat hankaloittaneet toteutusta. Kuten aiemmin on jo mainittu kappaleessa 2.4.4(Data Vault -tietovaraston rakentaminen), niin Data Vault -mallinnus mahdollistaa yhteyden kahden linkin välille, mutta sen käyttäminen ei ole suositeltavaa.

Lähdejärjestelmässä (K) ei siis valmiina ole olemassa tietoa käsittelijän tehtävään liittyen, vaan sieltä ladatuista tiedoista muodostetaan liiketoimintakerrokseen tietyillä säännöillä käsittelijäntehtävä, johon yhdistetään tietyn roolin omaava henkilö sekä tieto siitä,

mitä hän on tehnyt. Ladattaessa tietoja eläkekäsittely-satelliitista saadaan rivin muutos-päivä, joka kertoo, milloin tehtävä on tehty. Tämä tekopäivä, jolloin tehtävä on tapahtunut, yhdistetään roolin mukaiseen käsittelijä-linkkiin. Esimerkiksi ratkaisija-linkin (TK_KET_CKAS_RATKAISIJA_L) kautta saadaan tietää, kuka käsittelijä teki kyseessä olevaan eläkeasian käsittelyyn liittyvän ratkaisun ja statussatelliitin (TK_KET_CKAS_RATKAISIJA_SS) kautta, milloin hän sen teki. Tästä muodostuu rivi eläkekäsittelytehtävä-linkkiin siten, että tekopäivä tallennetaan linkin satelliittiin tekopäiväksi tietyillä säännöillä, jotka varmistavat, että tehtävä kirjautuu oikealle päivälle. Näin saadaan raportille se ajanhetki, jolloin tehtävä on tapahtunut.

Tehtävä syntyy, kun käyttäjä on vienyt (tai järjestelmä on automaattisesti päivittänyt) päivämäärätiedon ja käyttäjätunnuksen lähdejärjestelmässä (K) oleviin kenttiin. Joissakin tehtävissä edellytetään myös muiden järjestelmästä löytyvien ehtojen täyttymistä ennen kuin siitä muodostuu käyttäjälle tehtävä. Seuraavissa taulukoissa on kuvattu ehdot, joiden perusteella esimerkiksi tiedonkeruutehtävä muodostetaan. Ylemmässä taulukossa 7 olevat tiedot selventävät liiketoiminnan määrittelijälle, mihin kokonaisuuteen tehtävä liittyy. Alempi taulukko 8 sisältää ehdot, joiden tulee täytyä, jotta tehtävä voidaan muodostaa. Määrittelyssä on käytetty lähdejärjestelmän (K) kentänimiä, jotta määrittelijöiden olisi helpompi tunnistaa, mistä tiedosta on kyse.

Taulukko 7. Tiedonkeruutehtävän määrittelyä

Tieto	Selitys
Liittyvä asia	Eläkekäsittely
Tunnus	EK_TIEDONKERUU
Nimi	Tiedonkeruu
Määrittely	Sisältää Käsittelyn perustamisen, virittämisen ja muita toimenpiteitä ennen K160/K161 päivittämistä.
Tehtäväryhmä	Alkukäsittely
Liittyvät käsitteet	
Toteutuksen status	Toteutettu

Taulukko 8. Tiedonkeruutehtävään liittyvät ehdot

Tieto	Selitys	Kuvaus
Tehtävän käsittelijä	Perustajatunnus	KET_PERTUN
Seurattava pvm	Perustamispäivä	KET_PERPV
Tekopvm	K-järjestelmän muuttumispvm: KET -taulusta, KTA:n ja KET:n järjestyksellä ei ole väliä, pääasia että se löytyy Oletus: KTA keskustelu päivittää myös KET_MUPV 1.11.2011: seurattavissa tehtävissä K160/K161 tehty aina jos PERPV asetettu	KET_MUPV
Muut rajaukset	K160 tai K161 tehty KET-VIRETK=2	KET_VIRETK=2 KTA -taulusta K160 tai K161

Edellä kuvatut uudet käsitteet on myös määriteltävä ja toteuttava tietovarastoon, jotta niitä voidaan käyttää raportoinnissa. Näin ollen käsittelijän eläkekäsittelytehtävät talletetaan päivittäin tietovarastoon. Niihin liittyvät tiedot on ladattu kantaan jo lokakuusta 2011 lähtien. Käsittelijän eläkekäsittelytehtävien ensimmäinen virallinen raportointivuosi on 2012. Ne ovat käytössä uudesta tietovarastosta toteutetuilla osastokohtaisilla kokoomilla sekä käsittelyn virittäminen ja päättäminen raporteilla. Lisäksi tietokantaan lisätty lukitusominaisuus varmistaa sen, että aikaisempien kausien tiedot eivät muutu raporteilla sekä sen, että tekohetken organisaatio on oikein. Näin varmistetaan, että käsittelijän tekemä työ näkyy raporteilla siinä tiimissä tai osastolla, jossa hän tekohetkellä on ollut.

Raporttien käytöstä pidettiin koulutus kohdeyrityksen eläkekäsittelijöille. Liitteenä 3 olevassa koulutusmateriaalissa on esimerkkejä tilanteista, joissa käsittelijätehtävää on hyödynnetty raporteilla. Kyseinen liite sisältää luottamuksellista tietoa ja sitä voi erikseen tiedustella tutkimuksen laatijalta. Koulutusmateriaalissa olevilla esimerkeillä havainnollistetaan eroja vanhan ja uuden tietovarastoraportoinnin välillä. Niistä selviää muun muassa, miten vanhassa raportoinnissa kuukauden lopun tilanne lähdejärjestelmässä (K) määräsi kenelle "tekeminen" kirjautui. Lähdejärjestelmään (K) viety tehtäväpäivä, kuten käsittelyn päättymispäivä tai ratkaisupäivä määräsi, minkä kuukauden raportilla "tekeminen" näytettiin. Kun taas uudesta tietovarastosta toteutetuilla raporteilla, tehtäviä seurataan päivätasolla ja ne kirjautuvat tekopäivän mukaiselle kuukaudelle. Tekopäivä on todellinen päivä, jolloin käsittelijä on tehnyt tehtävän lähdejärjestelmässä

(K). Tehtäväpäivä on päivämäärä, joka kuvaa käsittelyn toteuttamiseen liittyvää päivämäärää, kuten mille päivämäärälle käsittely päätettiin tai ratkaistiin.

Uuden raportoinnin osalta on päätetty, että riittävä tarkkuustaso raporteille on päivätaso, vaikka Data Vault -menetelmä mahdollistaisi päivämäärissä kellonaikojen käytön. Päätökseen vaikuttaa, että latausajot perustuvat operatiivisten järjestelmien aikatauluihin sekä työn toteutukseen. Tästä johtuen tietovarastolataukset tehdään kerran vuorokaudessa ja vasta, kun kaikki edellisen päivän käsittelyt on tehty. Tallennuskerroksessa oleva raakadata sisältää alkuperäiset päivämäärätiedot ja tarvittaessa muokkaukset päivämääriin tehdään joko liiketoimintakerroksessa tai Data Martti -raportointikerroksessa.

Raportilla esille tuoduista esimerkkitalanteista käy ilmi, että Data Vault -menetelmällä toteutettu tietovarastoratkaisu on tuonut huomattavia parannuksia kohdeyrityksen tietovarastoraportointiin eläkekäsittelyn osalta. Liiketoiminnan luottamus raporteihin on kasvanut, kun tietosisältö on luotettavaa eikä muutu raportointiajankohdan mukaan. Tästä hyvänä esimerkkinä ovat tietojen historiointi ja jäljitettävyyys sekä lukitusominaisuudet, jotka saatiin käyttöön uuden tietovaraston myötä. Lisäksi Data Vault -menetelmän avulla tehty ratkaisu tietovaraston liiketoimintakerrokseen on ratkaissut tarpeen saada selville, ketkä käsittelijät ovat osallistuneet kyseessä olevan tehtävän käsittelyyn ja milloin. Tämä uusi käsite tuo lisäarvoa liiketoiminnalle ja mahdollistaa uudenlaisen käsittelijän eläkekäsittelytehtävään liittyvän historiaraportoinnin liiketoiminnan näkökulmasta.

5 Empiria tutkimuksen kohteena olevasta raportointialueesta

Kokemuksia raportoinnin kehittamisestä ja muutostarpeiden läpiviennistä selvitettiin strukturoitujen teemahaastattelujen avulla. Tällöin haastateltavilla oli mahdollisuus käsitellä asiaa laajasti, monitahoisesti ja syventävästi. Haastattelurunko koottiin tutkimusongelman näkökulmasta huomioiden Data Vault -menetelmään perustuvan konstruktion toteuttamisen yhteydessä esiin nousseet keskeiset tekijät. Haastattelut toteutettiin kahtena eri verkkokyselynä tutkimuksen alussa ja lopussa. Tutkimuksen avulla selvitettiin sekä raportoinnin käyttäjien että kehittäjien kokemuksia tarkastelun kohteena olleen eläkepalvelujen tietovarastopohjaisen raportoinnin kehittamisestä ja muutostarpeiden läpiviennistä. Lisäksi haastattelujen lopputuloksena saatiin arvokasta tietoa kohdeyritykselle Data Vault -menetelmän hyödyntämisestä.

5.1 Haastattelujen toteutus

Kohdeyrityksessä tutkimuksen aikana käynnissä olleen raportoinnin kehittämishankkeen yhteydessä rakennettiin uusi raportointitietovarasto hyödyntäen Data Vault -menetelmää tietojen mallintamisessa. Tutkimuksen ensimmäisessä haastatteluvaiheessa selvitettiin kokemuksia tietovarastoraportoinnin nykytilasta ja toisessa vaiheessa tavoiteltiin sen jälkeen, kun uuden mallinen tietovarasto oli otettu käyttöön. Tutkimuksen haastattelukyselyt koostuivat enimmäkseen monivalintakysymyksistä ja vastaaminen vei aikaa noin kymmeneen minuuttia yhtä kyselyä kohti. Kyselyjen vastaukset käsiteltiin luottamuksellisesti ja tuloksia hyödynnettiin vain tutkimuksen johtopäätösten tekemisessä ja kohdeyrityksen tietovarastoraportoinnin kehittämisessä.

Tutkimuksen kokonaisaikataulusta johtuen vertailuraporttien määrä jäi vain kolmasosaan suunnitellusta, sillä kohdeyrityksen kehittämishanke, jossa raportit toteutettiin, viivästyi aikataulustaan noin puolella vuodella. Tutkimuksen aikataulusta johtuen kuitenkin päätettiin, että tutkimus viedään loppuun vielä vuoden 2012 aikana. Haluttaessa tutkimuksen tavoitetilan haastattelu voidaan myös uusida vuoden tai pari kuluttua, mikäli se nähdään tarpeelliseksi tietovaraston kehittämisen näkökulmasta.

5.2 Haastateltavien valinta

Haastateltavaksi valittiin tietohallinnosta, toimittajalta sekä tutkimuksen kohteena olevan raportointialueen liiketoiminnasta sellaisia henkilöitä, joilla nähtiin olevan keskeinen rooli tietovarastoraportoinnin käyttämisessä ja kehittämisessä liiketoiminta tiedon hyödyntämisen näkökulmasta. Sopivien henkilöiden valinta tehtiin yhdessä kohteena olleen raportointialueen ylläpidosta vastaavan sovellusyhteyshenkilön ja IT-vastaavan kanssa, jotka myös itse osallistuivat tutkimuksen haastattelukyselyihin vastaamiseen.

Haastateltavaksi valitut henkilöt jaettiin rooleihin sen mukaan, miten he osallistuvat raporttien käyttöön ja ylläpidon toteutukseen. Haastattelukysymykset kohdistettiin henkilötasolle, jotta vastauksia voitiin käsitellä vastaajien roolien näkökulmasta. Roolit on kuvattu tarkemmin taulukossa 9.

Taulukko 9. Haastateltavien roolit

Rooli	
Liiketoimintatilaaja	Hyödyntää raportteja päivittäisessä työssään ja välittää asiakkailta tulleita kehitysehdotuksia.
Ylläpito / liiketoiminta	Osallistuu raporttien hyödyntämisen lisäksi aktiivisesti niiden ylläpitoon.
Ylläpito / tekninen / liiketoiminta	Henkilö tietohallinnosta, joka osallistuu aktiivisesti raporttien ylläpitoon yhdessä liiketoiminnan kanssa.
Ylläpito / tekninen	Henkilö toimittajalta, joka osallistuu aktiivisesti raporttien ylläpitoon määritysten pohjalta.

Lisäksi verkkokyselyn toimivuus ja sisältö testattiin etukäteen kahden kyselyyn osallistuneen henkilön kanssa. Näiden ennen varsinaisia haastattelukyselyjä suoritettujen harjoitusten avulla haastattelurunko hiottiin mahdollisimman toimivaan ja tehokkaaseen muotoon.

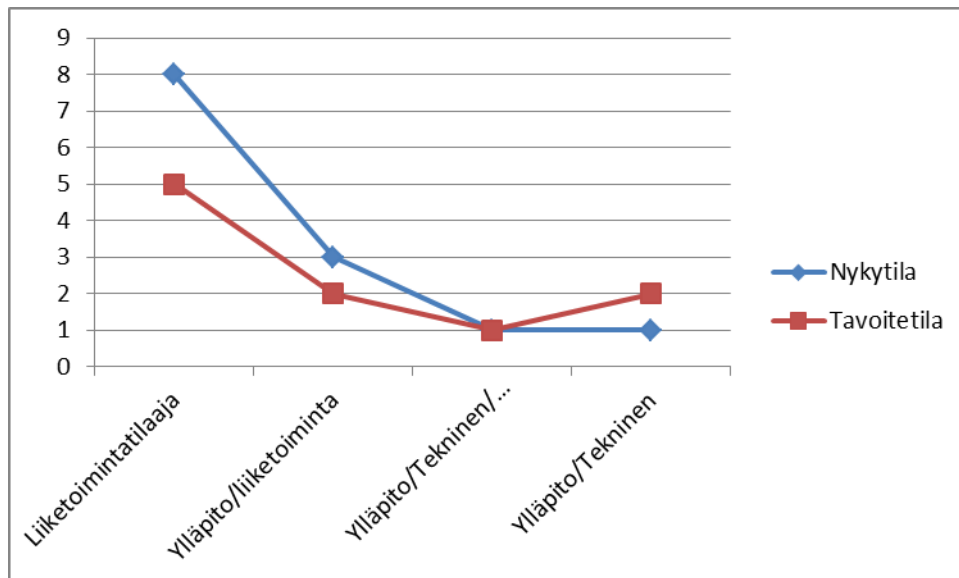
5.3 Nyky- ja tavoitetilan haastattelujen vertailu

Tähän raporttiin on koottu vain vertailu nyky- ja tavoitetilan haastatteluanalyysien lopputuloksista. Varsinaiset nyky- ja tavoitetilan haastatteluanalyysit ovat luettavissa liitteessä 4. Kyseinen liite sisältää luottamuksellista tietoa ja sitä voi erikseen tiedustella tutkimuksen laatijalta.

Eläkepalvelujen tietovarastopohjaisen raportoinnin kehittämisen ja muutostarpeiden läpiviennin nyky- ja tavoitetilaa selvitettiin strukturoitujen teemahaastattelujen avulla. Nykytilan haastattelu toteutettiin verkkokyselynä maaliskuussa 2011 ja tavoitetilan haastattelu lokakuussa 2012. Kyselyissä lähetettiin osallistujille saatekirjeet sähköpostitse sekä linkki, joka ohjasi kyselylomakkeisiin kohdeyrityksen sisäisessä intranetissä. Tutkimuksen kohteen nyky- sekä tavoitetilaa selvittävät kyselyt kohdistettiin raportoinnin käyttäjille ja kehittäjille. Nykytilan kyselyajankohta kohdistui vuosille 2005 - 2010, jotta vastauksiin saatiin enemmän kokemuspohjaa. Tavoitetilan kyselyajankohta kohdistui ajalle 2012 kevät ja syksy, sillä tutkimuksen kohteena olevat Data Vault -menetelmään perustuvat raportit otettiin käyttöön kevään ja kesän 2012 aikana.

Tutkimuksen nykytilakysely lähetettiin 18 henkilölle, joista viisi osallistui tutkimuksen kohteina olleiden raporttien hyödyntämisen lisäksi myös aktiivisesti niiden kehittämiseen. Nykytilakyselyyn vastasi 13 henkilöä, vastausprosentin ollessa siten 72 %. Tavoitetilakysely kohdistettiin osittain samoille henkilöille ja se lähetettiin 13 henkilölle, joista myös viisi osallistui raporttien hyödyntämisen lisäksi aktiivisesti myös niiden kehittämiseen. Tavoitetilakyselyyn vastanneita oli 10 ja vastausprosentti oli 76 %. Kummankin kyselyn vastausprosenttia voidaan pitää hyvänä tuloksena näin rajatussa otannassa. Molemmista kyselyistä saadut vastaukset koottiin yhteen ja analysoitiin.

Alla olevassa kuviossa 18 havainnollistetaan vastaajien jakautumista liiketoiminnan käyttäjiin ja ylläpitäjiin sekä teknisen ylläpidon edustajiin. Tavoitetilan kyselyyn osallistui liiketoiminnasta viisi henkilöä vähemmän, mutta muuten kyselyjen vastaajamäärissä ei ollut suurta eroa.



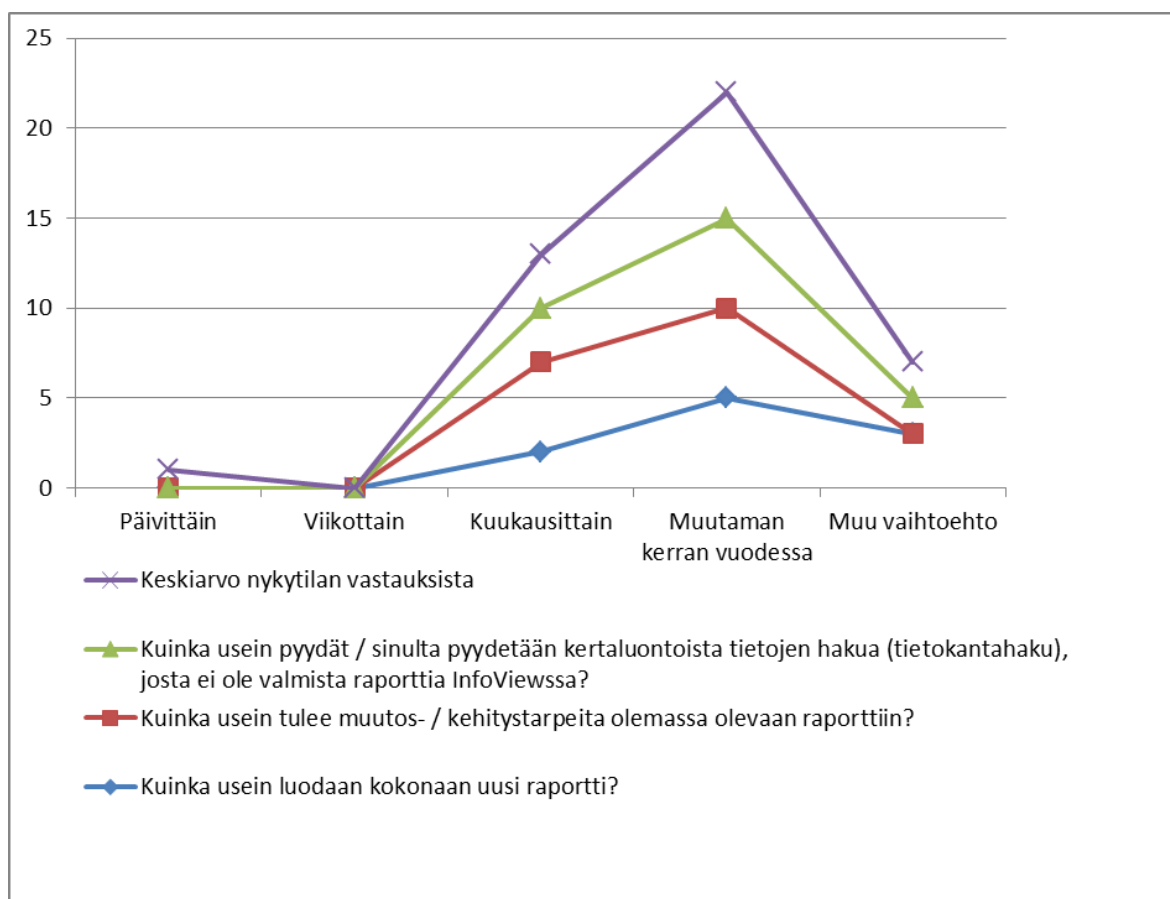
Kuvio 18. Haastattelukyselyjen vastaajamäärät rooleittain nykytilan (N = 13) ja tavoite-tilan (N=10) osalta

Eläkepalvelujen uudistettua tietovarastoa hyödyntävillä raporteilla korvattiin aiemmin käytössä olleita raportteja sekä luotiin uusia raportteja. Näin ollen molemmissa kyselyissä olivat kysymykset, jotka liittyivät ylläpidettävien raporttien ilmenemistiheyteen, raportoinnin yleisimpiin ongelmiin, ylläpidon toteutuksen muutosjoustavuuteen ja helpouteen sekä kokemuksiin raportoinnista. Näiden lisäksi kummassakin kyselyssä oli "sana on vapaa" -kysymys, johon tulleet vastaukset liittyivät lähinnä kohdeyrityksen ylläpidon resursointiin sekä palautteeseen ylläpidon toteuttajille. Edellä mainittujen kysymyskokonaisuuksien lisäksi nykytilan kyselyssä selvitettiin, millaisia raportteihin tulevat muutostarpeet ovat sekä tavoitetilan kyselyssä selvitettiin, miten Data Vault -menetelmän käyttöönotto on vaikuttanut raportointiin. Seuraavissa kappaleissa vertailaan näiden kysymyskokonaisuuksien tuloksia toisiinsa huomioiden tutkimusongelmaan näkökulma siitä, parantaako Data Vault -menetelmä tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta.

5.3.1 Ylläpidettävien raporttien muutospyyntöjen ilmenemistiheys

Nykytilan kyselyssä selvitettiin, että keskimäärin yksi henkilö hyödyntää 11 - 25 raporttia työssään ja ylläpitoon osallistuvat henkilöt saattavat hyödyntää jopa yli 50 raporttia. Sekä nyky- että tavoitetilan kyselyjen vastauksista selviää, että sen jälkeen kun uudistettu raportointi otettiin käyttöön, on ylläpidettävien raporttien muutospyyntöjen ilmenemis-

tiheys vähentynyt ja niitä tulee harvemmin kuin aiemmin. Vastauksien jakaumaa on havainnollistettu kuviossa 19, jossa verrataan nykytilan vastausten keskiarvoa raporttimuutosten ilmenemistiheydestä tavoitetilan vastaaviin arvoihin. Samasta kuviosta selviää myös, että muutospyyntöjen ilmenemistiheydessä ei ole suurta eroa sen välillä, luodaanko kokonaan uusi raportti, tuleeko muutos olemassa olevaan raporttiin vai onko kyseessä kertaluontoinen tietojen haku.



Kuvio 19. Tavoitetilan (N = 10) vastausarvot suhteessa nykytilan (N = 13) vastausten keskiarvoon raporttimuutosten ilmenemistiheydessä

Lisäksi tavoitetilan kyselyssä selvitettiin, ovatko raporttimäärät joko vähentyneet tai pysyneet samoina uudistetun raportoinnin myötä. Hieman alle kahden kolmasosan vastanneen mielestä myös raporttimäärät ovat joko vähentyneet tai pysyneet samoina. Tämä on luultavasti seurasta siitä, että uudistetun raportoinnin myötä olemassa olevien raporttien tietoja koottiin laajemmille kokoomaraporteille, mikä osaltaan vähensi erillisten raporttien lukumäärää. Toisaalta kehittämishankkeen yhteydessä luotiin myös ko-

konaan uusia raportteja, jonka seurauksena raportoitavien tietojen osuus laajeni ja raporttimäärät kasvoivat esimerkiksi eläkekäsittelyraporttien osalta.

5.3.2 Nyky- ja tavoitetilan raportoinnin yleisimmät ongelmat

Sekä nyky- että tavoitetilan kyselyssä selvitettiin raportoinnissa ilmeneviä suurimpia ongelmia. Vastaajille annettiin yhdeksän eri vaihtoehtoa, joista heidän tuli valita kolme suurinta ongelmaa raportoinnissa. Lisäksi heillä oli mahdollisuus antaa jokin muu syy, jota ei vielä ollut vaihtoehtoissa. Annetut vaihtoehdot olivat seuraavat:

- Uuden raportin tekeminen vie liian kauan aikaa.
- On vaikea tietää, mitä tietoja raportille voi poimia.
- Raportille pyydettyjä tietoja ei löydy tietovarastosta.
- Raporttia ei saa halutuilla poimintaehdoilla.
- Raportin tietosisältö on epäselvä.
- Määrittelyjen tekeminen vie liian paljon aikaa.
- Raportin toteutus ei pysy aikataulussa.
- Raporttipyynnöt ovat epämääräisiä.
- Raportin toteutukselle ei anneta aikataulua.
- Muu syy.

Neljäs osa nykytilan kyselyn vastaajista koki suurimmaksi ongelmaksi, että on vaikea tietää, mitä tietoja raporteille voi poimia. Tämä ongelma oli tavoitetilan kyselyssä vasta toisena ja ainoastaan vähän yli kymmenesosa vastaajista enää koki siten. Seuraavaksi merkittävimmät ongelmat nykytilassa liittyivät raportoitaviin tietoihin ja niiden saataavuuteen, sillä kumpaakin vaihtoehtoa oli valittu yhden viidesosan verran. Nämäkään eivät enää olleet niin merkittäviä tavoitetilan kyselyssä. Edellisen pohjalta voidaan päätellä, että Data Vault -menetelmän käyttöönotto on helpottanut liiketoiminnan ymmärrystä tiedoista, jotka ovat käytettävissä tietovarastossa. Yksi osasy tähän on varmasti myös uuden liiketoiminnan käsitelmän käyttöönotto, joka vastaa tietovaraston käsitteistöä.

Tavoitetilan suurimmaksi ongelmaksi koettiin, että raporttien tekeminen vie liian kauan aikaa, kun taas nykytilan vastauksissa vain hieman yli yksi kymmenesosa oli kokenut sen ongelmaksi. Tämä kertoo siitä, että tavoitetilan merkittävimmät ongelmat heijastavat ensisijaisesti uuden raportoinnin käyttöönottoon liittyviä asioita. Näin ollen ongelmat eivät keskity tietojen saatavuuteen tietovarastosta kuten nykytilan raportoinnissa, vaan enemmänkin toteutuksen haasteisiin sekä raportointitarpeiden kuvaamiseen ja määrittelyyn. Varsinkin liiketoimintatilaajat edustajat kokivat, että uuden raportin tekeminen vie liian kauan aikaa, johon yhtenä osasyynä vaikutti selvästi kehittämishankkeen resurssien vähyys. Tätä koskien oli tullut useampi kommentti tavoitetilan "sana on vapaa"-kysymykseen.

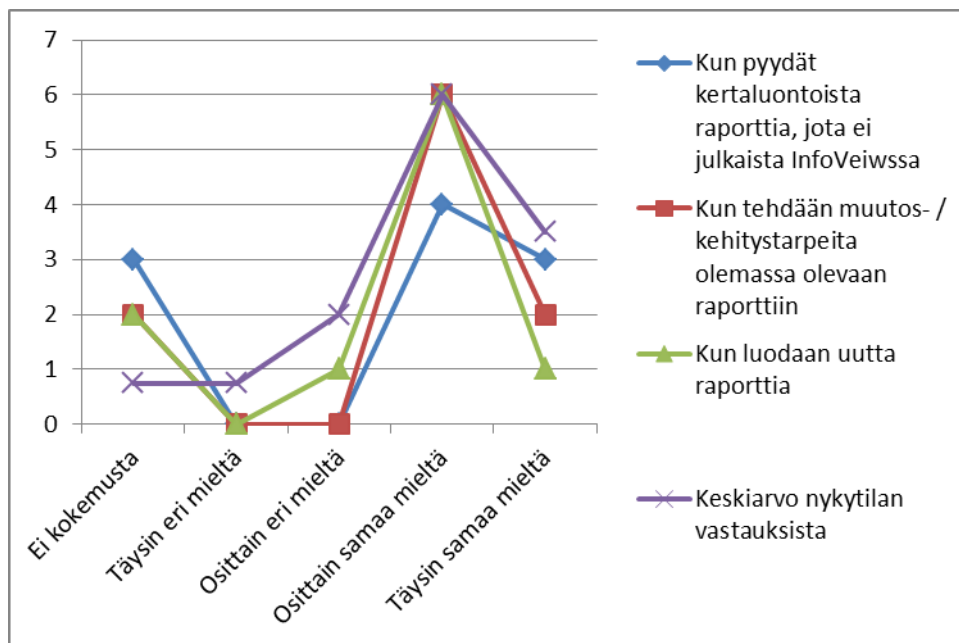
Nykytila-analyysin pohjalta voidaan todeta, että ongelmat aiemmassa raportoinnissa keskittyvät ensisijaisesti tietojen saatavuuteen tietovarastosta. Tietovaraston tulisi vastata juuri tähän tarpeeseen, jotta sitä hyödyntävät saisivat käyttöönsä tarvitsemansa tiedot oikea-aikaisesti. Vastauksista ilmenee myös, että nykyinen tietovaraston mallinnustapa ei ole kovin muutosjoustava, jolloin tietojen lisääminen tietovarastoon on hankalampaa ja kynnyks muutoksen toteuttamiseksi nousee. Näihin haasteisiin on selvästi vastattu uuden Data Vault -menetelmään perustuvan tietovaraston avulla, sillä tavoitetila-analyysistä selviää, että tietojen saatavuutta tietovarastosta ei enää koettu yhtä suureksi ongelmaksi.

5.3.3 Ylläpidon toteutuksen muutosjoustavuus ja helppous

Molemmissa kyselyssä selvitettiin raportoinnin ylläpidon toteutuksen muutosjoustavuutta ja helppoutta. Lisäksi selvitettiin, joudutaanko uuden raportin tai muutos- / kehitystarpeen yhteydessä lisäämään raportille tietoja, joita ei vielä löydy tietovarastosta ja kuinka helppoa ja nopeaa se on.

Raporttien toteutuksen muutosjoustavuutta ja helppoutta tietojen saannin osalta selvitettäessä, ei nyky- ja tavoitetilan välillä ollut merkittäviä eroja, kuten kuvioista 20 ilmenee. Vastajat arvioivat edellä mainittuja ominaisuuksia uuden raportin, muutos- tai kehityspyynnön sekä kertaluontoisen kyselyn näkökulmasta. Heistä kolme neljäsosaa oli sitä mieltä, että tiedot saadaan raporteille helposti ja joustavasti. Nykytilan kyselyssä

viidesosa vastaajista oli osittain eri mieltä siitä, että tietojen saanti raporteille olisi helpoaa ja joustavaa, kun taas tavoitetilan vastaajista enää vain alle kymmenesosa oli sitä mieltä. Näiden vastausten perusteella on nähtävissä muutosjoustavuuden paranemista juuri tietovarastosta saatavien tietojen osalta.

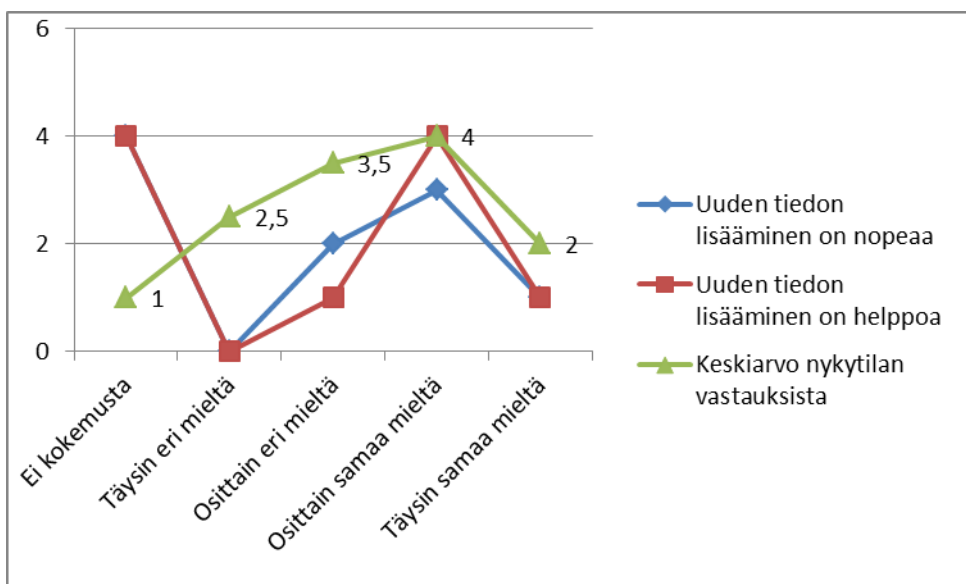


Kuvio 20. Tavoitetilan (N = 10) vastausarvot suhteessa nykytilan (N = 13) vastausten keskiarvoon tiedon saamisesta raportille helposti ja joustavasti eri toteutustarpeiden näkökulmista katsottuna

Seuraavaksi selvitettiin sitä, miten usein raporteille tarvitaan tietoa, jota ei vielä löydy tietovarastosta. Nykytilan vastauksista ilmenee, että vaihtoehtoja usein (23 %), toisinaan (23 %) tai harvoin (31 %) oli tullut keskimäärin saman verran kutakin, kun taas tavoitetilan vastaajista puolet koki, että tietoja on lisättävä vain toisinaan. Lisäksi tavoitetilan vastaajista jopa viidesosa oli sitä mieltä, että tietoja ei ole tarvinnut lisätä ollenkaan. Tähän vaikuttaa osaltaan yksi Data Vault -menetelmän ominaisuuksista, jossa tietovarastoon ladataan kaikki lähdejärjestelmissä oleva tieto raakadatkerrokselle. Siitä seuraa, että uusia tietoja ei tarvitse lisätä latauksiin, vaan riittää, että tietovarastossa jo valmiina oleva tieto nostetaan liiketoimintakerrokselle halutulla tavalla yhdisteltynä ja muokattuna.

Ylläpidon muutosjoustavuuteen ja helppouteen liittyen kysyttiin vielä, kuinka nopeaa ja helppoa uuden tiedon lisääminen tietovarastoon yleensä on (kuvio 21). Huomioitavaa

on, että tavoitetilan liiketoiminta vastaajista 40 % ilmoitti, että heillä ei ole kokemusta asiassa. Näin ollen tavoitetilan vastaajat tämän kysymyksen osalta kuuluivat ensisijaisesti henkilöihin, jotka osallistuvat raporttien ylläpitoon. Nykytilan kyselyssä vastaukset jakaantuivat tasaisemmin eri roolien kesken. Sekä nyky- että tavoitetilan kyselyyn vastanneista melkein puolet oli osittain tai samaa mieltä siitä, että uuden tiedon lisääminen on nopeaa ja helppoa. Vain tavoitetilan osalta tuli jonkin verran eroa sen välille millaisena he kokivat tiedon lisäämisen, sillä vastaajista 10 % enemmän oli sitä mieltä, että uuden tiedon lisääminen tietovarastoon on enemmän helppoa kuin nopeaa. Tästä voi päätellä, että Data Vault -menetelmän rakenne tekee tiedon lisäämisestä helppoa, mutta tietojen lisäämisen nopeus on hieman laskenut nykytilaan verrattuna. Vastaajien kokemus nopeuden laskusta liittyy todennäköisesti kehittämishankkeen resursoinnin haasteisiin tai siihen, miten hyvin vastaajat tuntevat kannan tietosisällön.



Kuvio 21. Tavoitetilan (N = 10) vastausarvot suhteessa nykytilan (N = 13) vastausten keskiarvoon tiedon saamisesta raportille helposti ja joustavasti

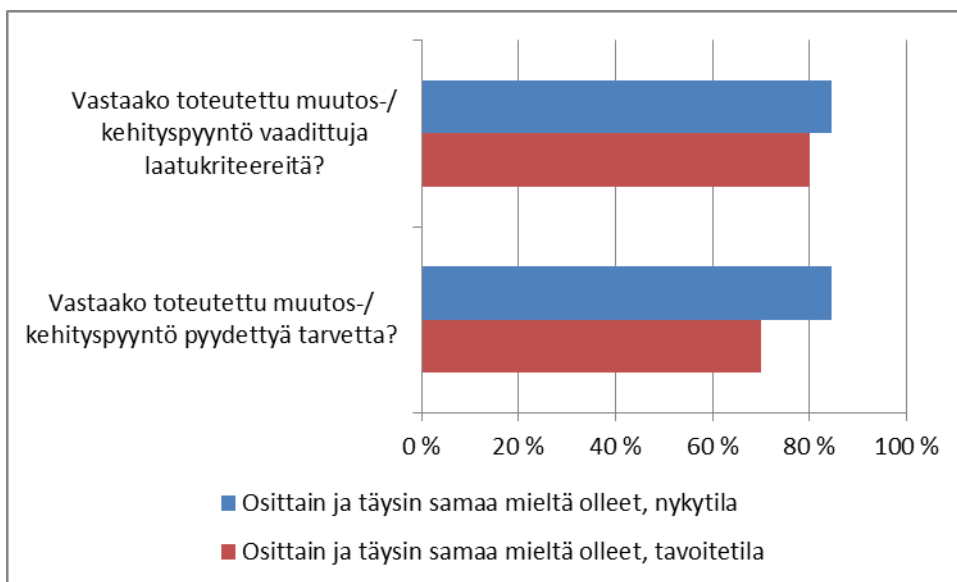
5.3.4 Kokemuksia raportoinnin ylläpidosta

Kummankin haastattelukyselyn lopuksi kartoitettiin vastaajien kokemuksia raportoinnin kehittämisestä ja ylläpidosta sekä siitä, onko se kaiken kaikkiaan nopeaa ja joustavaa. Lisäksi vastaajilta kysyttiin yleisarvosana raportoinnin ylläpidosta, johon sisältyvät kokonaan uudet raportit sekä muutos- ja kehitystarpeet olemassa oleviin raportteihin.

Selvitettäessä kokemuksia nyky- ja tavoitetilan raportoinnin ylläpidosta ja niihin liittyvistä muutos- ja kehityspyynnöistä, annettiin vastaajille kolme väittämää, joista he olivat joko eri tai samaa mieltä. Väittämät olivat:

- Tarvitseeko raportteihin tulevia muutos- / kehityspyynnöjä suunnitella?
- Vastaako toteutettu muutos- / kehityspyyntö vaadittuja laatukriteereitä?
- Vastaako toteutettu muutos- / kehityspyyntö pyydettyä tarvetta?

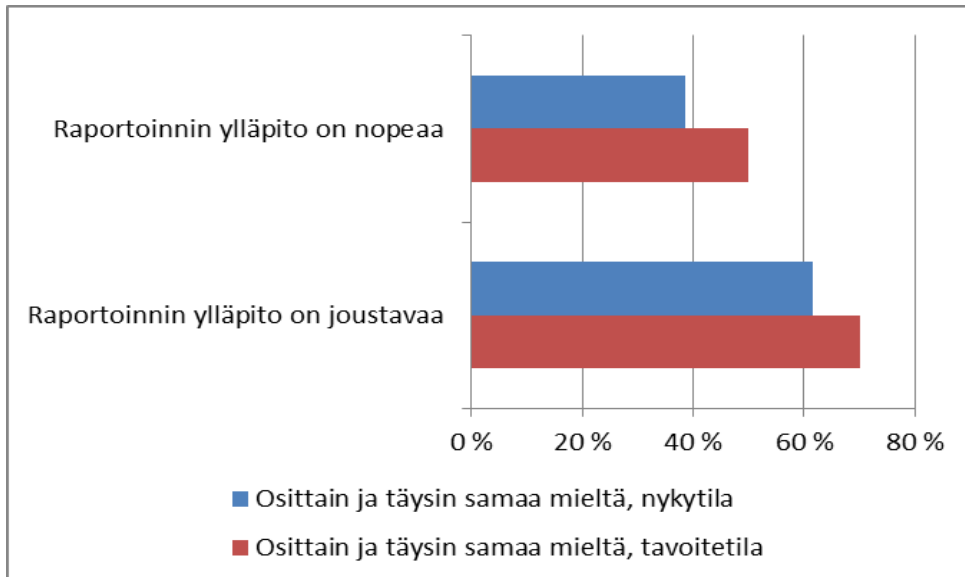
Väittämä, joka koski muutos- ja kehityspyynnöjen suunnittelua kysyttiin vain nykytilan kyselyn yhteydessä. Vastaajista 85 % koki, että pyynnöt tulee suunnitella, sillä he olivat joko osittain tai täysin samaa mieltä väittämän kanssa. Kahden muun väittämän osalta, joita kysyttiin sekä nyky- että tavoitetilan kyselyn yhteydessä, tuli arvoihin hieman eroja. Vastausarvot laskivat jonkin verran tavoitetilakyselyn yhteydessä, kun tarkastellaan asiaa joko täysin tai osittain samaa mieltä olleiden vastaajien osalta. Kuviossa 22 havainnollistetaan, miten tavoitetilan kyselyssä laatukriteereiden osalta laskua oli 5 % ja pyydetyn tarpeen toteutumisen osalta 15 % nykytilaan nähden.



Kuvio 22. Osittain ja täysin samaa mieltä olleiden vastaajien erot tavoitetilan (N = 10) ja nykytilan (N = 13) vastausten välillä muutos- ja kehityspyynnöihin liittyvien väittämien osalta

Vastausarvojen laskuun vaikutti osaltaan myös se, että tavoitetilan vastaajista keskimäärin 20 %:lla ei mielestään ollut kokemusta esitetyistä väittämistä. Vaikka nyky- ja tavoitetilan vastaukset erosivat toisistaan edellä mainittujen prosentiosuuksien verran, niin näyttää siltä, että kohdeyrityksen käytössä olevat toimintatavat raporttimääritysten tekemiseen tukevat edelleenkin liiketoimintatarpeiden kartoitusta kattavasti ja laatukriteerit täyttäen.

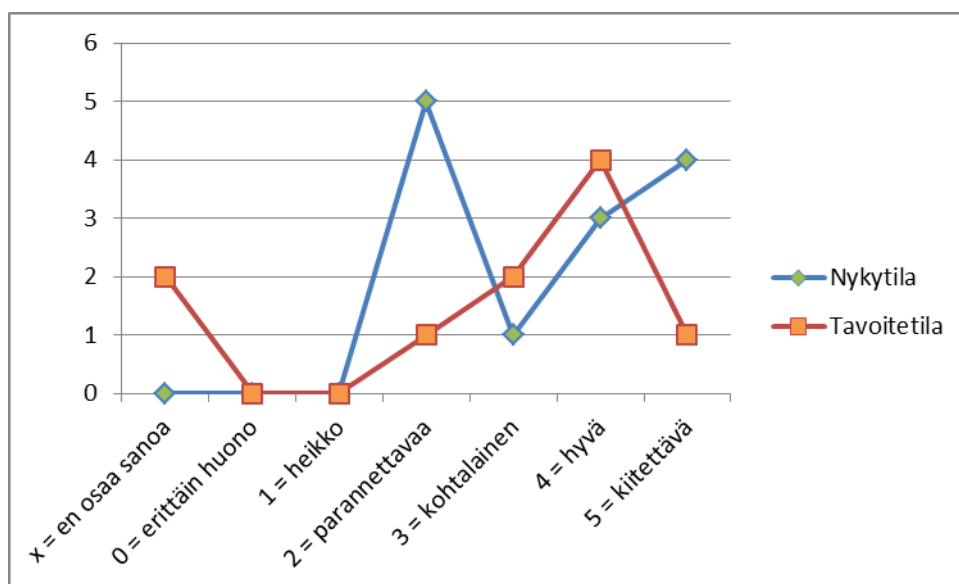
Seuraavaksi kysyttiin, onko raportoinnin kehittäminen ja ylläpito nopeaa ja joustavaa. Vastauksia tarkastellaan taas siitä näkökulmasta kuinka paljon vastaajat olivat asiasta osittain tai täysin samaa mieltä. Alla olevasta kuviosta 23 ilmenee, että kehittäminen ja ylläpito koettiin erittäin joustavaksi, mutta ei kovin nopeaksi. Kummatkin arvot ovat parantuneet uudistetun raportoinnin käyttöönoton myötä. Tästä voidaan päätellä, että tietovarastosta tehtävän raportoinnin muutosjoustavuus on lisääntynyt uuden Data Vault -menetelmän avulla. Varsinkin joustavuuden pysyminen hyvällä tasolla tukee tätä väittämää. Nopeuteen vaikuttaa hyvin usein muutkin asiat kuin vain tietovarastotasolle tehdyt kehittämistoimet.



Kuvio 23. Osittain ja täysin samaa mieltä olleiden vastaajien erot tavoitetilan (N = 10) ja nykytilan (N = 13) vastausten välillä ylläpidon nopeuden ja joustavuuden osalta

Vastaajia pyydettiin vielä antamaan yleisarvosana sekä nyky- että tavoitetilan raportoinnille erittäin huonosta (0) kiitettävään (5). Näiden lisäksi vastausvaihtoehtona oli "En osaa sanoa". Kuviosta 24 ilmenee, että vastausten perusteella raportoinnin taso on

enemmän kiitettävää ja hyvää, sillä keskimäärin 52 % vastaajista valitsi arvosanaksi hyvän (4) tai kiitettävän (5).



Kuvio 24. Raportoinnin yleisarvosanat tavoitetilan (N = 10) ja nykytilan (N = 13) vastausten osalta

Vastausten keskiarvo molempien kyselyjen osalta oli kuitenkin 3,55, joka tarkoittaa sitä, että tasossa on vielä parantamisen varaa. Hyvänä merkinä voidaan kuitenkin pitää sitä, ettei kummankaan kyselyn vastaajista kukaan pitänyt raportoinnin tasoa heikkona eikä erittäin huonona. Ne vastaajat (16 %), jotka olivat aiemmin kokeneet tason kohtalaiseksi tai sellaiseksi, että siinä on parannettavaa, olivat nyt todenneet tason parantuneen hyväksi tai kiitettäväksi.

Yhteenvetona kehittämisen ja muutosjoustavuuden osalta voidaan todeta, että niihin liittyvät käytännöt eivät ole huonontuneet nykytilanteeseen verrattuna. Tähän luultavasti vaikuttaa omalta osaltaan myös eläkepalveluissa tutkimuksen aikana toteutettu raportoinnin toimintatapojen, vastuiden ja roolien selkeyttäminen. Tämä toiminnan kehittämisprojekti on omalla tavallaan selkeyttänyt kehittämisen ja ylläpidon toteutusta ja tukenut uudistetun Data Vault -menetelmään perustuvan tietovaraston käyttöönottoa.

5.3.5 Millaisia raportteihin tulevat muutostarpeet ovat?

Nykytilan kyselyn yhteydessä selvitettiin myös yleisesti kehittämiseen ja ylläpitoon liittyvää asiaa. Vastaajilta kysyttiin, millaisia raportteihin tulevat muutostarpeet ovat. Vastauksista kävi ilmi hyvin tavanomaiset syyt sille, miksi raportteja tarvitsee muuttaa tai uudistaa. Vastaajat näkivät, että muutokset lähtevät sekä liiketoiminnan että asiakkaan tarpeista. Heidän mielestään, muutoksia aiheuttavat muun muassa seuraavat tilanteet:

- Raporteille halutaan lisätä luokittelutekijöitä, muutetaan ehtoja tai ryhmitellään tiedot uudella tavalla.
- Olemassa olevat raportit eivät täytä sen hetkisiä tarpeita ja on luotava kokonaan uusi raportti.
- Työn seurannan mittareita saatetaan muuttaa ja tietoja tulee saada asioista, joita aikaisemmin ei ole seurattu.
- Lähdejärjestelmiä saatetaan ryhtyä käyttämään uudella tavalla.
- Organisaatioon tai liiketoimintaprosesseihin tehdyt muutokset vaikuttavat myös olemassa oleviin raportteihin.

Kysymyksen yhteydessä selvisi myös haasteita, jotka liittyvät raporttimuutosten toteuttamiseen. Raportoinnin kehittäjät kokevat, että he joutuvat karsimaan toteutettavia muutoksia kohdeyrityksen vähäisten resurssien vuoksi. Lisäksi he kokevat, että uusien raporttien kehittäminen on hankalaa, sillä kohdeyrityksen lähdejärjestelmäosaamisesta on viime vuosien aikana vähentynyt paljon tietotaitoa. Myös tiedon mallintamiseen liittyen eräs vastaajista koki, että raporteille kaivataan selkeyttä. Hänen mielestään tällä hetkellä raporteilla on toisistaan poikkeavia lukuja samoista asioista, hakutekijöissä on puutteita ja tiedoissa turhia eroja. Lisäksi nykyisten raporttien tiedot saattavat muuttua, sillä nykyisessä tietovarastossa ei ole historiointia ja lähdedatasta näytetään vain uusin versio.

5.3.6 Data Vault -menetelmän tuomat vaikutukset raportointiin

Tavoitetilan kyselyn ehkä tärkeimpänä osana oli selvittää Data Vault -menetelmän tuomia vaikutuksia eläkepalvelujen raportointiin. Kyselyn yhteydessä kysyttiin kaikilta

(N = 10) vastaajilta uuden historiointiominaisuuden vaikutuksia raporteilla sekä poimitavien tietojen määrän muutosta raporteilla. Seuraavat Data Vault -menetelmän vaikutuksia selvittävät kysymykset kohdistettiin henkilöille, joiden rooli oli ylläpitäjä joko liiketoiminnassa, tietohallinnossa tai toimittajalla (N = 5). Tälle joukolle esitetyt kysymykset keskittyivät uudistetun raportoinnin myötä käyttöön otettuun eläkepalvelujen käsitelmalliin sekä liiketoiminnan käsitteiden yhteneväisyyteen tietokannan tietojen kanssa. Lisäksi selvitettiin kokemuksia mahdollisuudesta luoda uusia käsitteitä tietovarastoon liiketoimintatasolle.

Uudistetussa tietovarastossa otettiin käyttöön historiointiominaisuus, joka on yksi Data Vault -menetelmässä mukana tuleva ominaisuus. Menetelmässä historioidaan kaikki lähdejärjestelmistä tuleva tieto raakadatakerrokselle. Näin nämä historiointitiedot ovat haluttaessa käytettävissä myös liiketoiminta- ja raportointikerroksella. Luvussa 4.7 (Esimerkkejä Data Vault -menetelmän vaikutuksesta raportointiin) on myös käytännön esimerkkejä historioinnin vaikutuksesta raportointiin. Kyselyssä selvitettiin, miten historioinnin käyttöönotto on vaikuttanut raporttien tietosisällön oikeellisuuteen. Vastajat arvioivat kolmea väittämää skaalalla 1 - 4 (täysin eri mieltä - täysin samaa mieltä). Väittämät olivat seuraavat:

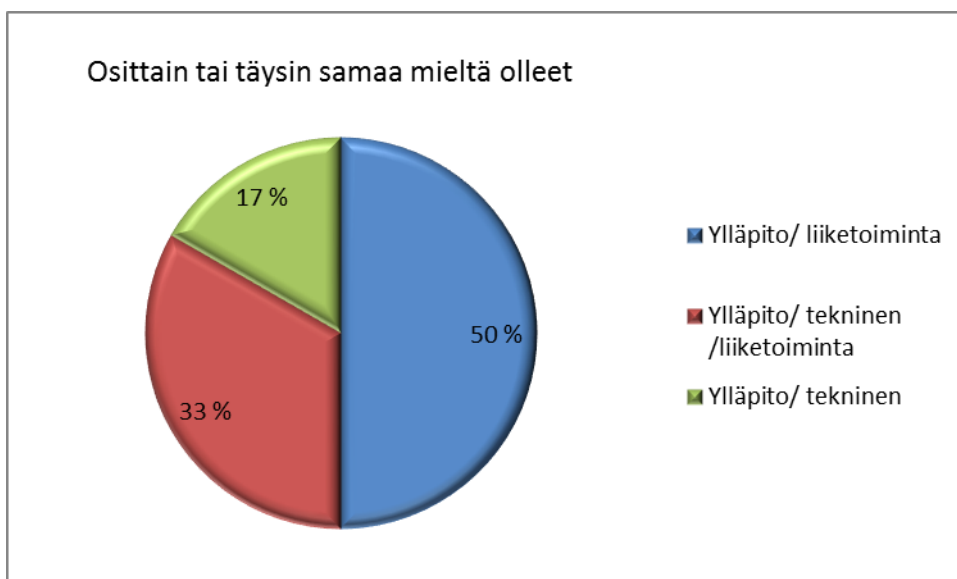
- Raporttien tiedot käsittelijän tekemästä työstä ovat tarkempia.
- Voin luottaa, että raporteilla olevat tiedot ovat oikein.
- Voin luottaa, että raporttien tiedot pysyvät samoina kuin aikaisempien kausien raporteilla.

Vastajat kokivat, että raporttien tiedot käsittelijän tekemästä työstä ovat tarkempia kuin aiemmin ja että raporttien tiedot eivät muutu aiempiin raportteihin nähden. Tämä näkyy siinä, että kaikki (100 %) vastaajat olivat joko täysin tai osittain samaa mieltä väittämien kanssa. Tietojen oikeellisuuden suhteen vastaajat olivat hieman kriittisempiä, sillä joko täysin tai osittain samaa mieltä väittämän kanssa olevia vastaajia oli vain 80 %. Näin ollen vastausten pohjalta voidaan päätellä, että historiointiominaisuuden käyttöönotto on lisännyt raporttien tietosisällön oikeellisuutta huomattavasti.

Lisäksi kaikilta tavoitetilan kyselyyn osallistuneilta kysyttiin tietojen määrän muutosta raporteilla. Yhtä poikkeusta lukuun ottamatta (10/1) vastaajat kokivat tietojen määrän lisääntyneen raporteilla. Tähän vaikuttaa varmasti myös uuden eläkekäsittelijäkäsitteen lisääminen tietovarastoon, joka mahdollistui Data Vault -menetelmän käyttöönoton yhteydessä.

Seuraavat Data Vault -menetelmän vaikutuksia kartoittavat kysymykset kohdistettiin henkilöille, joiden rooli oli ylläpitäjä (N = 5). Näistä ensimmäiset kysymykset keskittyivät uudistetun raportoinnin myötä käyttöön otettuun eläkepalvelujen käsitelmalliin. Vastaajilta kysyttiin ovatko he hyödyntäneet käsitelmallia raporttimäärittelyä tehdessään ja sitä kertoi hyödyntäneensä jo neljä viidestä vastaajasta.

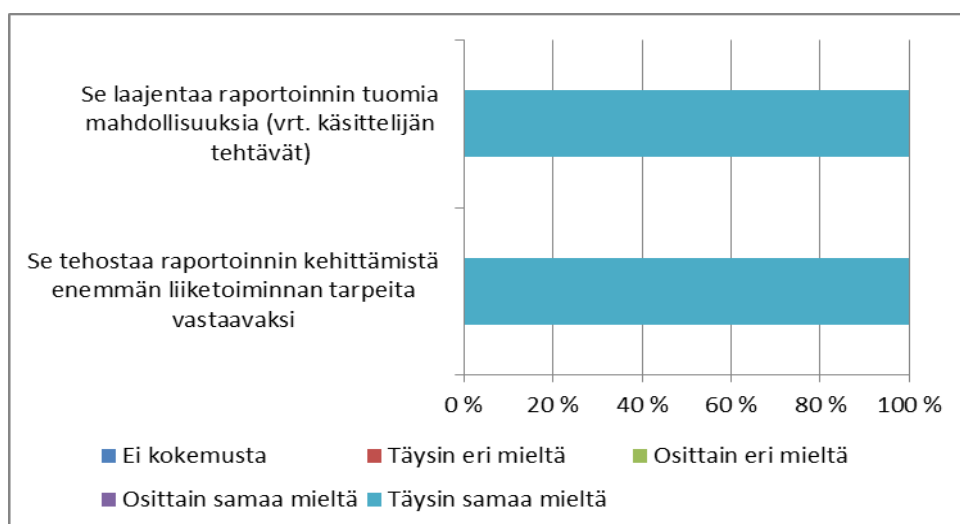
Data Vault -menetelmällä mallinnettu tietovarasto perustuu liiketoiminnan avainkäsitteisiin. Tämä on nähtävissä myös seuraavien kysymysten vastauksissa. Kun vastaajilta kysyttiin, vastaavatko uudistetun tietovaraston tiedot paremmin liiketoiminnan käsitteitä ja auttaako käsitelmä ymmärtämään paremmin tietokannassa olevia tietoja, niin keskimäärin 95 % vastaajista oli siitä osittain tai täysin samaa mieltä. Alla olevassa kuviossa 25 on havainnollistettu vastausten painottumista eri roolien kesken.



Kuvio 25. Vastaajien roolien suhde osittain tai täysin samaa mieltä olleiden vastaajien kesken (N = 5) siitä, että perustuuko tietovarasto liiketoiminnan avainkäsitteisiin

Yksi Data Vault -menetelmän tuomista mahdollisuuksista on luoda myös sellaisia käsitteitä, joita ei lähdejärjestelmien datassa vielä ole. Liiketoimintakerroksessa raakadatakerroksen tiedot voidaan yhdistellä uusilla tavoilla ja näitä yhdistelysääntöjä voidaan tarvittaessa joustavasti myös muuttaa. Luvussa 4.2. (Uudistettu Data Vault -tietovarasto-arkkitehtuuri) asiaa kuvataan tarkemmin.

Uusien käsitteiden luomisen mahdollisuus koettiin erittäin hyväksi raportoinnin kehittämisen näkökulmasta, kaikki vastaajat olivat sitä mieltä, että he tulevat jatkossa hyödyntämään tätä ominaisuutta. Lisäksi vastaajilta kysyttiin, laajentaako tämä ominaisuus raportoinnin tuomia mahdollisuuksia ja tehostaako se raportoinnin kehittämistä enemmän liiketoiminnan tarpeita vastaavaksi. Kuviosta 26 ilmenee, että kaikki ylläpitoon osallistuvat vastaajat olivat myös näiden väittämisen osalta täysin samaa mieltä asiasta.



Kuvio 26. Uusien käsitteiden luontiominaisuuden vaikutukset raportoinnin kehittämiseen (N = 5)

6 Tutkimuksen eteneminen

Tutkimuksen tarkoituksena oli vastata kysymykseen "Parantaako Data Vault -menetelmä tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta?". Tutkimus toteutettiin laadullisena tapaustutkimuksena ja siinä käytettiin konstruktivistista tutkimusotetta, joka kohdistui yhteen eläkealan yritykseen, Yritys A. Tutkimuksen aikana toteutettiin konstruktio ja verrattiin sen tuomia eroja aiempaan tapaan toimia tietovaraston toteutuksessa ja ylläpidossa.

Tutkimuksen aluksi koottiin teoreettinen viitekehys tutkimukselle kirjallisuuskatsauksen ja dokumenttianalyysin avulla. Tutkimusaiheesta ei vielä ollut saatavilla paljon kirjallista materiaalia. Lisäksi toinen tärkeimmistä menetelmään liittyvistä lähteistä julkaistiin vasta tutkimuksen aikana. Tutkimuksen yhteyteen kootut lähteet antavat kuitenkin kattavan kuvan Data Vault -menetelmästä ja sen sovellettavuudesta kohdeyritykseen, joka kertoo siitä, että lähteiden relevanssia on jouduttu pohtimaan tutkimuskysymysten kannalta. Kirjallisuuskatsauksen yhteydessä kartoitettiin myös, löytyykö tutkimusaiheesta aikaisempaa tutkimusmateriaalia, mutta sitä ei ollut saatavilla. Dokumenttianalyysissä hyödynnettiin tutkimuksen kontekstiin liittyvää kirjalliseen muotoon saatettua materiaalia kohdeyrityksestä, sen tietovarastoinnin nykytilasta sekä tutkimuksen kohteena olevasta rajatun liiketoiminta-alueen raportointitarpeista ja siihen liittyvästä kehittämishankkeesta.

Teoreettisen viitekehysten tarkentamisen jälkeen siirryttiin empiiriseen eli kokemukselliseen osaan tutkimusta. Empiirinen osa toteutettiin seuraamalla kohdeyrityksessä käynnissä olevaa kehittämishanketta, jossa toteutettiin rajatulle liiketoiminta-alueelle Data Vault -menetelmällä uusi tietovarasto raportointiin. Kehittämishanke oli osa laajempaa kehittämiskokonaisuutta, jossa mahdollistettiin eläketapausten käsittelyn automatisointi. Raportoinnin kehittämishanke jakaantui useampaan projektiin, joiden tuloksena oli kohdeyrityksen eläkepalvelujen raportoinnin laajentaminen tietovarastoksi ja siihen liittyvien raporttien toteuttaminen. Näihin raportteihin sisältyneet tietovarastomuutokset toimivat tämän tutkimuksen empiirisen osuuden vertailumateriaalina ja niiden avulla pystyttiin tutkimaan teorian soveltuvuutta kohdeyrityksen tietovarastomallinnukseen.

Tutkimuksen yhteydessä myös havainnoitiin Data Vault -menetelmän käyttöönoton vaiheita ja käytännön toteutusta kohdetyrityksessä, seuraamalla uuden tietovaraston toteutusta kehittämishankkeessa sekä soveltamalla tutkimuksen yhteydessä esille tullutta teoriaa käytäntöön. Havainnoinnin aikana tutkija seurasi toteutusta ja perehtyi eläkepalvelujen raportointiin tutkimuksen kohteena olevien raporttien osalta. Tutkimuksessa kehittämishankkeen asiantuntijoina olivat tietovarastoratkaisun toteuttavat konsultit, raportointia tuntevat ja sitä hyödyntävät liiketoiminta-asiantuntijat sekä raportointialueen vastaava tietohallinnosta. Tutkimuksen edetessä yhteistyö eri osapuolien kanssa toimi erittäin hyvin.

Lisäksi tutkimuksen aikana kerättiin käyttökokemuksia kehittämishankkeen osallistujilta ja loppukäyttäjiltä tietovarastomallinnuksen nykytilasta ja odotetusta tavoitetilasta sekä siitä, tuottaako uuden menetelmän käyttöönotto lisäarvoa kohdeyritykselle. Näitä käyttökokemuksia raportoinnin kehittämistä ja muutostarpeiden läpiviennistä selvitettiin strukturoitujen teemahaastattelujen avulla, jotka toteutettiin kahtena eri verkkokyselynä tutkimuksen alussa ja lopussa. Nykytilan haastattelujen toteuttaminen ja analysointi tutkimuksen alussa toteutui suunnitellusti, mutta tavoitetilan haastattelun aikataulu venyi kehittämishankkeen viivästymisestä johtuen, josta tarkemmin edempänä. Viivästymisestä johtuen tavoitetilan haastattelu päätettiin siirtää tutkimuksen loppupuolelle, jotta kehittämishankkeessa saataisiin mahdollisimman monta raporttia käyttöön kattavan vertailumateriaalin varmistamiseksi. Sen seurauksena tavoitetilan haastattelu perustui noin kolmasosaan alkuperäisestä raporttimäärästä sekä niihin liittyviin käyttökokemuksiin. Vertailumateriaaliin sisältyneiden raporttien määrä oli kuitenkin riittävä, sillä haastattelujen analyysien tuloksista pystyttiin tekemään johtopäätöksiä eläkepalvelujen tietovarastopohjaisen raportoinnin kehittämistä ja muutostarpeiden läpiviennistä.

Tutkimuksen edistyminen riippui uuden eläkekäsittelyjärjestelmän käyttöönottohankkeen ja raportoin kehittämishankkeen aikatauluista. Käyttöönottohankkeen aikataulun viivästymisestä ja lopulta sen keskeytyksestä johtuen raportoinnin kehittämishankkeen näkökulma muuttui vuoden 2011 lopulla. Sen seurauksena raporttien lähdejärjestelmänä oli edelleen nykyinen eläkekäsittelyjärjestelmä (K). Näin ollen kehittämishankkeeseen tuli muutoksia raportointitarpeisiin, tietovaraston rakentamiseen, aikatauluun ja

suunniteltuihin lopputuloksiin. Tämän lisäksi raportoinnin kehittämishanke viivästyi vielä aikataulustaan, jonka seurauksena raporttien hyödyntämiseen ja ylläpitoon liittyvää tutkimuksen vertailumateriaalia ei ollut saatavilla suunnitellussa aikataulussa.

Yhtenä osana tutkimusta tutkija osallistui Data Vault -menetelmäkoulutukseen. Koulutuksen tarjosi Ari Hovi Oy ja kouluttajana toimii konsultti Ari Hovi. Hän on ensimmäisiä henkilöitä Suomessa, jolla on Linstedtin sertifiointi, "Certified Data Vault Modeler" (kesä 2008) ja hän tekee yhteistyötä menetelmän kehittäjän Daniel Linstedtin kanssa. Koulutuksessa käsiteltiin yleisesti tietovarastoarkkitehtuureja ja mallinnusta sekä Data Vault -menetelmän taustaa ja rakennuskomponentteja. Lisäksi siellä otettiin kantaa tietojen käsittelyyn ja laatuun sekä niiden jäljitettävyyteen ja historiointiin. Kurssilla arvioitiin myös Data Vault -menetelmän soveltuvuusalueita sekä sen niveltämistä muuhun mallinnukseen. (Ari Hovi Oy.)

6.1 Tutkimuksen tekninen ympäristö

Tutkimuksen konstruktio toteutettiin kohdeyrityksen tietovaraston kehitys- ja tuotantoympäristössä. Data Vault -menetelmällä toteutetun tietomallin kuvaus tehtiin ER Studio -mallintamisvälineellä ja suunnittelun apuna käytettiin Microsoft Excel-ohjelmistoa. Tietovaraston tietokanta sijaitsee Microsoft SQL Server 2008 -tietokantapalvelimella. Tietovaraston latausvälineenä käytetään Informatica-ohjelmistoa, joka on ETL-väline. Tietovaraston raportointivälineenä käytetään BusinessObjects XI -ohjelmistoa ja raporttien julkaisuun Infoview-ohjelmistoa.

Tutkimuksen yhteydessä suunniteltiin Data Vault -menetelmään perustuva looginen ja fyysinen tietomalli tutkimuksen kohteena olevan liiketoiminta-alueen käsitteistä. Fyysisen tietomallin mukaiset taulurakenteet luotiin ensin tietovaraston kehityskantaan, johon testauksessa tarvittavat tiedot ladattiin operatiivisista järjestelmistä Informatica-ohjelmistolla. Tämän jälkeen tietovarastoon vietyjä tietoja ja niiden käytettävyyttä testattiin raportoinnin avulla. Kun konstruktion toimivuus oli testattu kehitysympäristössä, siirrettiin se tuotantoympäristöön.

6.2 Tutkimuksen vaiheet

Varsinainen tutkimus toteutettiin projektina, jossa tutkija toimi sekä projektinvetäjänä että sen jäsenenä. Tutkimusprojekti toteutettiin raportoinnin kehittämishankkeen yhteydessä. Tutkimuksen aikana tutkija seurasi aktiivisesti myös kehittämishankkeen etenemistä ja siinä saatuja tuloksia. Koska tutkimus tehtiin yritykseen, jossa tutkija työskentelee, niin erillistä lupaa tutkimuksen tekemiseen ei tarvittu. Kohdeyritys oli sitoutunut tutkimuksen lopputulokseen, sillä kyseessä oli tietovarasto ympäristön toiminnan parantaminen, joka tuki yrityksen tavoitteita EDW:n kehittämiseksi. Tutkimuksen lopuksi kukin osa raportoitui siten, että ymmärrettiin, mitä oli tehty ja mitä tulokset olivat olleet. Nämä tulokset on koottu tähän tutkimusraporttiin. Tutkimusprojektin dokumentit on erotettu omaksi kokonaisuudekseen tästä tutkimusraportista. Osa tämän tutkimuksen liitteistä sisältää luottamuksellista tietoa ja ovat kohdeorganisaation pyynnöstä luottamuksellisia.

Tutkimuksen etenemistä seurattiin Haaga-Helian ohjaajan, kohdeyrityksen sisältöohjaajan ja tutkimuksen tekijän yhteisissä tutkimusprojektin johtoryhmän kokouksissa, joita pidettiin kahdeksan kertaa. Johtoryhmätyöskentely tuki hyvin tutkimusprojektin etenemistä ja ohjaajien asiantuntemusta hyödynnettiin tutkimuksen edetessä erilaisissa tilanteissa. Lisäksi tutkimuksen etenemistä seurattiin Haaga-Helian kehittämistehtävään liittyvissä työpajoissa, joihin tutkija osallistui aktiivisesti.

Lisäksi tutkimuksen ratkaisun soveltamista kohdeyritykseen selvittiin siten, että projektiasiantuntijoiden kanssa käytiin yhdessä läpi oppimisprosessia, analysoiden tuloksia ja ennakkoehtoja. Lopuksi tutkija pyrki vaihtamaan näkökulman enemmän analysoijan rooliin ja reflektoi havaintojaan aiempaan teoriaan. Näin saatiin helpommin selville projektin tuoma teoreettinen panos tutkimukselle.

Tutkimus toteutettiin projektina vuosien 2010 - 2012 aikana. Siihen liittyvät vaiheet on kuvattu taulukossa 10. Taulukosta ilmenee myös vaiheiden aikataulutus sekä suunnitellut että toteutuneet työmääräarviot henkilötyötunteina.

Taulukko 10. Tutkimusprojektin vaiheet

Vaihe	Aloitus- pvm	Lopetus- pvm	Suunn. työ- määrä	Tot. työ- määrä
Projektinhallinta	1.11.2010	30.11.2012	70 h	70 h
Syvällisen teoreettisen ja käytännöllisen tiedon hankinta Data Vault -menetelmästä	1.11.2010	15.6.2011	168 h	118 h
Nykytilanteen selvitys tietovarastomallinnuksen muutosjoustavuudesta ja liiketoiminnan tarpeista sekä kerättyjen tietojen analysointi	1.11.2010	6.5.2011	134,5 h	79 h
Data Vault -toteutuksen suunnittelun seuranta	24.1.2011	30.11.2011	52,5 h	36 h
Data Vault toteutuksen seuranta	1.2.2011	20.1.2012	37,5 h	35,50 h
Käyttökokemusten kerääminen konstruktion rakentamisesta ja kerättyjen tietojen analysointi	1.2.2012	31.10.2012	124,5 h	121,5 h
Tutkimusraportin laadinta ja johtopäätösten tekeminen teorian soveltamisesta sekä Data Vault -menetelmän käyttöönotosta	22.8.2011	9.11.2012	210 h	206 h
Tulostusaineiston viimeistely	12.11.2012	30.11.2012	11 h	11 h
Koko projekti			808 h	677 h

Tutkijan mielestä tutkimukselle asetetut rajaukset pitivät tutkimuksen fokuksen selkeänä ja asiasisällön sopivana. Myös tutkimuksen tekeminen projektina auttoi tutkimukselle asetettujen tavoitteiden saavuttamista toisin sanoen tutkimuskysymyksiin vastaamista ja helpotti tutkimuksen huolellista suunnittelua ja muutoksenhallintaa. Sen avulla pystyttiin hallitsemaan myös empiriaosuuteen liittyneen kehittämishankkeen aikaansaamat aikataulumuutokset ja kohdentamaan toimintaa uudelleen. Tavoitteiden asetanta oli kohdallaan ja niitä tarkennettiin tutkimuksen edetessä.

7 Tutkimuksen yhteenveto

Tutkimuksen lähtökohtana oli selvittää kohdeyrityksessä Data Vault -menetelmän tuomia eroja tietovarastopohjaisen raportoinnin ylläpidossa ja muutosjoustavuudessa. Tutkimus toteutettiin työeläkealan yrityksessä tapaustutkimuksena konstruktivistista tutkimusotetta hyödyntäen. Tutkimus sisälsi sekä teoreettisen että empiirisen osuuden. Kirjallisuuskatsauksen ja dokumenttianalyysin avulla haettiin teoretietoa Data Vault -menetelmästä ja sen käytettävyydestä tietovaraston mallinnuksessa kohdeyrityksen kaltaisessa yrityksessä. Niiden avulla selvitettiin, miten Data Vault -menetelmä eroaa aiemmin käytössä olleesta mallinnusmenetelmästä, tähtimallista ja etenkin siitä, miten sitä on hyödynnetty kohdeyrityksessä.

Empiirisessä osassa tutkimusta rakennettiin konstruktio kohdeyrityksen tietovarastoympäristöön rajatulle raportointialueelle. Sen avulla pystyttiin havainnollistamaan menetelmän toimivuutta käytännössä. Koska rakennettu konstruktio toteutettiin tuotantoympäristöön ja otettiin käyttöön tutkimuksen aikana, mahdollisti se aidon käyttäjäkokemuksen keräämisen teemahaastattelujen avulla. Haastattelujen avulla kerättiin ensisijaisesti kokemusperäistä tietoa eroista aiemman ratkaisun sekä uuden konstruktion välillä.

Tutkimuksen aikana toimittiin läheisessä yhteistyössä tietovarastoa hyödyntävien liiketoiminta-asiantuntijoiden kanssa. Näin saatiin käytännön edustajat mukaan kehittämään tutkimusaiheesta innovatiivista konstruktioita, joka soveltui kohdeyrityksen ympäristöön. Kehitetyn konstruktion toteuttamiskelpoisuutta havainnoitiin seuraamalla sen suunnittelua ja toteutusta sekä testaamalla sen toimivuutta tietovarastoasiantuntijoiden kanssa. Tämä antoi myös mahdollisuuden tutkijalle markkinoida uutta tietovaraston mallintamismenetelmää kohdeyrityksessä ja sen tuomia mahdollisia hyötyjä.

Käytetyt tutkimusmenetelmät soveltuivat hyvin tämän tyyppisen tutkimuksen tekemiseen. Tutkijan näkökulmasta konstruktivistinen lähestymistapa lisäsi kokemuksellista perehtymistä asiaan. Kirjallisuuskatsauksen ja dokumenttianalyysin avulla saatiin kerättyä riittävästi teoretietoa sekä tietoa rakennettavan konstruktion kohteena olevasta raportointialueesta. Ainoana rajoituksena menetelmien osalta oli tavoitetilan teemahaastatte-

lujen toteuttamisen aikataulun viivästyminen ja siitä johtuvan vertailumateriaalin vähäisempi saatavuus. Jatkossa kannattaa kiinnittää huomiota näin lähellä käytännön toteutusta olevaa tutkimusta suunnitellessa etenkin aikataulun laadintaan, jotta lopputulokseen tarvittavat edellytykset olisivat saatavilla. Tämän tutkimuksen osalta vertailumateriaalin vähyys ei kuitenkaan vaikuttanut merkittävästi haastatteluanalyysojen tulokseen, sillä vastausten perusteella tehdyissä johtopäätöksissä on nähtävissä selviä vastauksia Data Vault -menetelmän käytettävyydestä ja soveltuvuudesta kohdeyrityksen tarpeisiin.

Tutkimuksen aihevalinnasta näkee, että tutkija itse on hyvin kiinnostunut liiketoimintatiedon hyödyntämisestä yrityksessä sekä tietovaraston kyvystä vastata alati muuttuvan tiedon ylläpitoon. Lisäksi esille tulee selkeästi myös kohdeyrityksen tietovaraston nykytila ja motivaatio kehittämiselle. Kohdeyritys oli sitoutunut tutkimuksen lopputulokseen, sillä kyseessä oli kehittämissanke tietovarasto-ympäristön toiminnan parantamiseksi, joka tuki yrityksen tavoitteita EDW:n kehittämiseksi.

Tutkimusongelman kysymyksiin saadaan vastaukset tässä tutkimusraportissa seuraavasti:

- Pääkysymys: Parantaako Data Vault -menetelmä tietovarastoratkaisun skaalautuvuutta ja muutosjoustavuutta? Tapaustutkimus, Yritys A.
Luvuissa 2.4.5 (Data Vault -tietokannan skaalautuvuus) ja 2.4.8 (Business Data Vault) tuodaan esille teoriaa Vault -menetelmä tietovarastoratkaisun skaalautuvuudesta ja muutosjoustavuudesta. Lisäksi luvussa 5 (Empiria tutkimuksen kohteena olevasta raportointialueesta) on haastattelujen avulla selville saatuja käytännön kokemuksia Data Vault -menetelmän tuomista parannuksista tietovarastoon.
- Alikysymys: Miten Data Vault -menetelmä eroaa nykyisin käytössä olevasta tietovarastojen mallinnusmenetelmästä?
Luvuissa 2.3 (Aiempien mallinnusmenetelmien ongelmat EDW:ssä) ja 2.4 (Data Vault -menetelmä) ja sen aliluvuissa kuvataan teoreettisella tasolla Data Vault -menetelmän eroa nykyisin kohdeyrityksessä käytössä olevasta tähtimallista. Lisäksi eroja tietomallitasolla kuvataan luvuissa 4.3 (Nykyinen tietomalli) ja 4.4 (Uudistettu Data Vault -tietomalli).

- Alikysymys: Mitä muutoksia uusi Data Vault -menetelmä tuo tietovarastoarkkitehtuuriin ja -raportointiin nykyiseen verrattuna?
Empiriaosuuden luvuissa 4.1 (Nykyinen tietovarastoarkkitehtuuri) ja 4.2 (Uudistettu Data Vault -tietovarastoarkkitehtuuri) kuvataan Data Vault -menetelmän tuomia muutoksia tietovarastoarkkitehtuuriin. Raportointiin tulleet muutokset, selviävät luvusta 4.7 (Esimerkkejä Data Vault -menetelmän vaikutuksesta raportointiin) ja sen aliluvuista, joissa perehdytään Data Vault -menetelmän toteutukseen käytännössä.

7.1 Yhteenveto tutkimustuloksista ja niiden kontribuutio käytäntöön

Tutkimuksen tavoitteena oli saada käytännön ongelmaan teoreettisesti perusteltu ja uudenlainen ratkaisu, joka tuo kohdeyrityksen liiketoiminnalle ja mahdollisesti myös tiedeyhteisölle uutta tietoa Data Vault -menetelmän hyödyntämisestä käytännössä. Tutkimuksen teoreettisessa osassa koottiin perustietoa Data Vault -menetelmästä sekä perusteluja sen käytettävyydestä tietovarastoinnissa etenkin EDW:n näkökulmasta. Tämä teoreettinen viitekehys auttoi tutkijan tutustumista Data Vault -menetelmään sekä sen tuomiin mahdollisuuksiin ja eroihin tähtimalliin verrattuna.

Tutkimuksen aikana rakennettiin kohdeyrityksessä Data Vault -menetelmää hyödyntävä tietovarasto eläkepalvelujen johdonraportoinnin käyttöön. Aiempi ratkaisu perustui tähtimallinnukseen. Periaatteena uuden tietovaraston rakentamisessa oli, että siinä hyödynnetään puhdasta Data Vault -menetelmään perustuvaa mallinnustapaa, sillä siten tietovarasto on helpommin ylläpidettävissä myös jatkossa. Muutos ei koskenut Data Martteihin perustuvaa raportointikerrosta, vaan tietovarastoon tallennettuja tietoja, jota Data Martit hyödyntävät.

Tutkimuksen tuloksena on tämä raportti, joka sisältää teoriatietoa Data Vault -menetelmästä sekä sen eroista tähtimalliin, jota on aiemmin käytetty laajalti tietovarastojen mallintamisessa. Tutkimuksen yhteydessä koottiin myös teoriaa tietovaraston kehittämisestä enemmän keskitettyyn malliin. Näitä tutkimustuloksia voidaan hyödyntää sekä kohdeyrityksessä että sen ulkopuolella ja sen seurauksena, että teoriaosuudesta kirjoitettiin kattava raportti, on tästä tutkimuksesta hyötyä myös alan tiedeyhteisölle.

Teorian pohjalta vahvistuu Data Vault -menetelmän tuomat edut tietovarastoinnissa, jotka lisäävät etenkin sen muutosjoustavuutta ja skaalautuvuutta. Lisäksi teoriassakin esille nostettu tietojen historiointi osoittautui tutkimuksen empiirisessä osassa juuri niin hyväksi ominaisuudeksi, kun asiantuntijat sen mainitsevat. Kohdeyrityksessä se laajensi raportoinnin mahdollisuuksia ja samalla paransi raporttien tietosisällön laatua ja luotettavuutta. Näin ollen tutkimuksen lopputulokset tukevat sitä käsitystä, että Data Vault -menetelmä tukee hyvin tietojen historiointia.

Data Vault -menetelmä tukee myös liiketoiminnan ja tietohallinnon yhteisten tavoitteiden saavuttamista. Se mukautuu nopeasti liiketoiminnan muutoksiin ja sen avulla tietovaraston muutoshallinta helpottuu. Tätä puoltaa menetelmän rakenne, joka mahdollistaa helposti ja joustavasti uuden liiketoimintatiedon tai uuden tietolähteen lisäämisen tietovarastoon kuin myös niiden poistamisen. Etenkin kun sitä verrataan tähtimalliin, muutokset tiedoissa ja tietolähteissä eivät vaikuta niin laajalle alueelle kuin tähtimallissa. Muutosjoustavuutta lisää, että raakadata pysyy muutosten aikana tallessa, joten se on helppo yhdistää uuteen liiketoimintakerrokseen luotuun tietorakenteeseen. Tämä johtuu siitä, että menetelmässä osa muutoksen kohteista on siirretty menetelmän sisältöön menetelmän rakenteen sijaan, kuten Martin jo aiemmin toteaa.

Yksi huomattava etu Data Vault -menetelmässä on uuden tiedon luomisen mahdollisuus käytettävissä olevasta datasta. Raakadakerroksen päällä olevassa liiketoimintakerroksessa liiketoiminnan tarvitsemat tiedot on yhdisteltävissä tavalla, joka mahdollistaa myös sellaisten käsitteiden luonnin, joita ei lähdejärjestelmien datassa vielä ole. Tällöin tietovarastosta tuotettava tieto ei välttämättä ole sidottu operatiivisiin tietolähteisiin ja niissä jo olemassa oleviin tietoihin ja käsitteisiin, sillä menetelmä mahdollistaa tarvittaessa myös olemassa olevasta datasta uusien käsitteiden luonnin, jotka tukevat paremmin liiketoiminnan tarpeita ja päätöksentekoa.

Lisäksi Data Vault -menetelmä tehostaa Data Martti -raportointikerroksen suorituskykyä, kun osa aiemmin siellä tehdystä laskennasta siirretään liiketoimintakerrokseen. Tämä edellyttää huolellista Data Marttien suunnittelua, sillä vastuu laaduntarkastuksesta siirtyy raportointikerrokseen. Tämä hyöty ei kaikilta osin toteutunut eläkepalvelujen raportoinnin kehittämishankkeessa, sillä laajat kokoomaraportit sisältävät tosiinsa kyt-

keytyneiden raporttitarpeiden yhdistelyä ja se vaikuttaa myös raportointivälineeseen luotaviin uneiverseihin. Toisin sanoen jatkossa, kun tulee muutos, niin siitä seuraa muutostarpeita sekä liiketoimintakerrokseen että tähtimallilla toteutettuun raportointikerrokseen.

Haasteena Data Vault -menetelmän käytössä on, että tiedot ovat raakadatasolla, jolloin ne eivät ole sellaisenaan liiketoiminnan käytettävissä. Lisäksi kannan rakentaminen on suoritaisempää kuin perinteisellä tähtimallilla toteutettu, sillä tietojen rakeisuus on tarkemmalla tasolla. Siitä huolimatta siihen sijoitettu panos tuottaa itsensä takaisin ylläpidon yhteydessä, sillä muutokset tietokantaan vaativat huomattavasti vähemmän työtä kuin tähtimallinnettuun kantaan, jossa jo olemassa oleviin rakenteisiin joudutaan tekemään muutoksia.

Lisäksi haastetta tuo raakadakerros, johon tiedot ladataan operatiivisista järjestelmistä. Raakadakerros ei huomioi, miten tietoja käytetään tai miten niitä analysoidaan eikä siinä oteta kantaa siihen, mitkä tiedot ovat oikein tai mitkä väärin. Tietovarastoon ladattua raakadataa ei siis ole laaturkastettu eikä puhdistettu eikä se sovellu loppukäyttäjän kyselyjen kohteeksi. Vaikka tiedot olisivat virheellisiä, niitä ei muuteta, jotta tietojen jäljitettävyyttä säilyisi. Näin ollen tietojen laatua on ylläpidettävä operatiivisissa järjestelmissä.

Tutkimuksen aikana käynnissä olleen raportoinnin kehittämishankkeen myötä saatiin tietoa myös Data Vault -menetelmää hyödyntävän tietovaraston rakentamisen kustannuksista. Näitä kokemuksia pystytään hyödyntämään, kun kohdeyrityksessä Data Vault -menetelmällä toteutettuun tietovarastoon tullaan tekemään vuosien 2013 ja 2014 aikana muutoksia, jotka liittyvät uuteen käyttöön otettavaan eläkekäsittelyjärjestelmään. Näiden muutosten myötä tietovarastoon on luotava uusia käsitteitä. Tämä kehitystyö tuo aikanaan esille Data Vault -menetelmän muutosjoustavuuden, sillä käsitteiden lisääminen ei aiheuta vanhan koodin uudelleensuunnittelua tai käytössä olevien latausprosessien muutoksia, vaan uutta koodia tarvitaan ainoastaan, kun luodaan uudet käsitteet ja liitetään ne jo olemassa oleviin. Uudesta tietovarastosta saatavat kustannussäästöt perustuvat siis ensisijaisesti siihen, että olemassa oleviin rakenteisiin ei tarvitse tehdä

muutoksia, vaan ainoastaan uuden luominen ja linkittäminen vanhaan vaatii kustannuksia.

Kuitenkin Data Vault -menetelmän varsinaiset kustannussäästöt tulevat esiin vasta vaikiintuneen ylläpidon myötä ja uusien päivitys- ja laajennustarpeiden ilmetessä. Menetelmä vaatii alussa enemmän työtä, sillä latauksia tulee suhteessa enemmän kuin tähtimallissa tietojen rakeisuudesta johtuen. Tämä lisää kuitenkin ylläpidettävyyttä, sillä uusia osia on helpompi lisätä, kun toteutusta voidaan tehdä pienemmissä osissa. Toisin sanoen, päinvastoin kuin tähtimallissa, jossa muutokset vaikuttavat aina olemassa oleviin tietorakenteisiin ja dataan ja joka näin ollen saattaa vaatia miljoonien datarivien päivittämistä. Tulevaisuudessa uusien raportointialueiden lisääminen uuteen tietovarastoon ei vaadi yhtä isoa investointia, sillä tietojen yhdistäminen olemassa olevaan on jo paljon joustavampaa ja nopeampaa. Tämä tuli esille myös haastatteluanalysien tuloksissa, sillä kummatkin arvot olivat parantuneet uudistetun raportoinnin käyttöönoton myötä. Sen seurauksena voidaan päätellä, että tietovarastosta tehtävän raportoinnin muutosjoustavuus on lisääntynyt uuden Data Vault -menetelmän avulla.

Uuden tietovaraston myötä ei myöskään tekniseen ympäristöön eikä laitteistoihin tarvinnut tehdä investointeja, vaan kustannuksia kertyi lähinnä Data Vault -menetelmän mukaisen tietovaraston toteutuksesta sekä raportteihin tehtävistä määrittelyistä, toteutuksesta ja testauksesta. Huomion arvoista on, että vaikka Data Vault -menetelmän käyttöönoton myötä latausten ja tietokantataulujen määrät kasvoivat, niin se ei vaikuttanut kohdeyrityksen tietovaraston suorituskykyyn. Tämä johtui siitä, että Data Marttin määrä säilyi suunnilleen ennallaan ja siitä, että latausajojen ajoajat laskivat vähintään kolmannekseen aiemmasta, koska osa latauksista pystytään jatkossa ajamaan rinnakkain.

7.2 Tutkimuksen rajoitukset ja yleistettävyys

Tutkimuksen aihe on ajankohtainen ja liittyy läheisesti työelämään, sillä nykyään tieto on yksi yritysten tärkeimmistä resursseista. Ajankohtaisuutta lisää myös näkökulma keskitettyihin tietovarastoihin. Tulevaisuudessa tiedon määrä tulee vain kasvamaan ja siksi tiedon hallinta ja laatu tulevat merkitsemään enemmän. Organisaation tietopää-

oman tärkeys ja muutosjoustavuus korostuvat etenkin silloin, kun tietoa hyödynnetään päätöksenteon tukena. Tutkimuksen kannalta tärkeitä olivat etenkin Data Vault -rakenteen perustaulut hubi-, satelliitti- ja linkkitaulut, jotka muodostavat perustavaa laatua olevan muutoksen verrattuna aiempiin tietovarastojen mallinnusmenetelmiin.

Tutkimuksen rajoituksena oli, että se suoritetaan vain yhdessä työeläkealan yrityksessä Suomen markkinoilla ja tutkimus keskittyi Data Vault -menetelmän hyödyntämiseen kohdeyrityksen tietovarastoinnissa rajatulla liiketoiminta-alueella. Menetelmä on niin uusi, että vertailudataa muista organisaatioista ei vielä ollut saatavilla. Vaikka tutkimuksen kohteena olikin työeläkesektorin yritys, niin teoreettisessa osassa esille tuodut Data Vault -menetelmään liittyvät ominaispiirteet ja käyttömahdollisuudet ovat päteviä millä tahansa toimialalla. Empiirisen osan käytännön toteutuksen ja haastattelujen tulokset sisältävät näin ollen myös työeläkesektorin erityispiirteitä, mutta tietovarastoon ja tietokannan rakenteisiin sekä ylläpitoon liittyvät käytänteet ovat sovellettavissa myös muissa yrityksissä.

7.3 Jatkoimenpide-ehdotukset

Tutkimuksen aikana saatujen lopputulosten, kokemusten ja kohdeyrityksen sisäisten keskustelujen pohjalta voidaan esittää seuraavia toimenpide-ehdotuksia:

– **Tavoitetilan tutkimushaastattelun uusiminen**

Data Vault -menetelmän hyödyntäminen eläkepalvelujen johdonraportoinnissa oli ensimmäinen tämän tyyppinen kehittämishanke kohdeyrityksen tietovarastoinnissa. Tutkimuksen aikataulusta tutkijan opintoihin liittyen ja kehittämishankkeen viivästymisestä johtuen ei tutkimuksen kuluessa saatu vielä kokemuksia loppuun asti vietyä kehittämishankkeesta eikä näin ollen myöskään vakiintuneesta ylläpidosta.

Näin ollen tutkimuksesta saatavia tuloksia pystytään täydentämään myöhemmin, mikäli katsotaan tarpeelliseksi uusia tavoitetilan haastattelututkimus tai sitä soveltava tutkimus jonkin ajan kuluttua.

– **Data Vault -menetelmän hyödyntäminen kohdeyrityksessä**

Tutkija osallistuu aktiivisesti kohdeyrityksen tietovarastoraportoinnin kehittämiseen.

Sen tähden tutkimuksessa asiaa on lähestytty keskitetyn tietovaraston kehittämisen näkökulmasta. Tutkimuksen pohjalta voidaan todeta, että Data Vault -menetelmän käytön laajentaminen on suositeltavaa kohdeyrityksessä, sillä Data Vault -menetelmä mahdollistaa skaalautuvan ja iteratiivisen tietovarastokehityksen. Tutkimuksen lopputulokset tukevat vahvasti sitä, että menetelmän ominaisuudet tulevat parhaiten esille juuri tietovarastoinnissa. Data Vault -menetelmä mahdollistaa muutosjoustavuuden, jota missään aikaisemmassa mallinnusmenetelmässä ei vielä ole sekä skaalautuvuuden etenkin sellaisten käsitteiden luonnissa, joita lähdejärjestelmistä ei vielä löydy. Tämä avaa varsinkin liiketoiminnalle mahdollisuuden uudenlaiseen raportointiin ja analysointiin.

– **Data Vault -menetelmän käytön laajentaminen kohdeyrityksessä**

Seuraava vaihe Data Vault -menetelmän hyödyntämisessä on jo päätetty, sillä uuden operatiivisen järjestelmän kehittämishankkeen jatkovaiheessa on tarkoitus lisätä uuden eläkekäsittelyjärjestelmän tiedot uudistettuun tietovarastoon yhteisraportoinnin mahdollistamiseksi. Uudistettu tietovarasto pystyy hallitsemaan samanaikaisesti sekä aiemman että uuden eläkekäsittelyjärjestelmän tiedot, joita hyödynnetään raporteilla.

Tämän lisäksi kohdeyrityksessä tulisi lisätä keskitetympään tietovaraston kehittämistä. Sen mahdollistaa joko olemassa olevien tai uusien raportointialueiden liittäminen osaksi uutta tietovarastoa. Tämä tulee ajankohtaiseksi muun muassa siinä vaiheessa, kun jo nyt käytössä oleviin raportointialueisiin kohdistuu laajempia muutostarpeita tai lähdetään rakentamaan uutta raportointialuetta. Selvää on, että Data Vault -menetelmän käyttöönotto useammilla raportointialueilla ohjaa tietovaraston käyttöä keskitetympään suuntaan. Siitä seuraa myös, että kaikki yrityksen tarvitsemat tiedot ovat samassa paikassa ja hyödynnettävissä yli organisaatiorajojen.

8 Johtopäätökset

Data Vault -menetelmän uutuudesta johtuen tutkimus oli haasteellinen, mutta samalla myös antoisa. Tutkimuksen aihetta lähestyttiin kirjallisten aineistojen analysoinnilla ja perehtymällä nykyiseen ja uuteen tapaan toimia. Varsinkin kattavan teoreettisen tiedon löytyminen vaati työtä, kun toinen menetelmään liittyvistä lähteistä julkaistiin vasta tutkimuksen aikana. Aiheen uutuus toi mukanaan myös sen, että tutkijan asiantuntijuus kasvoi tietovaraston mallinnusmenetelmien ja EDW:n osalta. Etenkin kun Data Vault -menetelmä on kehitetty juuri EDW-mallinnukseen ja perustuu Bill Inmonin DW2.0™-arkkitehtuuriin.

Havainnollistavaa oli tutustuminen Data Vault -menetelmän ja tähtimallin eroavaisuuksiin tietovarastossa ja raporteilla. Varsinkin kun se on tuonut perustavaa laatua olevan muutoksen aiempiin tietovarastojen mallinnusmenetelmiin verrattuna. Tutkimuksen empirian yhteydessä seurattiin tietovaraston toteutusta ja verrattiin sitä teoriaan Data Vault -menetelmästä. Lisäksi käytännön havainnointia tehtiin teemahaastattelujen avulla. Tutkimusprosessi oli siis kokemuksellinen, sillä edellä mainituissa vaiheissa tutkija ja käytännön edustajat tekivät hyvin läheistä yhteistyötä. Etenkin kun tiedonkeruutapoina hyödynnettiin myös kohdeyrityksen sisäisiä materiaaleja sekä avainhenkilöiden avoimia haastatteluja. Muun muassa luvussa 4.7 (Esimerkkejä Data Vault -menetelmän vaikutuksesta raportointiin) olevien tutkimuksen empiriaan liittyvien esimerkkien tekemisessä hyödynnettiin tietovarastototeuttajien asiantuntemusta haastatteleamalla heitä ja perehtymällä jo tehtyihin ratkaisuihin.

Tutkimuksen aikana tutkijan tietämys tietojen mallintamisesta ja tietovarastoinnista lisääntyi. Data Vault -menetelmästä hankitulle tiedolle on kysyntää myös tämän hetken työelämässä, joten tutkimukseen käytetty aika oli tutkijalle palkitsevaa ja hyödyllistä. Lisäksi tutkimuksen rajausta oli toimiva tutkimuksen tuloksia ajatellen, sillä tutkimuksen lopputulokset tukivat hyvin sille asetettuja tavoitteita ja verrattaessa tutkimuksen tuloksia teoriaan on selvää, että esitetyt tulokset ovat valideja. Ainoastaan tutkimuksen kohteena olleen raportoinnin kehityshankkeen aikataulun viivästyttämisestä johtuen tutkimuksen haastattelujen kautta saatu tulos jäi hieman odotettua kapeammaksi ja siksi tutkimuksen jatkotoimenpiteissä on ehdotus haastattelukyselyn uusimisesta jonkin ajan

kuluttua. Vaikka raportoinnin osalta vertailumateriaali jäikin vähäisemmäksi kuin oli suunniteltu, niin haastatteluanalyyysien tuloksesta on kuitenkin nähtävissä selviä vastauksia Data Vault -menetelmän käytettävyydestä ja soveltuvuudesta kohdeyrityksen tarpeisiin.

Lisäksi tutkimuksen yhteydessä tuli hyvin esille Data Vault -menetelmän joustavuus ja miten sen avulla pystyttiin reagoimaan muutokseen. Tutkimuksen aikana käynnissä olleen operatiivisen järjestelmän käyttöönottohankkeen aikataulu ja näkökulma muuttui, joten uusi käsite, käsittelijän eläkekäsittelytehtävä olikin saatava käyttöön raporteille jo nykyisin käytössä olevasta eläkekäsittelyjärjestelmästä. Ongelmana oli, että nykyisessä järjestelmässä ei ollut olemassa edellä mainittua käsitettä. Tämä saatiin ratkaistua liiketoimintakerroksessa, jonne luotiin tietyillä säännöillä nykyisen järjestelmän tiedoista edellä mainittu uusi käsite.

Haastattelujen tuloksena selvisi, että aiempi tähtimalliin perustuva tietovaraston mallinustapa ei ole yhtä muutosjoustava kuin Data Vault -menetelmään perustuva, vaan tietojen lisääminen tietovarastoon on hankalampaa ja kynnyks muutoksen toteuttamiseen nousee. Näihin haasteisiin vastaa Data Vault -menetelmä. Lisäarvo, jonka se tuo kohdeyritykselle, tuli hyvin esille tavoitetilan haastatteluanalyyysin tuloksista. Pääosa vastaajista oli sitä mieltä, uudistetun tietovaraston tiedot vastaavat paremmin liiketoiminnan käsitteitä. Lisäksi vastauksista selviää, että etenkin liiketoimintakerroksen tuomat ominaisuudet, joilla on mahdollista rikastaa perusjärjestelmän käsitteistöä, laajentavat raportoinnin tuomia mahdollisuuksia sekä tehostavat raportoinnin kehittämistä liiketoiminnan tarpeita vastaavaksi. Raportoinnin kehittämisen näkökulmasta vastaajat kokivat uusien käsitteiden luomisen mahdollisuuden erittäin hyväksi ja kaikki kysymykseen vastanneet olivat sitä mieltä, että he tulevat jatkossa hyödyntämään tätä ominaisuutta.

Kohdeyritys oli hyvin sitoutunut tutkimuksen lopputulokseen, koska se tuki yrityksen tietovarastoinnin kehittämistavoitteita. Sen tähden tutkijan mielestä tutkimuksen lopputulosten antama arvo toimeksiantajalle oli merkittävä, sillä niiden avulla pystytään paremmin suunnittelemaan keskitetyn tietovaraston ja raportoinnin kehittämistoimia jatkossa. Kohdeyrityksen aiempi eläkepalvelujen tietovarasto on rakennettu useammasta erillisestä tähtimallilla toteutetusta Data Martista. Tähtimallin etuna on ollut nopea to-

teutus, jolloin paikallisille käyttäjille on saatu toimitettua helposti raportteja. Tämän arkkitehtuurin ongelmana on kuitenkin ollut, että Data Marteista on tullut erillisiä kunkin raportointialueen tarpeiden mukaisesti, joka ei näin ollen tue keskitetymmän tietovaraston kehittämistä.

Tutkimuksen lopputulosten kautta saatiin myös kokemuksiin perustuvaa tietoa siitä, millaisia vaatimuksia ja haasteita Data Vault -menetelmän käyttöönotto ja hyödyntäminen kohdeyritykselle tuo. Lisäksi tutkimuksen yhteydessä selvisi, että Data Vault -menetelmä on hyödynnettävissä sellaisenaan kohdeyrityksen tarpeisiin, sillä uudistetun tietovaraston toteutus noudattaa muun muassa Linstedtin ja Hovin antamia suosituksia menetelmän käytöstä tietovarastoarkkitehtuurissa ja tietojen mallintamisessa, viitaten lukuun 2.4.4 (Data Vault -tietovaraston rakentaminen). Vaikutukset kustannuksiin tuliaan kuitenkin näkemään vasta ajan kuluessa, sillä kuten teoriastakin nousee esille, niin Data Vault -menetelmän käyttöönoton aloituskustannukset ovat selvästi korkeammat kuin vastaavan toteutuksen tekeminen tähtimallilla. Teoriaan nojautuen voidaan kyllä odottaa, että uudistetun tietovaraston myötä kustannukset tulevat vähenemään ylläpidon muutosjoustavuuden ja skaalautuvuuden kautta.

Tutkimuksen lopputulosten perusteella voidaan todeta, että Data Vault -menetelmä tuo merkittäviä etuja tietovaraston muutosjoustavuuteen ja skaalautuvuuteen, kun sitä verrataan tähtimalliin. Kustannusten osalta Data Vault -menetelmän hyödyntämisen aloituskustannukset ovat korkeat, mutta ylläpidon vakiinnuttua kustannushyödyt saadaan paremmin käyttöön. Lisäksi tutkimuksen lopputulokset puoltavat käsitystä, että Data Vault -menetelmä tukee hyvin tietojen historiointia ja parantaa tietovaraston suorituskykyä.

Tulevaisuudessa on mielenkiintoista nähdä, miten kohdeyrityksen liiketoimintakäyttäjien ymmärrys tietovaraston tietosisällöstä kasvaa, kun Data Vault -menetelmän tuomat muutokset tietovarastointiin vakiintuvat käyttöön. Oletettavaa on, että liiketoiminnan asiantuntijoiden kyky hyödyntää tietovarastossa olevaa tietoa ja näin ollen myös yrityksen tietopääomaa paranee, sillä menetelmän tuomien ominaisuuksien myötä heillä on käytettävissä kaikki se tieto, joka lähdejärjestelmistä löytyy sekä mahdollisuus luoda uusia käsitteitä tarpeen mukaan. Kohdeyrityksen raportoinnin ylläpidon muutosjousta-

vuotta lisää myös se, että tulevaisuudessa liiketoiminnan itsenäiskäyttö raportoinnin suhteen kasvaa, kun liiketoiminnan asiantuntijoiden raportointivälineosaaminen ja ymmärrys tietovaraston tietosisällöstä lisääntyy. Tätä edesauttaa etenkin se Data Vault -menetelmän ominaisuus, että tietojen rakenne tietovarastossa on hahmotettavissa liiketoiminta-avaimista käsin.

Lähteet

Ari Hovi Oy. EDW-suunnittelu Data Vault -menetelmällä. Koulutus. Luettavissa:
<http://www.arihovi.com/edw-suunnittelu-data-vault-menetelmalla/>. Luettu
23.11.2011.

GobIT 4.1. 2007. The Control Objectives for Information and related Technology. IT
Governance Institute. USA.

Hovi, A. 2010. Data Vault - menetelmä EDW-kantojen suunnitteluun. Kurssimateriaali
2.12.2010. Ari Hovi Oy. Helsinki.

Hovi, A., Hervonen, H. & Koistinen, H. 2009. Tietovarastot ja Business Intelligence.
WSOYpro/Docendo-tuotteet. Jyväskylä.

Hovi, A., Huotari, J. & Lahdenmäki, T. 2005. Tietokantojen suunnittelu & indeksointi.
2. laitos. 1. painos. Docendo Finland Oy. Jyväskylä.

Hovi, A. Ylinen, J. & Koistinen, H. 2001. Tietovarastot liiketoiminnan tukena. Satku -
Kauppakaari. Helsinki.

Koistinen, H. 2002. Tietojärjestelmien ylläpito. Talentum Media Oy. Helsinki.

Laamanen, K. & Tinnilä, M. 2009. Prosessijohtamisen käsitteet. 4. uudistettu painos.
Teknologiainfo Teknova Oy. Helsinki.

Linstedt, D. Daniel Linstedt Homepage. Luettavissa:
<http://danlinstedt.com/about/data-vault-basics/>. Luettu: 29.3.2010.

Linstedt, D. 2002. Data Vault Series 1. Artikkelisarja. Luettavissa:
<http://www.tdan.com/view-articles/5054/>. Luettu: 29.3.2010.

Linstedt, D. 2011a. Data Vault Modeling is not a Product. Luettavissa:
<http://danlinstedt.com/datavaultcat/data-vault-is-not-a-product/>. Luettu 13.2.2012.

Linstedt, D. 2011b. Defining the Business Data Vault. Luettavissa:
<http://LearnDataVault.com>. Luettu 10.3.2012.

Linstedt, D. 2011c. Super Change Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault. Linstedt, D.

Linstedt, D., Graziano, K. & Hultgren, H. 2009. The New Business Supermodel: The Business of Data Vault modeling. 2. painos. Linstedt, D.

Lukka, K. 2001. Konstruktiivinen tutkimusote. Luettavissa:
http://www.metodix.com/fi/register_helia/. Luettu: 5.3.2010.

Martin, J. 2008. Uusi tietovarastojen mallinnusmenetelmä - Data Vault. Systemityölehden artikkeli. Luettavissa: <http://www.pcuf.fi/sytyke/lehti/kirj/st20082/ST082-23A.pdf>. Luettu: 27.4.2010.

Ojasalo, K., Moilanen, T. & Ritalahti, J. 2009. Kehittämistyön menetelmät: Uudenlaista osaamista liiketoimintaan. WSOYpro OY. Helsinki.

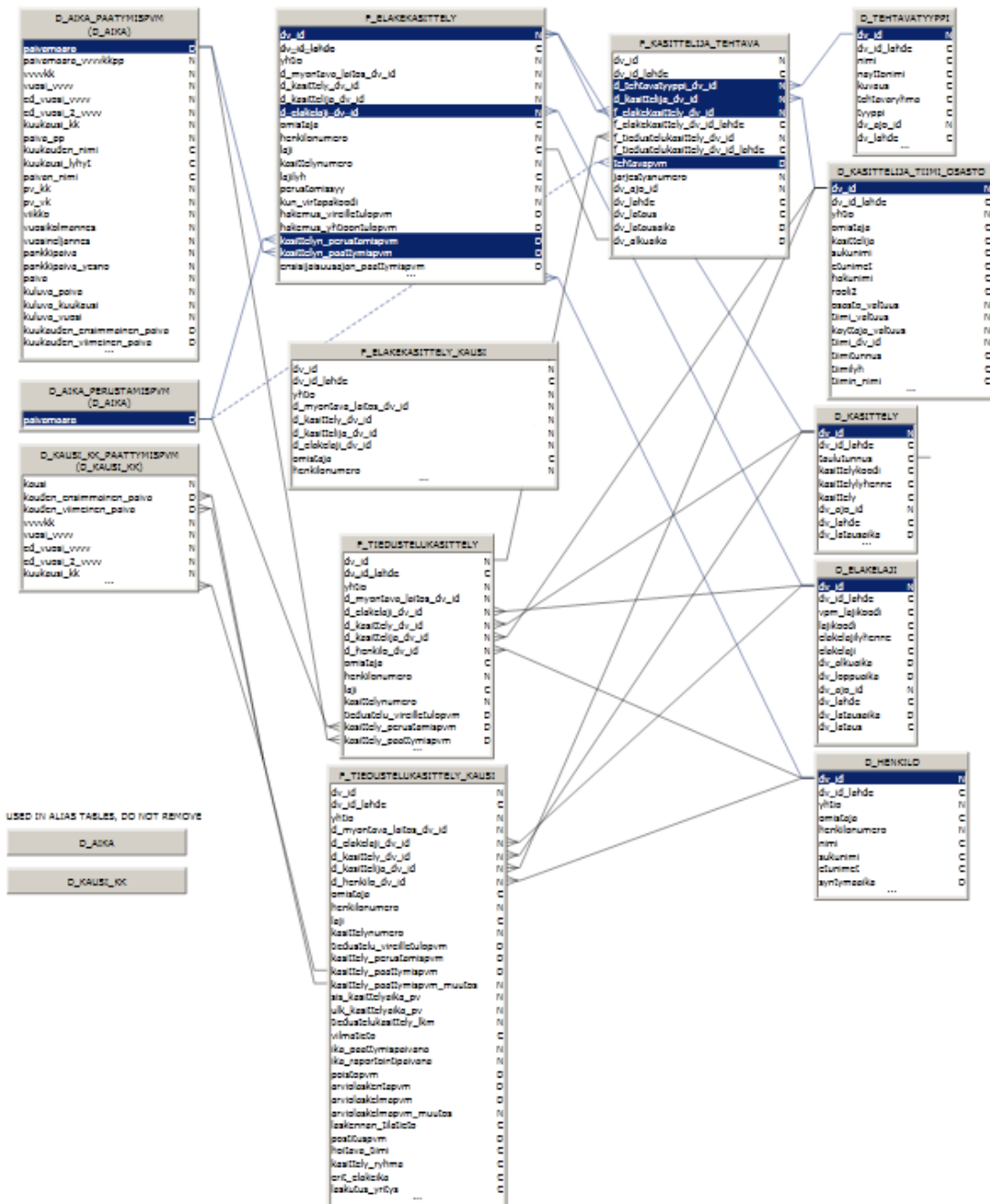
Wikipedia. Data Vault Modeling. Luettavissa:
http://en.wikipedia.org/wiki/Data_Vault_Modeling. Luettu: 29.3.2010.

Liitteet

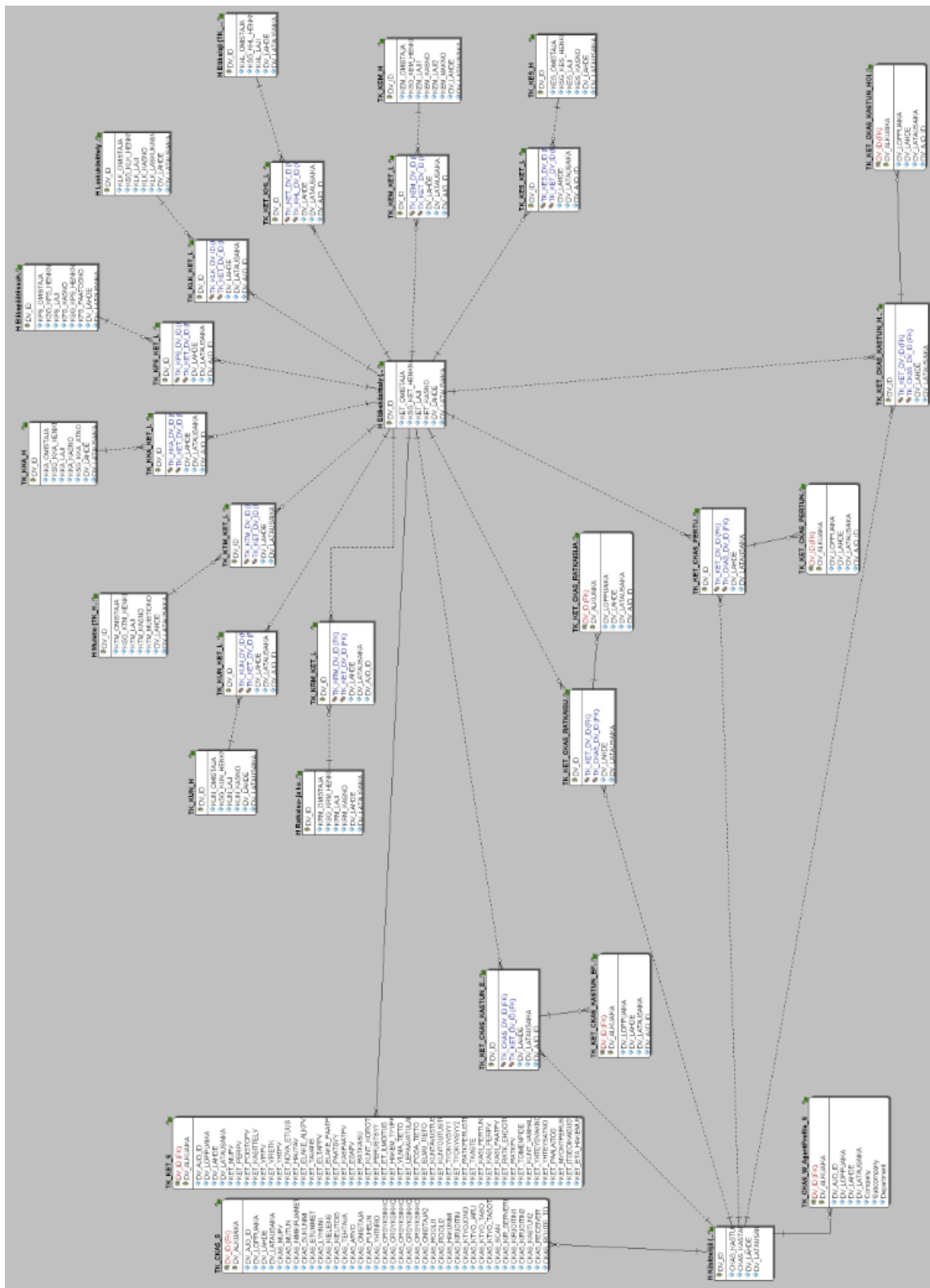
Liite 1. Eläkekäsittelyn tähtimalliin perustuva Universe BI-välineessä

Universe: XIR3_KEH-ELAKE_AFF-ETO

6/23/2011



Liite 2. Eläkekäsittelyn Data Vault -rakenne tallennuskerroksessa



Liite 3. Käsittelijätehtävän hyödyntäminen raporteilla (luottamuksellinen)

Liite 4. Nyky- ja tavoitetilahaastattelujen analyysit (luottamuksellinen)