



TAMPEREEN TEKNILLINEN YLIOPISTO

**KRISTIAN AF HÄLLSTRÖM**  
**AGILE BI - LIKETOIMINTATIEDON HALLINTAYMPÄRISTÖN**  
**KETTERÄ RAKENTAMINEN**

Diplomityö

22. maaliskuuta 2011

Tarkastaja: professori Mika Hannula

Tarkastaja ja aihe hyväksytty

Teknis-taloudellisen tiedekuntaneuvoston  
kokouksessa 6. lokakuuta 2010

# TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Tietojohtamisen koulutusohjelma

af HÄLLSTRÖM, KRISTIAN: Agile BI - Liiketoimintatiedon hallintaympäristön

ketterä rakentaminen

Diplomityö, 64 sivua, 2 liitesivua

Maaliskuu 2011

Pääaine: Tiedonhallinta

Tarkastaja: professori Mika Hannula

Avainsanat: Tiedonhallinta, liiketoimintatiedon hallinta, business intelligence, ketterät menetelmät, agile BI, ketterä liiketoimintatiedon hallinta, tietovarasto, ETL-prosessi

Liiketoimintatiedon hallinta, eli business intelligence on noussut tärkeäksi kilpailutekijäksi ja merkittäväksi osaksi päätöksentekoprosessia. Liiketoimintatiedon hallintaprojektit ovat monimutkaisia ja vaativat paljon resursseja. Tämän tutkimuksen tavoitteena oli löytää keinoja ketterään liiketoimintatiedon hallintaan. Työssä käytetty tutkimusstrategia oli toimintatutkimus. Teemahaastattelujen lisäksi työn tekijä keräsi tietoa osallistamalla projekteihin.

Liiketoimintatiedon hallinnan ohjelmistoprojektit ovat tyypillisesti toteutettu vesiputousmallin mukaisesti. Ketterät menetelmät ovat yleistyneet voimakkaasti ohjelmistokehityksessä ja seuraavaksi niitä tullaan soveltamaan BI-projekteissa. Agile BI:llä, eli ketterällä liiketoimintatiedon hallinnalla tarkoitetaan useimmiten ketterien menetelmien käyttöä liiketoimintatiedon hallinnan projekteissa. Ketteryys viittaa kuitenkin myös tietovaraston ketterään rakentamiseen sekä muutoksille ketterään liiketoimintatiedon hallintaympäristöön.

Haastattelujen perusteella ketteriä menetelmiä ei ole vielä hyödynnetty kohdeyrityksen tietovarastointiprojekteissa, mutta niiden potentiaali on tiedossa. Kohdeyrityksessä ei olla varmoja ketterien menetelmien soveltumisesta sellaisenaan liiketoimintatiedon hallintaan. Suuria mahdollisuuksia on automatisoiduissa ETL-prosesseissa ja valmiissa rajapinnoissa järjestelmiin.

Tässä tutkimuksessa toteutettiin käytännön sovellus, joka automatisoi osan ETL-prosessista Salesforce-järjestelmästä tietokantaan. Työssä toteutettu sovellus luo automaattisesti tietokannan taulut, lataa halutut tiedot Salesforcesta ja tallentaa ne tietokantaan. Sovellus on erittäin joustava muutoksille ja helposti kopioitavissa. Se vähentää huomattavasti rutiiniväilyä, vaikka sillä toteutetaan vain osa koko BI-ympäristöstä. Automatisoitu ETL-prosessi tallentaa metatietoa, jolloin tiedon läpinäkyvyys paranee ja muutokset ovat helpommin havaittavissa. Metatiedon ansiosta tietovaraston rakentaminen ja siihen liittyvät ETL-rutiinit on helpompi toteuttaa.

## ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Information and Knowledge Management  
at HÄLLSTRÖM, KRISTIAN: Agile BI - Business Intelligence Environment, the  
Construction of Agile

Master of Science Thesis, 64 pages, 2 Appendix pages

March 2011

Major: Business Information Management

Examiner: Professor Mika Hannula

Keywords: Business intelligence, agile methods, agile BI, data warehouse, ETL-  
process

Business intelligence has become an important competitive factor and a significant part of the decision-making process. Business intelligence projects are complex and require a lot of resources. This study was aimed to find ways to adapt agile business intelligence. The research strategy used was action research. The writer collected information through interviews and by participating in projects.

Business intelligence software projects are typically carried out based on the waterfall model. Agile methods have emerged as a strong software development and next they will be applied for BI projects. By agile BI people most often mean BI projects that use agile methods. Agility, however, also refers to the construction of agile BI environment, and reacting to changes.

Based on interviews, agile methods have not yet been utilized in data warehousing projects in the case company, but the potential is realized. People in the case company are not sure of compatibility of agile methods in business intelligence projects. However, automated ETL processes and complete interfaces with the information systems have great potentials.

A Practical application was carried out in this study. It automates part of the ETL process from Salesforce information system into the database. Constructed application automatically creates the database tables, extracts the desired information from Salesforce and loads it into the database. The application is very flexible for changes and easily copied. The using of the application will significantly reduce the routine work, even when it is used only with a part of the BI environment. Automated ETL process saves metadata, which improves the transparency of information and the changes are more easily detected. Due to metadata the data warehouse construction and related ETL routines are easier to implement.

## ALKUSANAT

Haluan kiittää IPSS Oy:tä mahdollisuudesta suorittaa tämä diplomityö. Ohjaajaani Sami Helin ja esimieheni Antti Paappanen ehdottivat diplomityöni aihetta kesällä 2010. Kiitän heitä kannustamisesta ja hyvistä neuvoista. Kiitokseni kuuluvat myös haastatteluihin osallistuneille ja koko IPSS Oy:n henkilöstölle mukavasta työilmapiiristä.

Eryityisesti haluan kiittää työni tarkastajaa, professori Mika Hannulaa tärkeästä palautteesta työn loppuvaiheessa. Kiitän diplomityöseminaarin Jussi Okkosta sekä opponenteja hyvistä neuvoista. Kiitos myös teille, jotka olette olleet muuten kannustamassa ja tukemassa opinnäytetyöni tekemistä.

Suurimmat kiitokset kuuluvat rakkaalle morsiamelleni, Suville tuesta ja rakkaudesta diplomityöni sekä opintojeni aikana.

Espoossa, 22.3.2011

Kristian af Hällström

## SISÄLLYS

Tiivistelmä .....	I
Abstract .....	II
Alkusanat .....	III
Termit ja niiden määritelmät .....	VI
1. Johdanto .....	1
1.1. Tutkimuksen tavoite ja rajaukset .....	2
1.2. Tutkimusote ja -strategia .....	4
1.3. Työn rakenne.....	5
2. Liiketoimintatiedon hallinta .....	6
2.1. Tiedon tasot ja ulottuvuudet.....	6
2.2. Liiketoimintatiedon hallinnan määrittely ja termien monimuotoisuus .....	9
2.3. Liiketoimintatiedon hallinnan hyödyt ja tavoitteet.....	12
2.4. Tietovarastoinnin perusarkkitehtuuri.....	14
3. Liiketoimintatiedon hallintaprojekti ja ketterät menetelmät .....	17
3.1. Vesiputousmalli .....	17
3.2. Syklisiä malleja.....	19
3.3. Tietovarastointiprojektin vaiheet .....	22
3.4. Ketterät menetelmät.....	24
4. Ketterä liiketoimintatiedon hallinta .....	27
4.1. Ketterät menetelmät liiketoimintatiedon hallinnassa .....	27
4.1.1. Ketterien menetelmien hyödyt ja haitat liiketoimintatiedon hallinnan projekteissa.....	28
4.1.2. Milloin ketterät menetelmät sopivat liiketoimintatiedon hallinnan projekteihin .....	29
4.2. Ketterä liiketoimintatiedon hallintaympäristö .....	30
4.2.1. Muuttuva lähdedata .....	30
4.2.2. Ketterä raportointi.....	31
4.3. Liiketoimintatiedon hallintaympäristön rakentaminen ketterästi .....	31
5. Tutkimusaineisto ja menetelmät .....	35
5.1. Kohdeyritys .....	35
5.2. Työn taustaa .....	36
5.3. Tutkimusmenetelmät .....	37

6.	Haastattelut ja niiden tulokset .....	39
6.1.	Haastatteluprosessi .....	39
6.1.1.	Haastatellut henkilöt .....	39
6.1.2.	Haastattelukysymykset .....	40
6.1.3.	Haastattelujen kulku .....	40
6.2.	Haastattelujen tulokset.....	41
6.2.1.	Liiketoimintatiedon hallintaprojektit kohdeyrityksessä .....	41
6.2.2.	Tietovaraston rakentaminen kohdeyrityksessä .....	43
6.2.3.	Ketterä liiketoimintatiedon hallinta ja tulevaisuudennäkymät.....	44
6.3.	Yhteenveto haastatteluista.....	45
7.	Käytännön sovellus.....	47
7.1.	Kohdeympäristön kuvaus .....	47
7.1.1.	Salesforce .....	48
7.1.2.	Pentaho Data Integrator – PDI .....	48
7.1.3.	ETL-prosessi kohdeympäristössä .....	49
7.2.	Tiedon latauksen automatisointi .....	49
7.2.1.	Ohjelman toiminta .....	49
7.2.2.	Generoituvat tiedostot ja niiden toiminta .....	50
7.3.	Tulokset .....	53
8.	Päätelmät .....	55
8.1.	Yhteenveto .....	55
8.2.	Tutkimuksen tarkastelu .....	56
8.3.	Työn reliabiliteetin ja validiteetin arviointi.....	58
8.4.	Suosituksukset sekä kohdeyritykselle ja tiedeyhteisölle.....	59
	Lähteet.....	61

## TERMIT JA NIIDEN MÄÄRITELMÄT

Ad hoc -kysely	Ennalta määrittelemättömät kyselyt, joilla haetaan vastauksia yleensä vain yksittäisiin tietotarpeisiin.
Agile Methods	Ks. ketterät menetelmät
Avoin lähdekoodi	Tuottamis- ja kehittämismenetelmä tietokoneohjelmistoille, jotka ovat vapaasti levitettävissä ja välitettävissä, joiden lähdekoodi on vapaasti saatavilla ja käyttötarkoitukset ja -oikeudet ovat vapaita (engl. Open Source).
BI, Business Intelligence	Ks. liiketoimintatiedon hallinta.
CRM-järjestelmä	Tietojärjestelmä, joka sisältää asiakaslähtöisen ajattelutavan, suomeksi asiakkuudenhallintajärjestelmä (engl. Customer Relationship Management).
Dashboard	Mittaristo on keskeisten tunnuslukujen seurantaan tarkoitettu BI-työkalu, joka sisältää visuaalisia komponentteja.
Data	Data on alimman tason tietoa, kuten yksittäinen merkki tai bitti. Se on tiedon muoto, jolla ei itsessään ole merkitystä.
Data Mart	Tietovarastoa suppeampi tietokanta, joka tukee tietyn aihealueen tai käyttäjäryhmän raportointia. Suomeksi se on tietokomero, paikallistietovarasto tai alitietovarasto. Myös OLAP-kuutio voi olla tietokomero.
Data Mining	Ks. tiedonlouhinta
Dimensio	Dimensio eli ulottuvuus on tähtimallin ja kuution osa, joka tallennetaan dimensiotauluihin. Kuvaa näkökulmaa, kuten aika, tuote, asiakas tai alue, joiden kautta halutaan faktoja tarkastella.
DW, Data Warehouse	Ks. tietovarasto
ERP	Tietojärjestelmä, jonka avulla hallinnoidaan yrityksen tai organisaation tärkeimpiä toimintoja, suomeksi toiminnanohjausjärjestelmä (engl. Enterprise Resource Planning).
ETL-prosessi	Prosessi, jossa tietoa kerätään, jalostetaan ja ladataan tietovarastoon (engl. Extract Transfor Load).
Fakta	Tähtimallin osa, joka kuutiossa sisältää mitattavan suureen, mittarin. Fakta on tietokannan faktatauluihin tallennettava tapahtumatyyppinen tieto, kuten myynti-, varastosaldo- tai asiakaskäyntitieto.
Informaatio	Informaatio on toiseksi alimman tason tietoa, dataa, johon liittyy jokin merkitys taso.
Java	Oliopohjainen, laitteistoriippumaton ohjelmointikieli

Ketterät menetelmät	Joukko ohjelmistokehityksen menetelmiä, jotka perustuvat iteratiiviseen ja inkrementaaliseen kehitykseen. Yhteistä näille on toimivan sovelluksen ensisijaisuus, suora viestintä ja nopea reagointi muutoksiin.
Liiketoimintatiedon hallinta	Päätöksenteon tukiprosessi, jossa tietoa kerätään eri lähteistä, jalostetaan, esitetään ja analysoidaan päätöksenteon tueksi.
Metadata	Metadata eli metatieto on tietoa tiedosta. Metadata kuvaa tietojen sisällön ja merkityksen.
OLAP-kuutio	Moniulotteisen tiedon käsittelyyn ja analysointiin suunniteltu rakenne. Mahdollistaa tiedon analysoimisen ja porautumisen tehokkaasti (engl. Online Analytical Processing).
SaaS	Ohjelmiston hankkimista ja tarjoamista palveluna lisenssiperustaisen tavan sijaan (engl. Software as a Service).
Scrum	Projektinhallinnan menetelmä, jota käytetään yleisesti ketterässä ohjelmistokehityksessä.
SQL	Kyselykieli, jolla tietokannan dataa tai metadataa haetaan, päivitetään, lisätään (engl. Structured Query Language).
Surrogaattiavain	Tiedon yksilöivä keinotekoisesti luotu avain tietokannan taulussa.
Tiedonlouhinta	Menetelmä, jolla tietovaraston tietomassasta etsitään trendejä, riippuvuuksia ja korrelaatioita.
Tietovarasto	Tietokanta, joka on integroitu, yhdenmukainen kokoelma yrityksen tai organisaation tietoja. Se on suunniteltu tukemaan raportointia ja analysointia (engl. Data Warehouse).
Tähtimalli	Dimensionaalinen suunnittelumenetelmä, jota käytetään OLAP-kuutioissa. Nimi tulee tähteä muistuttavasta rakenteesta, jonka keskellä on tietokannan faktataulu ja sen ympärille kytketty dimensiotaulut (engl. Star schema).
XML	Yleisesti tuettu standardi ja tunnettu merkintäkieli, jota käytetään formaattina tiedonvälitykseen järjestelmien välillä ja formaattina dokumenttien tallentamiseen (engl. eXtensible Markup Language).



# 1. JOHDANTO

Liiketoimintatiedon hallinta ja tietovarastointi ovat nousseet koko ajan merkittävämmiksi yritysten menestymisen kannalta. Maailmanlaajuisesti tietohallintojohtajien tärkeimpinä prioriteetteina on ollut jo vuosia liiketoimintatiedon hallinta (Wailgum 2009; Kauppi 2011). Suomalaisista suuryrityksistä lähes jokainen, jopa 98 % käytti systemaattisesti liiketoimintatiedon hallintaa hyödyksi vuonna 2007 (Halonen & Hannula 2007, s. 5) ja pari vuotta myöhemmin jokainen suuryritys (Vuori & Hannula 2009, s. 4). Nykyään entistä enemmän myös pk-yrityksissä otetaan käyttöön tietovarastoja ja niiden analysointivälineitä, raportointia ja mittaristoja.

Globalisaatio on aiheuttanut markkinoiden yhä nopeampaa muuttumista, jota on kuvattu kasvavalla kellotaajuudella (Halonen & Hannula 2007, s. 3). Liiketoiminnan tarpeet muuttuvat kellotaajuuden kasvun takia entistä nopeammin, samoin myös lähdejärjestelmät ja niiden tietosisältö. Tämä tuo paineita kehittää liiketoimintatiedon hallintaa. Liiketoimintatiedon hallintaympäristön, eli BI-ympäristön tulee olla entistä joustavampi ja ketterämpi kyetäkseen vastaamaan jatkuviin muutoksiin. Tietovaraston rakentaminen tulisi myös olla entistä nopeampaa ja tehokkaampaa, jotta resursseja voitaisiin keskittää joustavuuden parantamiseksi sekä säästää ylläpitoon. Lisäksi resurssien säästö tuo tietovarastoinnin yhä lähemmäksi pieniä yrityksiä.

Agile BI on uusi termi, jota on alettu käyttämään BI-ohjelmistotoimittajien keskuudessa ja hieman julkisessa keskustelussa. Se tarkoittaa suoraan suomennettuna ketterää liiketoimintatiedon hallintaa. Ketteryys viittaa ketteriin menetelmiin, joita on käytetty aikaisemmin ohjelmistotuotannossa. Se voi tarkoittaa myös ketterää tietovaraston rakentamista tai muutoksille ketterää ja joustavaa liiketoimintatiedon hallintaa.

Tämä diplomityö on tehty IPSS Oy:lle (Intelligence Precision Solution and Services Oy) syksyn 2010 ja sitä seuraavan talven aikana. IPSS Oy on asiakasjohtamisen ja -hallinnan teknologiaratkaisuihin keskittynyt konsulttiyritys. Helsingissä toimipaikkaa pitävä IPSS Oy on vuonna 1999 perustettu jatkuvasti kasvava pk-yritys. IPSS Oy tarjoaa omien asiakashallintajärjestelmien lisäksi yhteistyökumppaneiden järjestelmiä, joita tarvittaessa täydennetään omilla komponenteilla. Liiketoimintatiedon hallinta liittyy asiakashallintajärjestelmiin olennaisesti, tietovarastointia käytetään muun muassa asiakasymmärryksen kasvattamisessa, potentiaalin tunnistamisessa ja toiminnan mittaamisessa.

## 1.1. Tutkimuksen tavoite ja rajaukset

Liiketoiminnassa ketteryydellä (engl. agile) tarkoitetaan kykyä reagoida ja sopeutua muutoksiin nopeasti ja tehokkaasti. Tämän tutkimuksen tavoitteena on selvittää mitä ketteryys tarkoittaa liiketoimintatiedon hallinnassa ja miten ketterään liiketoimintatiedon hallintaan päästään. Työssä tarkastellaan ketteriä menetelmiä ja niiden hyötyjä liiketoimintatiedon hallinnan projekteissa. Uusia ratkaisuja pyritään löytämään myös tietovarastojen rakentamiseen. Tavoitteena on helpommin hallittava projekti, muutosvalmiuden kohottaminen ja tarvittavien resurssien minimoiminen.

Tutkimuksen tavoite voidaan esittää myös ongelman asettelun muodossa. Tämän työn tutkimusongelma on liiketoimintatiedon hallinnan kehittäminen ketterämmäksi. Tutkimuksen näkökulma on SaaS-liiketoiminnan (Software as a Service) liiketoimintatiedon hallintaprojektit, jotka kuuluvat tutkimuksen tilanteen yrityksen toimintaan. Erityisen lähellä on asiakashallintaprojektit, muun muassa Salesforce-projektit, johon työn empiirisen osan sovellus tulee liittymään.

Tutkimusongelma esitetään usein tutkimuskysymyksen muodossa. Tutkimuskysymys on laaja ja kuvaa työn tavoitteita hyvin. Siihen pyritään löytämään vastauksia ja ratkaisuja tutkimuksen aikana. Tämän työn tutkimuskysymys on seuraavanlainen:

- Miten liiketoimintatiedon hallintaa voisi kehittää ketterämmäksi?

On kuitenkin hyvin vaikea saada työn tavoitteita yhteen kysymykseen, joten pääkysymyksen alle lisätään usein alakysymyksiä tai alaongelmia. Tämän tutkimuksen alakysymykset ovat seuraavat:

1. Mitä tarkoittaa ketterällä liiketoimintatiedon hallinnalla ja kuinka se määritellään?
2. Miten ketteriä menetelmiä voidaan hyödyntää liiketoimintatiedon hallinnan projekteissa?
3. Miten voidaan suoraviivaistaa liiketoimintatiedon hallintaympäristöjen rakentamista ja rakentaa ympäristö ketterämmin?
4. Miten liiketoimintatiedon hallintaympäristöä voidaan kehittää ketterämmäksi?

Ensimmäinen alatutkimuskysymys on tärkeä, sillä se voi tuoda kokonaan uusia näkökulmia aiheeseen. Siihen liittyy myös seuraavat alatutkimuskysymykset. Kysymykseen pyritään löytämään vastauksia työn teoriaosan viimeisessä luvussa. Termin määrittämisen lisäksi työssä pyritään pohtimaan sen tarkoitusta, tavoitteita ja menetelmiä.

Toiseen alatutkimuskysymykseen pyritään myös löytämään vastauksia teoriaosassa, mutta lisäksi sitä tullaan pohtimaan myös työn empiirisessä osassa. Tähän kysymykseen haetaan vastauksia erityisesti työssä toteutettavien teemahaastattelujen pohjalta.

Haastateltavat henkilöt ovat työskennelleet liiketoimintatiedon hallinnan projekteissa ja ovat käyttäneet ketteriä menetelmiä enemmän tai vähemmän. Haastatteluilla pyritään löytämään sopivia käytäntöjä ketteristä menetelmistä liiketoimintatiedon hallinnan projekteissa. Työssä kuitenkin keskitytään tietovarastojen rakentamiseen käytännön toimina, ei projektinhallinnan tai johtamisen näkökulmasta.

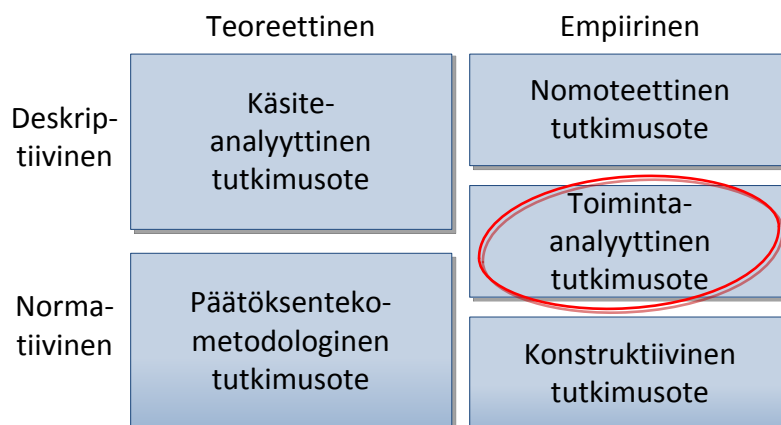
Kolmanteen tutkimuskysymykseen pyritään löytämään ratkaisuja teoriaosassa, haastatteluilla sekä käytännön sovelluksessa. Teoriaosan tutkimuksella ja teemahaastatteluilla haetaan uusia ideoita. Lisäksi haastatteluilla kartoitetaan kohdeyrityksen nykytilannetta ja etsitään liiketoimintatiedon hallintaprojektien kehittämiskohtia. Toimintaa pyritään suoraviivaistamaan ja rakentamaan tietovarasto ketterästi ja nopeasti, kuitenkin ylläpitoa unohtamatta. Toimintoja automatisoimalla vähennetään rutiinityötä sekä voidaan keskittyä haastavampiin ja tärkeämpiin projektin osiin.

Neljäs alatutkimuskysymys keskittyy puolestaan liiketoimintatiedon hallintaympäristöön, kuinka siitä saadaan ketterämpi ja kuinka muutoksiin voidaan reagoida paremmin ja nopeammin. Liiketoiminnan tarpeet ja lähdejärjestelmien tietorakenteet voivat muuttua, ja näihin muutoksiin on pystyttävä vastaamaan. Tutkimuskysymystä tarkastellaan teoriaosan lisäksi haastatteluiden kautta sekä myös käytännön sovelluksessa otetaan liiketoimintatiedon hallintaympäristön ketteryys huomioon.

Tutkimuksessa pyritään tarkastelemaan ketterää liiketoimintatiedon hallintaa kokonaisuutena. Työn empiriaosassa keskitytään kuitenkin liiketoimintatiedon hallintaympäristöön ja sen rakentamiseen. Tutkimuksen ulkopuolelle rajataan tarvemäärittely sekä pitkälti myös tietovaraston analysointi, raportointi, mittaristot ja tiedonlouhinta. Nämä vaativat lähes aina vahvaa asiakaskohtaista räätälöimistä, sillä tarpeet voivat olla hyvin erilaisia. Rajauksen ulkopuolelle jää myös toiminnan johtaminen, projektinhallinta sekä kustannuslaskenta. Työ tehdään yksittäiselle yritykselle ja empirian sovellus tietylle projektille.

## 1.2. Tutkimusote ja -strategia

Liiketaloustieteiden tutkimusotteet jaetaan kuvan 1.1 mukaan viiteen otteeseen empirian ja normatiivisuuden mukaan. Tämän tutkimuksen tutkimusote sijoittuu toiminta-analyyttiseen, eli pohjautuu vahvasti empiriaan sekä on deskriptiivisen ja normatiivisen välillä.



Kuva 1.1 Teollisuustalouden nelikenttä mukailten lähteestä (Kasanen et al. 1991, s. 317 ks. Okkonen 2010)

Työ pohjautuu empiriaan, eli käytännön työhön. Tavoitteena ei ole keskittyä kohteen kuvaamiseen, kuten deskriptiivisemmässä nomoteettisessa tutkimusotteessa. Tarkoituksena ei myöskään ole luoda normeja ja konkreettisia toimintamalleja miten tulisi toimia, kuten normatiivisessa otteessa. Tämän tutkimuksen tavoite on kuvailla tutkittavaa kohdetta ja luoda ohjeita tai suosituksia, kuten yleensä toiminta-analyttisessä tutkimusotteessa.

Tutkimusstrategiat jaetaan perinteisesti kolmeen ryhmään: kokeelliseen-, survey- ja tapaustutkimukseen. Viimeisen, niin sanotun case-tutkimuksen tarkoituksena on kerätä yksityiskohtaista tietoa ja ymmärrystä tietystä, yksittäisestä tutkimuskohteesta sen omassa toimintaympäristössä. (Hirsjärvi et al. 1997, s. 130.) Tämän tutkimuksen tutkimusstrategiaksi on valittu tapaustutkimus. Toimintaympäristö on IPSS Oy ja tutkimuskohde liiketoimintatiedon hallinnan projektit.

Saunders et al. (2009, s. 138) jakavat tutkimusstrategiat seitsemään ryhmään: edellä mainittujen kolmen lisäksi toimintatutkimukseen, aineistolähtöiseen analyysiin, etnografiaan sekä arkistotutkimukseen. Tähän työhön niistä sopii parhaiten toimintatutkimus. Tutkija kartoittaa kohdeyrityksen kehittämiskohtia ja ketterän liiketoimintatiedon hallinnan mahdollisuuksia. Lisäksi hän pyrkii tutkimuksessa muuttamaan ja kehittämään tutkittavaa kohdetta olemalla itse aktiivinen ja osallistujana.

Tutkimuksen yleinen tavoite voi olla kartoittava, kuvaileva tai selittävä (Saunders et al. 2009, ss. 139–140). Tässä tutkimuksessa tavoite on kartoittava. Tutkimuksessa ei

keskitytä kuvailuun tai selittelyyn, vaan kartoitetaan uusia toimintatapoja ketterämpään toimintaan ja ketterämpään liiketoimintatiedon hallintaan.

Työn teemahaastatteluilla kartoitetaan nykytilannetta ja etsitään kehittämiskohtia. Lopuksi toteutetaan pienimuotoinen käytännön sovellus, jolla pyritään vastaamaan joihinkin kehittämiskohtiin. Tutkimuksen tarkoitus ei ole vertailla nykyistä toimintaa aiempaan tai tulevaan. Liiketoimintatiedon hallinnan projektit ovat tyypillisesti hyvin erilaisia, joten voisi olla vaikea verrata niitä suoraan toisiinsa.

Suhde teoriaan on tässä tutkimuksessa induktiivinen. Teorian pohjalta pyritään löytämään uusia toimintatapoja ja käytäntöjä. Tämän takia on ensin perehdyttävä teorial tietoon, vaikka teoria- ja empiriaosaa tehdään rinnakkain.

Tutkimuksessa kerättiin tietoa monimetodisesti, mutta vain kvalitatiivisia menetelmiä käyttäen. Työ sisältää teemahaastatteluita, joilla kerättiin tietoa kohdeyrityksestä, sen toimintatavoista ja kokemuksista. Teemahaastattelut on valittu yhdeksi metodiksi, sillä ne ovat helposti toteutettavissa ja kohdeyrityksellä on pitkä kokemus alalta. Kohdeyrityksen toimintaa haluttiin kehittää, joten luontevin ratkaisu saada tietoa nykytilasta ja kehittämiskohdista oli haastattelut.

Lisäksi työssä hyödynnettiin tekijän kokemusperäistä tietoa. Kokemusta on kerrytetty työn aikana ja muutamia kuukausia ennen työn aloittamista. Työntekijä on toiminut erilaisissa liiketoimintatiedon hallinnan tehtävissä tutkimuksen tekemisen ohella. Kokemustieto oli erittäin tärkeää kehittämiskohteiden löytämisessä ja arvioimisessa sekä käytännön sovellusta rakennettaessa. Tutkimusmenetelmistä lisää työn empiriaosassa luvussa 5.3.

### **1.3. Työn rakenne**

Teoriaosan ensimmäisessä luvussa perehdytään liiketoimintatiedon hallintaan, käydään läpi tiedon tasot ja ulottuvuudet, määritellään liiketoimintatiedon hallinnan ja sen lähitermit sekä käydään läpi sen hyödyt, tavoitteet ja perusarkkitehtuuria. Toisessa teorialuvussa käsitellään projektimalleja ja menetelmiä liiketoimintatiedon hallinnan projekteista ja ketteristä menetelmistä. Viimeisessä teorialuvussa käsitellään ketterää liiketoimintatiedon hallintaa.

Empiriaosan aluksi esitellään kohdeyritys, kerrotaan työn taustaa ja tutkimusmenetelmät. Toinen empirialuku käsittelee haastatteluita ja niiden tuloksia. Seitsemäs luku, eli empirian kolmas luku käsittelee käytännön sovellusta, ketterää tiedonlatauksen toteutusta. Siinä esitellään kohdeympäristö, tehdyn sovelluksen toiminta ja tulokset. Lopuksi esitellään työn päätelmät.

## 2. LIKETOIMINTATIEDON HALLINTA

Tässä luvussa käsitellään liiketoimintatiedon hallintaa yleisesti, mitä se on ja mitä sillä tavoitellaan. Luvun tarkoituksena on rakentaa ymmärrystä aihepiiristä. Aluksi määritellään tieto, sen eri tasot ja ulottuvuudet. Sen jälkeen käsitellään liiketoimintatiedon hallintaa ja muita lähitermejä. Kolmannessa alaluvussa käsitellään liiketoimintatiedon hallinnan tavoitteita ja hyötyjä, neljännessä käydään läpi tietovarastoinnin perusarkkitehtuuria.

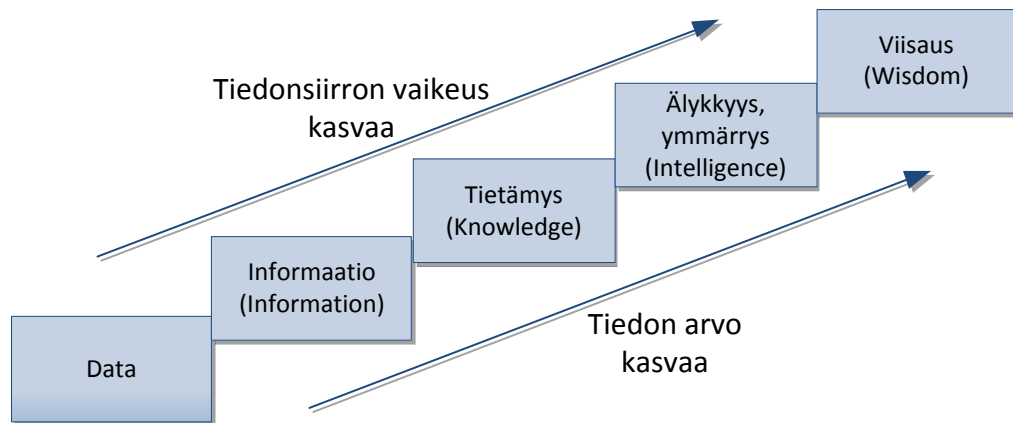
### 2.1. Tiedon tasot ja ulottuvuudet

Tunnetun, ja jo kuluneen sanonnan ”*Nam et ipsa scientia potestas est*”, eli suomeksi tieto on valtaa, sanoi ensimmäisen kerran Sir Francis Bacon jo 1500-luvun lopussa (Miettinen 2002). Tietoa on arvostettu aikojen alusta lähtien ja sen on uskottu tuovan rikkautta ja tiedon avulla on uskottu voitettavan sotia. Raamattuun on kirjoitettu, että parempi viisaus kuin hopea, tuottoisampi tieto kuin kulta (Sananl. 3:14).

Mitä tieto sitten on? Platon määritteli aikoinaan, että tieto on hyvin perusteltu tosi uskomus (Markkula et al. 2001, s. 33; Kaario & Peltola 2008, s. 6). Tämä vanha ja tunnettu määritelmä on kolmiosainen. Tiedon täytyy pystyä perustelemaan hyvin, tiedon pitää olla totta ja tietoon uskotaan. Tietoa on kuitenkin vaikea määritellä selkeästi ja tarkasti. Tietoteoria, eli epistemologia on tutkinut tietoa vuosisatojen ajan, eikä vielä ole löytänyt yksikäsitteistä, selkeää ratkaisua tiedon käsitteelle (Sydänmaanlakka 2007, s. 189). Tämän työn kannalta tärkeämpään on tarkastella tiedon eri muotoja ja ymmärtää mitä tieto ja liiketoimintatieto voi olla, kuin määritellä tieto lyhyesti ja eksaktisti.

Tiedon tasot (Kuva 2.1) koostuvat datasta, informaatiosta, tietämyksestä, älykkyydestä ja viisaudesta. Davenportin ja Prusakin (1998, s. 2) mukaan data on joukko irrallisia, objektiivisia ja jäsenettyjä faktoja tapahtumista. Sydänmaanlakan (2007, s. 187) mukaan data on numeroita, tekstiä, kuvia tai niiden yhdistelmiä, informaation raaka-ainetta. Data on irrallista, joten sitä ei voi ymmärtää ilman kontekstia (Pirttimäki 2007, s. 39). Tämä tarkoittaa sitä, että kuvat voitaisiin paremminkin luokitella informaatioksi, sillä niihin liittyy usein merkityksiä, kuvat voidaan ymmärtää yhdellä silmäyksellä ja niistä voidaan myös oppia. Kuvat kuitenkin sisältävät dataa. Toisaalta symbolit ja merkit ovat myös kuvia, jotka puolestaan ovat selkeämmin dataa. Informaatioon puolestaan liittyy merkitys ja sen voi oppimisen kautta muuttaa tietämykseksi. (Markkula et al. 2001, s. 31.)

Kaarion ja Peltolan (2008, s. 6) mukaan numero kolme on dataa, mutta siihen liitetty Celsius-merkintä antaa sille merkityksen ja se muuttuu informaatioksi. Toisaalta lämpötilasymboli; asteen merkki ja C-kirjain ovat myös dataa, eikä lämpötila vieläkään sisällä kontekstia. Data muuttuu informaatioksi vasta kun lisätään esimerkiksi paikka ja aika, tai kun ymmärretään sen konteksti.



Kuva 2.1 Tiedon tasot (mukaillen lähteistä Sydänmaanlakka 2008, s. 188; Thierauf 2001, s. 8; Kaario & Peltola 2008, s. 8)

Informaatio on dataa, johon on liitetty jokin merkitys. Se muuttuu tietämykseksi vasta, kun ihminen on vastaanottanut ja prosessoinut sen (Pirttimäki 2007, s. 39). Tietämys on aina ihmisen oman prosessoinnin tulosta ja siten kontekstisidonnaista (Markkula et al. 2001, s. 31). Tietämys on siis aina ihmisissä ja toisin kuin informaatiota, sitä ei voi tallentaa esimerkiksi paperille.

Älykkyys, ymmärrys tai näkemys sekä viisaus onkin hieman hankalampi määritellä täsmällisesti. Useimmiten älykkyyteen tai ymmärrykseen liitetään kokemus aiemmasta tietämyksestä. Markkula et al. (2001, s. 31) esittää ymmärrykseen sisältyvän myös selityksiä miksi tieto on sellaista ja miten se kytkeytyy muihin asioihin. Thierauf (2001, s. 8) kuvaa viisauden kyvyksi arvioida asioita järkevästi yli ajan. Michel de Montaigne on todennut että voimme olla tietäviä muiden tiedoilla, mutta emme viisaita muiden viisaudella (Sydänmaanlakka 2008, s. 191). Tämä pätee hyvin myös älykkyyteen.

Esimerkiksi ”+16 °C” on dataa, päivän ulkolämpötila tietyssä paikassa tiettyyn kellonaikaan on informaatiota, keskilämpötilan pitkäaikainen nousu tietämystä sekä lämpötilan nousun syiden ja seurausten ymmärtäminen voisi olla älykkyyttä tai ymmärrystä.

Kuten kuvassa 2.1 on esitetty, tiedon siirron vaikeus ja tiedon arvo kasvaa datasta viisauteen. Viisautta ei voi siirtää ihmiseltä toiselle edes koneiden avulla, vaan viisauden kartuttamiseen tarvitaan paljon aikaa ja perehtymistä (Kaario & Peltola 2008, s. 8). Davenportin ja Prusakin (1998, s. 5) mukaan informaatio muuttuu tiedoksi

kokemuksien, johtopäätöksien ja keskustelujen kautta. Tästä voidaan päätellä, että tietoliikenneverkkojen kautta voidaan välittää vain dataa ja informaatiota.

Tieto jaetaan perinteisesti myös hiljaiseen (piilevään, engl. tacit knowledge) tietoon ja eksplisiittiseen (havaittavaan, engl. explicit knowledge) tietoon (Lönngqvist et al. 2005, s. 36). Tämän esitti Michael Polanyi jo 1950-luvulla, mutta Nonaka ja Takeuchi tekivät sen tunnetuksi paljon myöhemmin (Sydänmaanlakka 2008, s. 192; Pirttimäki 2007, s. 40; Alavi & Leidner 2001, s. 110). Hiljainen tieto on asiantuntemusta, osaamista ja kokemusta, jota kaikilla ihmisillä on aivoissaan (Kaario & Peltola 2008, s. 7). Markkulan et al. (2001, s. 36) mukaan hiljainen tieto sisältää käsityksiä, tietotaitoa, tunteita, kokemuksia uskomuksia, arvoja, ideaaleja ja intuitiota. Eksplisiittistä tietoa puolestaan on kaikki kirjoitettu tieto ja tietokannoissa oleva tieto. Nonakan (1991, s. 98) mukaan eksplisiittinen tieto on muodollista ja järjestelmällistä tietoa, esimerkiksi tuotedokumentti, tieteellinen kaava tai tietokoneohjelma, joten sitä on helppo jakaa ja viestittää.

Tiedon tasoihin yhdistettynä eksplisiittinen tieto on dataa tai informaatiota, kun hiljainen tieto voi olla tietämystä, ymmärrystä tai viisautta. Michael Polanyi on sanonut: ”Me voimme tietää paljon enemmän, kuin osaamme kertoa” (Nonaka 1991, s. 98). Tämä kuvaa hyvin tiedon jakamisen vaikeutta mitä ylempänä tiedon tason hierarkiassa ollaan.

Alavi ja Leidner (2001, s. 113) jakavat hiljaisen tiedon vielä kognitiiviseen hiljaiseen tietoon eli mentaaliin malleihin, ja tekniseen hiljaiseen tietoon, ns. ”know-how”-tietoon. Tieto voidaan jakaa tyypeihin, jotka vastaavat eri kysymyssanoihin, kun mitä, miten, miksi, mistä ja milloin -tietämiset (engl. know-how, know-who, know-what, know-why, know-when ja know-where) (Markkula et al. 2001, ss. 34–35). Alavi ja Leidner (2001, s. 113) eriyttävät ne hiljaisesta ja eksplisiittisestä tiedosta omiksi tyypeikseen. Useimmiten ne kuitenkin jaetaan erikseen, ja sisältyvät tavallisesti hiljaiseen tietoon. Sydänmaanlakka (2008, s. 190) lisää vielä sosiaalisen puolen, tunneälykkyyden sekä intuitiivisen älykkyyden.

Tässä työssä keskitytään vain tiedon tasoihin sekä tiedon lajeista hiljaiseen ja eksplisiittiseen tietoon. Liiketoiminnassa ja liiketoimintaympäristössä syntyy jatkuvasti kaiken tasoista tietoa ja sekä eksplisiittistä että hiljaista tietoa. Näitä eritasoisia ja muotoisia tietoja pyritään liiketoimintatiedon hallinnan välineillä keräämään, jakamaan ja havainnollistamaan, sekä luomaan uutta tietoa. Tietovarastointi keskittyy kuitenkin vain eksplisiittiseen tietoon, dataan ja informaatioon, sen keräämiseen ja esittämiseen.



## 2.2. Liiketoimintatiedon hallinnan määrittely ja termien monimuotoisuus

Liiketoimintatiedon hallinta eli Business Intelligence (BI) on päätöksenteon tukiprosessi, joka pyrkii analysoimaan, jalostamaan ja esittämään eri lähteistä koottua tietoa päätöksen teon tueksi (Kaario & Peltola 2008, s.61). Liiketoimintatiedon hallinnassa on olennaista myös kerätä tietoa esimerkiksi tietovarastoon (engl. data warehouse), jotta sitä voidaan analysoida, jalostaa ja esittää. Myös Dishman ja Calof (2008, s. 767) esittävät prosessia informaation keräämiseksi, analysoimiseksi ja raportoimiseksi strategisissa päätöksenteoissa, jolloin se on olennainen perusta strategisessa päätöksentekoprosessissa. Thieraufin mukaan (2001, s. xi) liiketoimintatiedon hallinta tarjoaa päätöksentekijöille korkeamman tason ymmärrystä yrityksen toiminnasta.

Liiketoimintatiedon hallinnasta on käytetty paljon muita termejä, kuten strategic intelligence, competitor analysis, competitive technical intelligence, market intelligence ja competitive intelligence (Dishman & Calof 2008, s. 767). Suomalaisissa yrityksissä käytetään pääasiassa termiä business intelligence, mutta sen lisäksi myös termejä competitive intelligence, kilpailijaseuranta, markkinaseuranta, market intelligence tai jotakin muuta (Halonen & Hannula 2007, s. 6). Suomenkielinen melko vakiintunut termi on liiketoimintatiedon hallinta, mikä myös kuvaa toimintaa parhaiten. Suomennettuja vastineita on liiketoimintatiedon hallinnan lisäksi esiintynyt myös termit yritystiedon rikastus, analyttinen tiedon hallinta ja tiedon hallinnan prosessi (Hovi et al. 2009, s. 78).

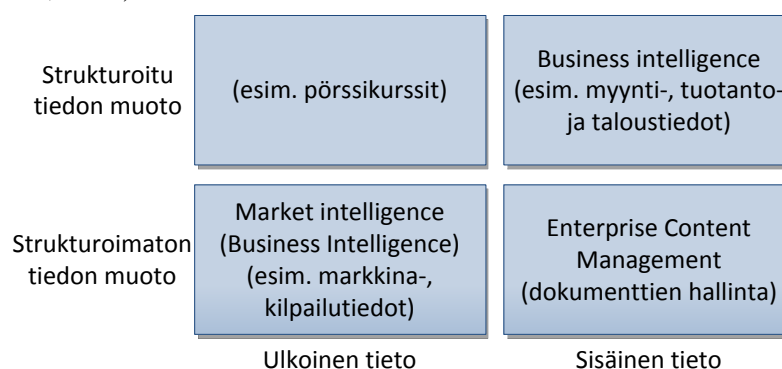
Liiketoimintatiedon hallinnan lähitermejä on muun muassa tietojohdaminen, aineettoman pääoman johtaminen, tietämyksenhallinta ja tiedonhallinta. Termejä on paljon, eikä kaikille ole yhteisesti hyväksytty määritelmää. Monien termien sisältö on hieman päällekkäin toisten kanssa, mutta termien näkökulmat ovat usein erilaiset. Tietojohdaminen (engl. knowledge management) luetaan usein pääkäsitteeksi, johon myös liiketoimintatiedon hallinta kuuluu. Toisinaan tietämyksenhallinta (engl. myös knowledge management) käsitetään samaksi kuin tietojohdaminen (Hovi et al. 2009, s. 190).

Aineeton pääoma koostuu inhimillisestä pääomasta (osaaminen, asenne, tieto, koulutus), suhdapäomasta (suhteet asiakkaisiin ja sidosryhmiin, maine, brandit, yhteistyösopimukset) ja rakennepääomasta (arvot ja kulttuuri, työilmapiiri, prosessit ja järjestelmät, dokumentoitu tieto ja immateriaalioikeudet) (Lönnqvist et al. 2005, s. 31). Aineettoman pääoman johtaminen linkittyy moniin yrityksen toimintoihin ja prosesseihin, ja liittyy läheisesti pääkäsitteen, tietojohdamiseen.

Lönnqvistin et al. (2005, s. 45) mukaan liiketoimintatiedon hallinta kuuluu osaksi rakennepääomaa, sillä liiketoimintatiedon hallintaan sisältyy prosesseja ja järjestelmiä, joilla tietoa kerätään, tulkitaan ja analysoidaan. Liiketoimintatiedon hallinta liittyy kuitenkin hyvin läheisesti myös inhimilliseen pääomaan ja suhdetietoon. Hiljainen tieto on aina inhimillistä pääomaa. Inhimillisen pääoman tietämystä pyritään kasvattamaan myös liiketoimintatiedon hallinnalla. Liiketoimintatieto liittyy usein asiakassuhteisiin ja muihin sidosryhmiin, joilta tietoa myös kerätään. Kuitenkin tekninen liiketoimintatiedon hallinta, eli lähinnä tietovarastointi kuuluu selkeästi rakennepääomaan.

Aineettoman pääoman johtaminen on lähellä tietämyksenhallintaa (engl. myös knowledge management), erityisesti aineettoman pääoman kehittäminen (Lönnqvist et al. 2005, s. 120). Näkökulma on vain hieman erilainen. Helanderin et al. (2007, s. 9) mukaan tietämyksenhallinta on toimintaa, jolla pyritään mahdollisimman tehokaaseen tiedon hyödyntämiseen yrityksen tai organisaation sisällä. Selkein ero onkin, että tietämyksenhallinta keskittyy jo yrityksessä olevaan tietoon, kun liiketoimintatiedon hallintaan liitetään liiketoimintaympäristöstä tietoa. Toinen keskeinen ero on se, että tietämyksenhallinnassa hyödynnetään jalostuneempaa tietoa, kuten tietämystä, kun taas liiketoimintatiedon hallinnassa dataa ja informaatiota (Koskinen et al. 2005, s. 6).

Tiedonhallinta (engl. Enterprise Content Management, ECM) voidaan nähdä kokonaisvaltaisena tai suppeampana käsitteenä. Laajemmin nähtynä se tarkoittaa kaiken organisaation liittyvän tiedon hallintaa, sisältäen sekä tietämyksenhallinnan että liiketoimintatiedon hallinnan (Kaario & Peltola 2008, s. 3). Tiedonhallinta (engl. information management) suppeampana käsitteenä tarkoittaa eksplisiittisen tiedon hallintaa. Toisaalta ECM voidaan kääntää dokumenttien hallinnaksi (kuva 2.2), jolloin se on huomattavasti suppeampi tarkoittaen strukturoimattoman, sisäisen tiedon hallintaa (Hovi et al. 2009, s. 79).



Kuva 2.2 Liiketoimintatiedon hallinnan tulkinnat (Hovi et al. 2009, s. 79).

Liiketoimintatiedon hallinnasta on Hovin et al. (2009, s. 79) mukaan kaksi näkemystä, sisäinen kvantitatiivinen ja ulkoinen, kvalitatiivinen näkemys (kuva 2.2). Toisaalta liiketoimintatiedon hallinta voidaan nähdä kokonaisvaltaisempana (engl. myös business

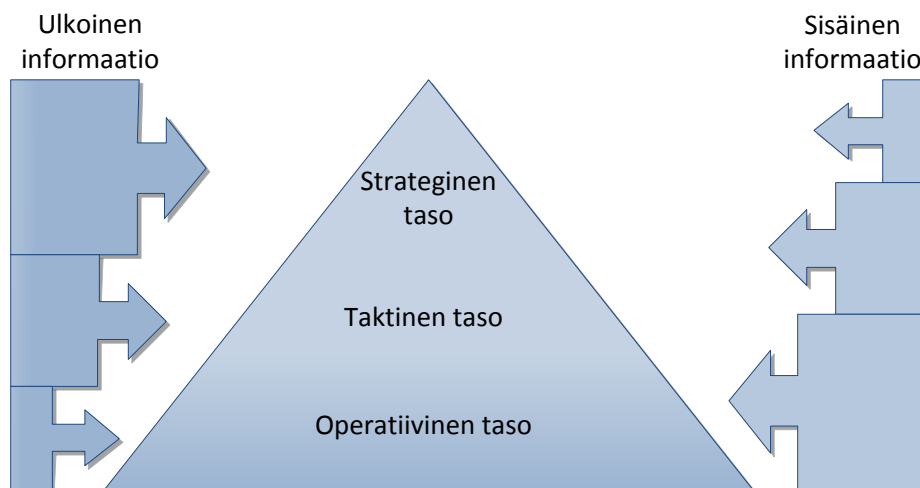
and competitive intelligence), jossa yhdistyy molemmat. Koskisen et al. (2005, ss. 5-6) mukaan kilpailutiedon hallinta (engl. competitive intelligence, CI) ja kilpailijaseuranta (engl. competitor intelligence) ovat liiketoimintatiedon hallinnan alaprosesseja. Toisin sanoen ne ovat suppeampia, ja sisältyvät liiketoimintatiedon hallintaan, eli he kannattavat myös kokonaisvaltaisempaa käsitystä. Kaario ja Peltola (2008, s. 61) näkevät myös liiketoimintatiedon hallinnan laajempänä terminä, joka yhdistää rakenteisen, kuten tuotantoluvut ja ei-rakenteisen tiedon, kuten kilpailijauutiset. Ja kukaan ei estä käyttämästä myös ulkoista, strukturoitua tietoa liiketoimintatiedon hallinnassa, esimerkiksi valuuttakursseja. On myös mahdollista yhdistää dokumenttien hallinta liiketoimintatiedon hallintaan, jolloin se sisältäisi kaikki kuvan 2.2 alueet.

Tietovarastointi (engl. data warehousing) on tiedon tallentamista tietovarastoon päätöksenteon hyödyntämiseksi. Tietovarastointi on kokoelma päätöksenteon tuen teknologioita, ja jonka tarkoituksena on mahdollistaa tietotyöntekijöiden (johtajien, esimiehien, analyytikoiden) paremmat ja nopeammat päätökset (Chaudhuri & Umeshwar 1997). Liiketoimintatiedon hallinta ja tietovarastointi tarkoittavat pitkälti samaa asiaa, molemmissa tietoa hyödynnetään päätöksenteossa. Tietovarastoinnissa korostuu tiedon kerääminen ja tallentaminen tietovarastoon, jotta sitä voidaan hyödyntää päätöksenteossa, kun taas liiketoimintatiedon hallinta sisältää sen lisäksi tiedon analysoimisen, jakamisen sekä tiedon analysoimisen myös muista, ulkoisista lähteistä. Liiketoimintatiedon hallinta on yleisempi ja laajempi termi, kuvaa paremmin toimintatapaa, kun taas tietovarastointi puolestaan teknisempi ja kapeampi termi.

Tässä työssä käytettävä termi liiketoimintatiedon hallinta, tai business intelligence tarkoittaa yllä mainittua kokonaisvaltaisempaa näkemystä, jossa kerätään rakenteista ja ei-rakenteista eksplisiittistä tietoa päätöksenteon tueksi. Tietovarastoinnissa keskitytään kuitenkin strukturoituun dataan ja ensisijaisesti sisäiseen liiketoimintatietoon.

### 2.3. Liiketoimintatiedon hallinnan hyödyt ja tavoitteet

Liiketoimintatiedon hallinnan ensisijainen tavoite on, kuten aiemmin on esitetty, kerätä ja analysoida tietoa päätöksenteon tueksi. Erilaista tietoa tarvitaan erilaisissa päätöksentekotilanteissa. Operatiivisen tason seurannassa ja päätöksentekotilanteissa tarvitaan pääosin sisäistä tietoa, taktisessa myös enemmän ulkoista ja strategisessa suunnittelussa pääosin ulkoista informaatiota (kuva 2.3).



Kuva 2.3 Ulkoisen ja sisäisen tiedon tarpeet eritasoisissa päätöksentekotilanteissa (Pirttimäki 2007, s. 45).

Operatiivisella tasolla keskitytään lisäksi kapeampaan ja yksityiskohtaisempaan tietoon ja historiaorientoituneeseen, strategisella tasolla puolestaan tietoa tarvitaan laaja-alaista, yhdisteltyä ja tulevaisuusorientoitunutta tietoa (Thierauf 2001, s. 66; Pirttimäki 2007, s. 59). Liiketoimintatiedon hallinta auttaa niin operatiivisen, taktisen kuin strategisen tason päätöksenteossa. Hyödyt jää pääosin operatiiviselle ja taktiselle tasolle jos tietovarasto sisältää ainoastaan sisäistä informaatiota. Operatiivisella tasolla riittää usein vakioraportit ja vakiomittarit, kun strategisella tasolla tarvitaan ad hoc -raportteja ja syvempää analyysiä, tiedonlouhintaa (engl. data mining).

Operatiiviset tietojärjestelmät sisältävät vain joitakin raportointiominaisuuksia, mutta tyypillisesti järjestelmät eivät keskustele keskenään. Lisäksi raportit ja raportointi hidastavat järjestelmän varsinaista käyttöä. (Hovi et al. 2009, s. 6.) Tämän takia liiketoimintatiedon hallinnassa tietoa kerätään tietovarastoon, jotta sitä voidaan paremmin analysoida ja hyödyntää. Raportointi operatiivisista järjestelmistä voi vaatia järjestelmätoimittajan tuen, tai vain järjestelmätoimittaja voi luoda raportteja. Lisäksi operatiivinen tietokanta ei yleensä ole suunniteltu raportointiin.

Operatiivisessa tietokannassa ja tietovarastossa on myös monia muita eroja (taulukko 2.1). Tietovarastossa tiedot tallennetaan aiheen perusteella, kun operatiivisessa järjestelmässä tapahtuman mukaan. Operatiivisissa järjestelmissä ei ole mielekästä eikä aina edes mahdollista tallentaa historiatietoa, joka on usein tärkeää tietovarastossa.

Tietovarasto on usein integroitu moniin järjestelmiin ja tietolähteisiin, kun operatiivista kantaa käyttää vain yksi järjestelmä.

Ominaisuus	Operatiivinen tietokanta	Tietovarasto
Tiedon perusta	Transaktioperustainen	Aihepohjainen
Tiedon pysyvyys	Nykyhetki	Nykyinen ja historiatieto
Pääsyoikeus	Luku ja kirjoitus	Luku
Tiedon tarkkuus	Yksityiskohtaista	Yhdistettyä
Integraation aste	Yksittäinen	Integroitu
Käytettävyys	Reaaliaikainen	Viiveellä

*Taulukko 2.1 Operatiivisen tietokannan ja tietovaraston erot (Winter 2001, s. 5; Rantanen 2007, s. 11)*

Tietovarasto päivittyy viiveellä, kun taas operatiiviseen kantaan tieto luodaan ensimmäisenä. Vaikka liiketoimintatiedon hallinnassa pyritään reaaliaikaisuuteen, se ei ole niin tärkeää kuin operatiivisissa järjestelmissä. Tietovarasto päivitetään usein joka yö, jolloin järjestelmän käyttö on vähäistä. Pienempiä kokonaisuuksia voidaan kuitenkin halutessaan päivittää useammin.

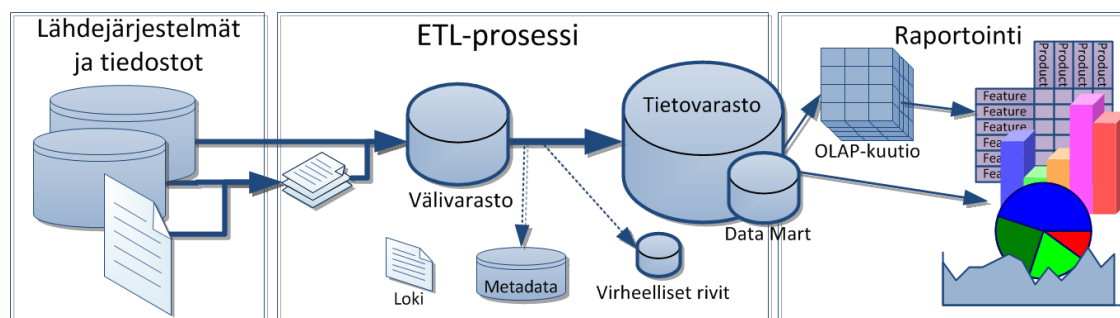
Hovin et al. (2009, ss. 80–81) mukaan liiketoiminnan hallinnan tavoitteita on nopeuttaa ja parantaa organisaatioiden kykyä tehdä päätöksiä, vastata käyttäjien tietotarpeisiin oikea-aikaisesti, tukea organisaation strategiaa ja tavoitteisiin pääsyä, parantaa käyttäjien omatoimisuutta tietotarpeen suhteen sekä vähentää kustannuksia ja parantaa operatiivista tehokkuutta. Tietovarastointi juontaa juurensa DSS-järjestelmistä (Decision Support System) eli päätöksenteon tukijärjestelmistä, joten myös liiketoimintatiedon hallinnan päätarkoitus on tukea päätöksentekoa.

Kustannussäästöjä tulee manuaalisen työn vähentymisen seurauksena ja mahdollisesti parempien päätösten ansiosta (ibid.). Liiketoimintatiedon hallinnalla avulla saadaan kustannussäästöjä, kun paremman tiedon avulla voidaan optimoida varastotasoja, ajoittaa ostoja paremmin sekä kohdistaa paremmin myyntiä ja markkinointia.

Liiketoimintatiedon hallinnalla pyritään kasvattamaan olemassa olevaa tietämystä yrityksen ja toimintaympäristön tilasta. Liiketoimintatiedon hallinta prosessiin syötteisiin kuuluu data, informaatio sekä aiempi tieto ja ymmärrys, tuloksena uusi tieto ja ymmärrys (Pirttimäki 2007, s. 73).

## 2.4. Tietovarastoinnin perusarkkitehtuuri

Tietovarastoinnin perusarkkitehtuuriin kuuluu raportointiin suunniteltu tietovarasto, kuten edellisessä alaluvussa mainittiin. Tietovarastoon ladataan dataa operatiivisista järjestelmistä ja mahdollisesti myös tiedostoista ja tietokannoista ETL-prosessilla (extract, transform, load) kuvan 2.4 mukaisesti.



Kuva 2.4 Tietovarastoinnin perusarkkitehtuuri (mukaillen lähteistä Chaudhuri & Dayal 1997, s. 2; Todman 2001, s. 253; Biere 2003, s. 13; Hovi et al. 2009, s. 14)

Tieto ladataan operatiivisista järjestelmistä yleensä sellaisenaan välivarastotietokantaan (engl. staging area, staging database), minkä jälkeen dataan tehdään vaaditut muutokset ja tarvittaessa jalostettu data tallennetaan vielä toiseen tietokantaan, mistä lopulta päivitetään tietovarastoon (Biere 2003, s. 13). Operatiivisen tietokannan tiedot voidaan tallentaa tiedostoihin, jotka siirretään tietovarastopalvelimelle toiseen ympäristöön, luettavaksi välivarastotietokantaan ja edelleen tietovarastoon. Toinen vaihtoehto on lukea suoraan operatiivisen tietojärjestelmän kannasta. (Hovi et al. 2009, ss. 50–52.)

ETL-prosessiin kuuluu kuusi vaihetta:

1. Valitaan tietolähteet, joista halutaan siirtää dataa ETL-prosessilla.
2. Muutetaan lähdedataa, kun tiedot on poimittu tietolähteistä. Tiedosta voidaan johtaa myös uutta tietoa.
3. Yhdistetään eri lähteistä koottu data.
4. Valitaan latauksen kohde.
5. Kohdistetaan lähdedatan kentät kohdedatan kenttiin.
6. Ladataan data kohdetietokantaan. (Trujillo & Luján-Mora 2003, s. 309.)

Ensimmäinen kohta on tiedon poimintaa tietolähteistä, eli ETL-prosessin ensimmäistä kohtaa, E-kirjainta (extract). Toinen ja kolmas on tiedon muuttamista, T-kirjainta (transform). Tässä vaiheessa tietoa suodatetaan, muunnetaan koodeja selkokieleiksi tai toiseksi koodiksi, muutetaan tietotyyppejä, generoidaan surrogaattivaimet (Trujillo & Luján-Mora 2003, s. 309). Tärkeää on yhdenmukainen koodisto tietovarastossa, esimerkiksi sukupuoli voi eri järjestelmissä olla joko M- tai N-kirjain, selkokielellä ”Mies” tai ”Nainen”, selkokielellä englanniksi, pienillä tai isoilla kirjaimilla kirjoitettuna, koodattuna numeroilla kuten 1 ja 0 tai boolean-tietotyyppinä (”true” tai

”false”). Tarkastukset voivat olla myös pakollisiksi määriteltyjen kenttien arvojen tarkastaminen, tuplarivien löytäminen, raja-arvojen määrittäminen (esim. ikä on väliltä 17–130), ja muototarkastukset, kuten postinumeron tai päivämäärän oikea muoto (Hovi et al. 2005, s 56). ETL-prosessi tarkastaa kaiken lähdedatan ja tallentaa virheelliset rivit virhetauluun (Todman 2001, s. 253). Lisäksi ETL-prosessin tulisi kirjoittaa lokia, mistä selviää ladattujen tietojen määrät, latausten kellonajat ja latausten epäonnistumiset. Nämä helpottavat huomattavasti valvontaa ja ylläpitoa.

Loput kohdat liittyvät ETL:n viimeiseen kirjaimeseen (load), lataamiseen tietovarastoon. Dimensionaalimallinnuksessa hitaasti muuttuvissa dimensioissa uudet tiedot lisätään sellaisenaan, olemassa olevat joko päivitetään, jos jokin kenttä on muuttunut, tai ei tehdä mitään, jos tieto on ajan tasalla. (Mundy et al. 2006, s. 230.) Menettely sopii useimpiin tapauksiin ja on yksinkertainen, mutta haittana vanhan tiedon menetyks. Mikäli halutaan säilyttää historiatietoja, muutosten kohdalla vanha rivi asetetaan passiiviseksi ja lisätään passivoitumispäivämäärä, sekä lisätään uusi rivi, mikä merkitään aktiiviseksi (Mundy et al. 2006, s. 230). Toinen, tosin hieman kömpelömpi vaihtoehto historiatiedon säilyttämiseen on lisätä tieto samalle riville, mutta uuteen sarakkeeseen (Hovi et al. 2009, s. 41). Lisäksi historiatiedot voidaan tallentaa erilliseen tietokannan tauluun, josta ne ovat saatavissa. Liiketoiminnassa syntyvät transaktiotiedot tallennetaan faktatauluihin, eikä niitä yleensä tarvitse edes päivittää. Dimensiot päivitetään ja historioidaan vain ne taulut ja kentät, mitä liiketoiminnan johto vaatii. (Hovi et al. 2009, s. 40.) Historiointi on aina työläämpää ja raportoinnin kannalta haastavampi toteuttaa.

Kimball esittää ETL-prosessin koostuvan 34 alisysteemistä. Ne voidaan ryhmitellä neljään pääluokkaan: latauksen eri tietolähteistä, datan puhdistamiseen vaadittavalle tasolle, tietojen lataaminen tietovarastoon ja ETL-ympäristön hallintaan. (Casters et al. 2010, ss. 114–126.) Alisysteemipaketti on hyvin kattava: se sisältää kaikki tarvittavat toimenpiteet ETL-prosessiin ja sen hallintaan liittyen aina virheiden käsittelystä, summaustauluihin ja rutiinin ajastuksesta toipumis- ja palauttamissuunnitelmaan. Kimballin esittämä malli ei ole ristiriidassa Trujillon ja Luján-Moran esittämään malliin, mutta on huomattavasti kattavampi. Kaikkia alisysteemejä ei kuitenkaan käytetä kuin laajoissa ja pitkälle kehitetyissä projekteissa.

Hovin et al. (2009, s. 59) mukaan ETL-prosessi voidaan suorittaa omalla palvelimellaan, tai samalla palvelimella kuin tietovarasto, jolloin se on ELT-prosessi. Vaikka tiedonlataukset ja muutokset toteutettaisiin samalla palvelimella missä tietovarasto sijaitsee, ELT on harhaanjohtava nimi, sillä muutokset (transform) täytyy tehdä ennen tietovarastoon latausta (load). Samalla palvelimella hoidetut datalataukset voivat hidastaa raportointia, varsinkin jos latauksia suoritetaan myös päiväsaikaan. Tieto voidaan tallentaa myös tietokomeroihin (engl. data mart) tietovaraston sijasta tai sen lisäksi. Tietoja voidaan laskea valmiiksi tietovarastoon, tai tallentaa summaukset

erillisiin tietokomeroihin tai kuutioihin (Hovi et al. 2009, s. 86). Valmiiksi lasketut summataulut nopeuttavat tiedon käsittelyä, kun tietovarastossa on paljon rivejä. Esimerkiksi ajan suhteen summaus on hyvin yleistä, tietovarastossa voi olla esimerkiksi myyntirivit kuukausitasolla ja kvartaalitasolla, joista voidaan tarvittaessa porautua tarkemmalle tasolle. (Todman 2001, ss. 50–51.)

Tietovarastoa hyödynnetään raportoinnin ja kyselyjen avulla. Tyypilliseen arkkitehtuuriin kuuluu analysointi, kyselyt, raportointi ja tiedonlouhinta tietovarastosta, kuutioista ja tietokomeroista (Chaudhuri & Dayal 1997, s. 2). OLAP-kuutiot (Online Analytical Processing) tarjoavat interaktiivista analysointia eri dimensioilla (aika, tuote, maantieteellinen alue) ja eri tasoilla (vuosi, kvartaali, kuukausi). OLAP keskittyy analysoimaan ja tutkimaan dataa ”miksi”-kysymyksillä, kun kyselyt ja raportointi keskittyvät tiedon saatavuuteen ja ”mitä”-kysymykseen. (Howson 2007, s. 40.) Kuutioissa voidaan tarkastella valittujen dimensioiden suhteen mittareita, lisätä aggregointitasoa, porautua, tarkastella yhtä tai useampaa hierarkiaa pitkin, viipaloida, projisoida ja tehdä pivot-*taulukko* (Chaudhuri & Dayal 1997, s. 2). Parikan (2009, s. 17) mukaan dimensioiden lisäksi, tulisi tietokannan taulussa olla vähintään yksi numeerinen sarake. Toisinaan voidaan kuitenkin raportoida rivien lukumäärää, jolloin numeerinen kenttä ei ole välttämätön. Esimerkiksi asiakaskäyntien lukumäärä voi olla kiinnostava mittari, eikä tietokannan taulussa tarvitse olla numeerista arvoa. Kuutioiden lisäksi tietovarastoa hyödynnetään parametrisoiduilla vakioraporteilla, mittaristoilla (engl. *dashboard*) ja tiedonlouhinnalla (Hovi et al. 2009, ss. 90, 95–96). Mittaristot voivat hakea tietoa suoraan tietovarastosta tai hyödyntää kuutioita.



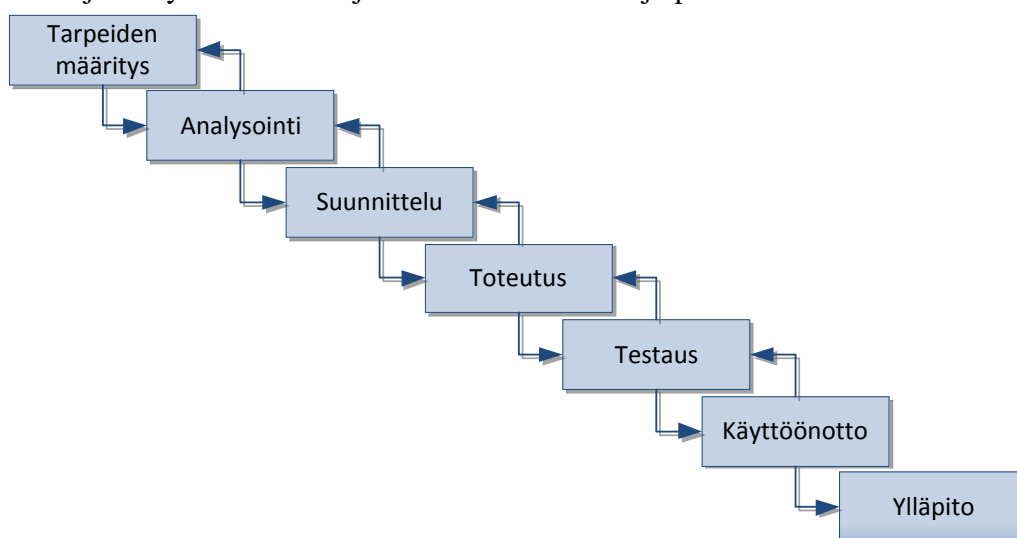
### 3. LIKETOIMINTATIEDON HALLINTAPROJEKTI JA KETTERÄT MENETELMÄT

”On paras tehdä asiat järjestelmällisesti, koska olemme vain ihmisiä ja epäjärjestys on pahin vihollisemme”, lausui Hesoid, 8. vuosisadan tiedemies (Forsberg et al. 2003, s. 3). Jotta myös liiketoimintatiedon hallinnan projektit onnistuisivat, tulee projektin olla hyvin suunniteltu ja hallittu.

Projektinhallintaan, ja erityisesti ohjelmistotuotantoon on kehitetty hyvin paljon erilaisia menetelmiä, mutta ne kaikki voidaan jakaa kahteen pääkategoriaan, vesiputousmalleihin ja nopean sovelluskehityksen lähestymistapaan (Todman 2001, s. 230). Tässä luvussa esitellään ensin perinteinen vesiputousmalli, minkä jälkeen käydään läpi muita projekti- ja prosessimalleja. Pääosin ne on suunniteltu ohjelmistokehitykseen, mutta useimpia voidaan soveltaa myös tietovarastoinnin projekteissa. Lopuksi käsitellään ohjelmistokehityksestä tuttua aihetta, ketteriä menetelmiä.

#### 3.1. Vesiputousmalli

Vesiputousmalli (kuva 3.1) on vanha ja paljon käytetty erityisesti ohjelmistotuotannossa. Sen kehitti tohtori Winston W. Royce jo vuonna 1970 suurien ohjelmistokehityksen hallintaan (Royce 1970, s. 1; Forsberg et al. 2003, s. 22). Vesiputousmallin ideana on toteuttaa ylävasemmalla oleva ensimmäinen askel kokonaan ja siirtyä seuraavaan ja näin edetä erillisinä ja peräkkäisinä vaiheina.



Kuva 3.1 Perinteinen vesiputousmalli (mukaihen lähteistä Royce 1970, s. 2; Todman 2001, s. 231; Moss & Atre 2003, s. 7; Forsberg 2003, s. 22)

Vesiputousmallia on käytetty hyvin paljon sen selkeyden ja yksinkertaisuuden vuoksi. Projektin vaiheet ja niiden määrä voidaan vapaasti valita, joten tämä malli on muunneltavissa projektin tarpeisiin. Projektimallin vaiheet ovat kuitenkin hyvin yleispäteviä, joten tämänkin takia sopii erilaisiin projekteihin.

Useimmiten ensimmäisinä vaiheina ovat tarpeiden määrittäminen ja vaatimukset. Alkuperäisen mallin mukaan alussa ovat järjestelmävaatimukset ja ohjelmistovaatimukset (Royce 1970, s. 2). BI-projekteihin sopisi paremmin liiketoiminnan tarpeiden määrittäminen, projektin suunnittelu ja toiminnalliset vaatimukset kuten Moss ja Atre (2003, s. 7) esittää.

Projektimallin seuraavina vaiheina tulee analysointi ja suunnittelu, kuten Roycen (1970, s. 2) alkuperäisessä mallissa, tai alustava ja yksityiskohtainen suunnittelu, kuten Forsberg et al. (2003, s. 22) asian näkevät. Liiketoimintatiedon hallinnan projekteissa analysointiin voisi kuulua esimerkiksi saatavilla olevan datan analysointi sekä erilaisten vaihtoehtojen vertailu. Tämän jälkeen tulee toteutus, ohjelmistoprojekteissa koodaus ja vianetsintä. Alkuperäisen mallin mukaan lopussa on testaus ja käyttö, mutta Moss ja Atre (2003, s. 7) esittää viimeisiksi vaiheiksi edellisten lisäksi myös ylläpidon. Tätä ei ole syytä unohtaa ohjelmistoprojekteissa, eikä myöskään BI-projekteissa.

Monet ovat tarkastelleet Roycen esittämää vesiputousmallin esikuvaa virheellisesti yhdensuuntaisena mallina. Todellisuudessa hänen suosittama lähestymistapa on hieman erilainen, kuin nykypäivän vesiputouskäsitteen malli tiukasta järjestyksestä vaatimusten, analyysin ja kehityksen vaiheista. (Larman & Basili 2003, s. 48.) Royce esitteli myös toiseen suuntaan kulkevat nuolet, ja lisäsi että tarvittaessa joudutaan valitettavasti palaamaan myös muihin vaiheisiin, ei ainoastaan edelliseen tai seuraavaan (Royce 1970, s. 3). Lisäksi hän suositteli tekemään prosessin kahteen kertaan. Jos esimerkiksi ohjelmaa ollaan kehittämässä ensimmäistä kertaa, kriittisten suunnittelu ja toteutusalueiden osalta asiakkaalle tarjottava sovellus on jo toinen versio. (Royce 1970, s. 7; Larman & Basili 2003, s. 48.) Tällöin ensimmäinen versio on vasta prototyyppi ja toinen varsinainen tuotos.

On tärkeää ymmärtää, että vesiputousmalli ei ole yhdensuuntainen, eikä kaikkia vaiheita tarvitse suorittaa järjestyksessä. Edellisiin vaiheisiin voidaan aina palata, jos niissä havaitaan puutteita. Esimerkiksi testatessa löytyy varmasti virheitä, jolloin joudutaan koodaamaan uudestaan eli palaamaan edelliseen tehtävään, joskus jopa suunnittelemaan uudestaan. Taaksepäin voidaan siirtyä myös hyppäämällä joidenkin vaiheiden ylitse, esimerkiksi testauksesta voidaan siirtyä suoraan suunnitteluun, ja edelleen vaatimusten määrittelyyn (Royce 1970, s. 3). Aina ei kuitenkaan voi siirtyä taaksepäin, esimerkiksi ylläpidosta ei varmastikaan mennä takaisin käyttöönnottoon. Käyttöönnotosta voidaan

vielä poikkeustapauksessa siirtyä takaisin testaukseen ja ottaa sovellus uudestaan käyttöön myöhemmin.

Vesiputousmalli on alun perin suunniteltu suuriin ohjelmistoprojekteihin. Ohjelmistoprojektit ovat kuitenkin neljänkymmenen vuoden aikana muuttuneet hyvin erilaisiksi ja kasvaneet entistä suuremmiksi. Tänä päivänä vesiputousmalli ei sovi monimutkaisiin ja suuria riskejä sisältäviin projekteihin, sillä se synnyttää epärealistisia kustannus- ja aikatauluarvioita. Tämä johtuu siitä, että usein on tarve aloittaa ohjelmiston suunnitteluja koodaus aiemmin, jotta varmistetaan että vaatimukset on ymmärretty oikein ja projekti on toteuttamiskelpoinen. (Forsberg et al. 2003, s. 23.) Toisinaan on siis tarve toteuttaa eri projektin vaiheita rinnakkain. Tämä voi myös nopeuttaa projektin etenemistä huomattavasti.

Vesiputousmallin heikkouksia on myös se, että projektia suunniteltaessa ei useinkaan tiedetä kovin tarkasti minkälaisia ongelmia ja haasteita koodatessa syntyy. Projektiin, johon tulee jatkuvasti muutoksia, ei tämä projektimalli ole kovinkaan toimiva. Vesiputousmalli toimii siis vain hyvin stabiilissa ympäristössä, silloin kun projektiin ei tule muutoksia ja lisävaatimuksia kun sitä on alettu suunnittelemaan ja toteuttamaan.

Toisin kuin erillisissä järjestelmissä, dynaamisia ja integroituneita liiketoimintatiedon hallintaympäristöjä ei ole mahdollista rakentaa yhdessä suuressa rykelmässä. Dataa ja toiminnallisuutta täytyy käsitellä iteratiivisissa jaksoissa, jolloin jokainen kehitysaskel toimii herätteenä uudelle vaatimukselle seuraavassa iteratiivisessa jaksossa. (Moss & Atre 2003, ss. 7-8.) Vesiputousmallin heikkouksien seurauksena on kehitetty paljon hyvinkin erilaisia projektimalleja, niin syklisiä kuin sellaisia malleja, joissa on rinnakkain eteneviä vaiheita.

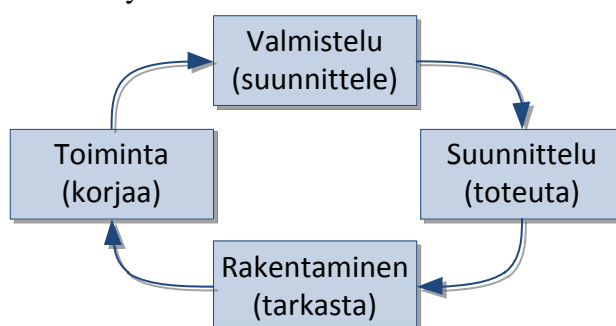
### **3.2. Syklisiä malleja**

Boehm (1988, s. 64) esittää spiraalisen ohjelmistoprosessin mallin, jossa tuotetaan aluksi prototyyppjä, kunnes viimeisellä kierroksella syntyy toiminnallinen ohjelmisto. Keskipisteestä alkava ja spiraalinmuotoisesti myötäpäivään kulkeva viiva kuvaa aikaa ja tehtävät ovat säteittäin. Kuvan ensimmäisellä neljänneksellä määritellään tavoitteet, selvitetään vaihtoehdot ja rajoitukset, toisella neljänneksellä toteutetaan riskianalyysiä, kolmannella tehdään vaatimukset, suunnitelmat, kehitetään ja toteutetaan sekä viimeisellä neljänneksellä suunnitellaan seuraavaa vaihetta. Spiraalimallin säteittäinen aikajana on epäjohdonmukainen, kun perinteisesti aika etenee vasemmalta oikealle ja malli hämärtaa kehittämisen hallitsemiseen tarvittavia tarkistuksia (Forsberg et al. 2003, s. 24). Lisäksi malli korostaa riskienhallintaa voimakkaasti, jota tulisi suorittaa jokaisella kierroksella. Kaikissa projekteissa riskienhallinta ei ole kovin keskeinen asia, eikä sitä muutenkaan tulisi pitää erillisenä asiana, vaan toteuttaa jatkuvana, rinnakkain projektin kanssa.

Boehmin spiraalimalli on myös kehitetty ohjelmistotuotantoon ja sitä on käytetty hyvin paljon ohjelmistojen kehittämiprojekteissa (Forsberg et al. 2003, s. 24). Mallilla on pyritty ratkaisemaan vesiputousmallin heikkouksia. Mallin kehittäjän mukaan tuottavuus on puolet parempi kaikissa projekteissa kyseistä mallia käytettäessä (Boehm 1988, s. 69). Spiraalimalli on kuitenkin suunniteltu ohjelmistokehitykseen, joten sen toimivuudesta liiketoimintatiedon hallinnan projekteissa ei ole tietoa.

Moss ja Atre (2003, s. 8) esittävät renkaanmuotoisen, syklisen mallin, johon on liitetty lähes samat vaiheet kuin vesiputousmallissa. Tässä mallissa korostuu liiketoimintatiedon hallintaprojektin iteratiivinen luonne ja BI-sovellusten ero erillisjärjestelmiin (ibid.). Kyseinen hieman triviaali ratkaisu ei ole kovin toimiva, sillä mallin mukaan jokaisella kierroksella tulisi pohtia muun muassa liiketoiminnan mahdollisuuksia, päätöksenteon tuen strategiaa, projektin suunnittelua ja vaatimuksia. Ainakin malli on melko raskas, jos yhdellä kierroksella kehitetään pienehköjä asioita. Tällöin joutuu turhan usein pohtimaan vaatimuksia yleisemmälläkin tasolla.

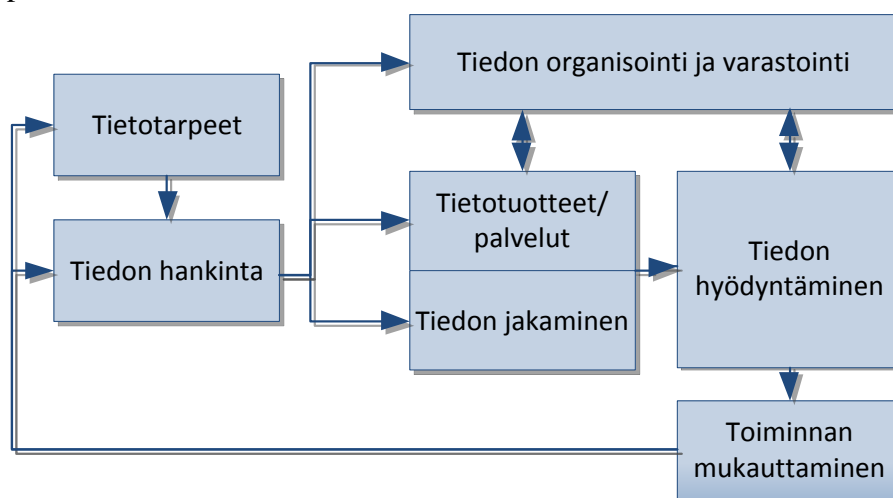
Brady (2004, s. 3) esittää Hadden-Kelly-mallin (kuva 3.2), jonka vaiheet koostuvat valmistelusta (preparing), suunnittelusta (planning), rakentamisesta (building) ja toiminnasta (operating). Valmistelu koostuu lähdejärjestelmien tiedon laadun parantamisesta, teknisen ympäristön suunnittelusta, resurssien kohdentamisesta ja toimintasuunnitelman laatimisesta. Suunnitelma sisältää mm. infrastruktuurin määrittelyn ja yksityiskohtaisen suunnitelman tietovaraston rakentamista varten. Rakentamisvaiheessa toteutetaan tiedon siirrot lähdejärjestelmistä tietovarastoon tai tietokomeroon (Data Mart) siltä osalta, joka vastaa liiketoiminnan tarpeisiin ja määrittelyihin. Lisäksi rakennetaan tarvittaessa datan hyödyntämissovelluksia, raportointivälineitä. Viimeisenä kohtana on koulutus, tiedon hyödyntäminen ja käyttö sekä ylläpito. (Brady 2004, s. 3.) Hadden-Kelly-malli on yksinkertaisuudessaan melko tehokas. Yhdellä kierroksella voi tehdä yhden kokonaisuuden valmiiksi, ja rakentamisvaihe voidaan myös tarvittaessa toteuttaa iteratiivisesti.



*Kuva 3.2 Hadden-Kelly-syklimalli yhdistettynä PDCA-sykliin, jälkimmäisen vaiheet sulussa (mukaillen lähteestä Brady 2004, s. 3)*

Tämä syklinen malli on hyvin samankaltainen, kuin perinteinen PDCA-sykli (plan, do, check, act). PDCA-syklissä (kuva 3.2 sulkeissa olevat) on ensin suunnittelu, sitten toteutus, tarkastaminen ja lopuksi toimintaa virheiden korjaamiseksi, minkä jälkeen alkaa uusi kierros. Samoin kuin Hadden-Kelly-mallissa, myös PDCA-syklissä voidaan toteuttaa pienempiä tai suurempia kokonaisuuksia kerralla, ja yksi mallin vaihe voidaan avata uuteen malliin, syklisteen tai vesiputoukseen.

Choon informaation hallinnan prosessimalli (kuva 3.3) on myös syklinen malli. Sitä voidaan tarkastella myös liiketoimintatiedon hallintaprojektin näkökulmasta, vaikka se on suunniteltu jatkuvaksi prosessimalliksi tiedonhallintaan yleisesti. Prosessi alkaa oikealta alhaalta, toiminnan mukauttamisesta, kun tietoa syntyy organisaation toiminnassa (Choo 2002, s. 24). Seuraavassa vaiheessa pohditaan tietotarpeita, minkä jälkeen tietoa hankitaan ulkoisista ja sisäisistä lähteistä. Kerättyä tietoa organisoidaan ja varastoidaan esimerkiksi tietovarastoon, tarjotaan tietotuotteita ja palveluita kuten raportteja ja mittaristoja sekä jaetaan sitä tarvitseville. Tämä ei ole passiivinen uudelleenpakkaamisvaihe tulevalle tiedolle, vaan lisäarvoa tuo tiedon laadun parantaminen ja sovitetaan tieto tietotarpeeseen käyttäjälle (Choo 2002, s. 25). Liiketoimintatiedon hallinnassa tämä tarkoittaa, että tiedon laatua on varmistettu jo tietovarastossa ja tietotuotteet, eli raportit ja mittaristot tehdään käyttäjän tarpeisiin. Lopuksi tietoa mallin mukaan tietoa hyödynnetään päätöksenteossa ja mukautetaan toimintaa päätösten mukaisesti.

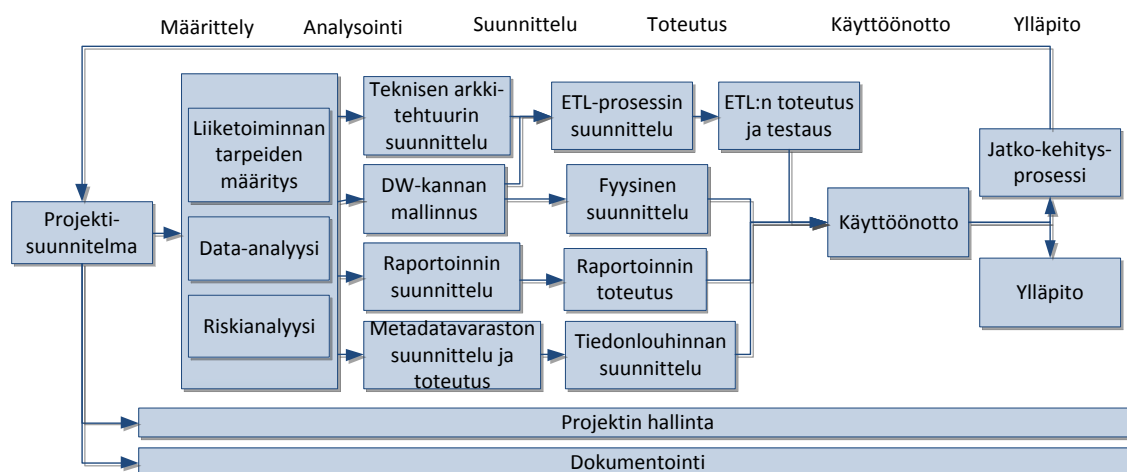


Kuva 3.3 Informaation hallinnan prosessimalli (mukaillen lähteestä Choo 2002, s. 24)

Choon informaation hallinnan syklimalli voisi toimia liiketoimintatiedon hallintaprojektissa siten, että projektin alussa mietitään tietotarpeet, ja kun tarvittava tieto on saatavilla, toteutetaan tietovarastoa ja samanaikaisesti suunnitellaan raportointia. Heti, kun tietoa alkaa olla saatavilla, sitä hyödynnetään ja käytetään päätöksen teon tuessa. Tällaisen toiminnan seurauksena implementointivaihetta ei olisi ollenkaan, vaan tietovarastoa ja raportointia hyödynnetään aina sen kehitettyä ja niitä kehitetään jatkuvasti lisää.

### 3.3. Tietovarastointiprojektin vaiheet

Tietovarastointiprojekteista löytyy lähes aina tietyt vaiheet. Projekteilla on elinkaari (kuva 3.4), joka alkaa projektisuunnitelmasta ja päättyy käyttöönottoon sekä ylläpitoon ja/tai jatkokehitysprosessiin. Moss ja Atre (2003, s. 15), Mundy et al. (2006, s. 4) sekä Hovi et al. (2009, s. 131) esittävät oman näkemyksensä tietovarastointiprojektin vaiheista. Kaikki ovat kehittäneet omanlaisensa mallin, mutta kaikista löytyy samoja asioita. Kaikissa malleissa on löydettävissä vesiputousmallin mukaiset vaiheet määrittelystä ja analysoinnista aina käyttöönottoon ja ylläpitoon asti. Selvänä erona vesiputousmalliin on rinnakkain etenevät vaiheet. Projektisuunnitelman ja liiketoiminnan tarpeiden määrittelyn jälkeen voidaan toteuttaa useita asioita rinnakkain käyttöönottoon asti. Mundy et al. (2006, s. 4) esittävät myös projektinhallinnan kulkevan kokoajan projektin rinnalla ja Hovi et al. (2009, s. 131) lisäävät vielä dokumentoinnin.



Kuva 3.4 Tietovarastointiprojektin elinkaaren vaiheet (mukaillen lähteistä Moss & Atre 2003, s. 15; Mundy et al. 2006, s. 4; Hovi et al. 2009, s. 131)

Kustakin mallista löytyy projektisuunnitelma, mutta Moss ja Atre (2003, s. 15) lisäävät liiketaloudellisen arvioinnin ja yrityksen infrastruktuurin arvioinnin alkuun, ennen projektisuunnitelmaa. Ne voitaisiin toisaalta sisällyttää myös muihin vaiheisiin, kuten itse projektisuunnitelmaan. Toisaalta tietovarastointiprojekti voidaan nähdä alkavaksi projektisuunnitelmasta vasta, kun se on perusteltu kannattavaksi ja hyväksytty. Lisäksi Mossin ja Atren esittämässä mallissa on selkeämmin esitetty riippuvuussuhteet, kuten tietokannan mallinnuksen jälkeen voidaan toteuttaa ETL:n suunnittelu, raportoinnin toteutus ja data mining. Toisaalta ETL-prosessia voidaan alkaa suunnittelemaan, vaikka tietovaraston taulut eivät olisi valmiita.

Projektisuunnitelman jälkeen kaikissa malleissa on liiketoiminnan tarpeiden määrittely. Hovi et al. (2009, s. 131) sisällyttävät siihen riskianalyysin ja tietojen laatuanalyysin. Tarpeet voidaan määrittellä asettamalla projektille tavoitteet. Liiketoiminnan näkökulmasta asetetut, hyvin määritellyt ja realistiset tavoitteet luovat pohjan

onnistuneelle projektille (Hovi et al. 2009, s. 144). Data-analyysi, prototyypit ja metadata-analyysit voidaan sisällyttää projektivaatimusten määrittelyn kanssa liiketoiminta-analyysivaiheeseen (Moss & Atre 2003, p 15). Mundy et al. (2006, s. 8) esittävät alaprosessin tarpeiden määrittämisestä, johon kuuluu valmistelu, liiketoiminnan ja tietotekniikan haastattelut sekä haastattelutulosten kirjaus, liiketoimintaprosessien tunnistus, dimensioiden määrittely, ylimmän johdon asettama priorisointi ja yleisvaatimusmäärittelyasiakirjan laatiminen. Liiketoiminnan tarpeet on määriteltävä hyvin, mutta tarpeita määriteltäessä tulee tietää, minkälaista dataa on saatavilla ja minkälainen tietovarastointiratkaisu on mahdollista toteuttaa. Nämä asiat tulevat hyvin ilmi kaikissa malleissa vaikkakin hieman eri termeillä ja eri näkökulmasta.

Suunnittelu voidaan aloittaa heti tarpeiden määrittämisen jälkeen. Suunnitteluun kuuluu mm. teknisen arkkitehtuurin suunnittelu, tietokannan mallinnus ja fyysinen suunnittelu, ETL-prosessin suunnittelu ja raportoinnin sekä mittaristojen suunnittelu. Tietokannan mallinnukseen sisältyy taulujen määrittely ja niiden suhteet, kun fyysiseen suunnitteluun kuuluu muun muassa kenttien nimeäminen, tietotyyppien ja surrogaattiavainten määrittelyt, indeksit ja faktataulujen ositussuunnitelma (Mundy et al. 2006, s. 157). Suunnittelun jälkeen tiedot poimitaan operatiivisista järjestelmistä, muokataan ja jalostetaan sekä tallennetaan tietovarastoon ETL-prosessin mukaisesti sekä toteutetaan raportointi. Moss ja Atre (2003, s. 14) lisäävät myös tiedonlouhinnan toteuttamisen, sillä heidän mielestään tietovarastointia ei hyödynnetä täysissä määrin, vaan usein käytetään vain ennalta suunniteltuja vakioraportteja, kun todellinen hyöty tulisi organisaation piilossa olevan tiedon löytämisestä. Toteutuksen ja testauksen ollessa valmiita, voidaan järjestelmä ottaa käyttöön, ja siirtyä ylläpitoon. Lisäksi jatkokehitysprosessia voidaan alkaa tarvittaessa suunnittelemaan.

Semanttiseen metatietoon (kuten nimi, aihe, tiivistelmä), rakennemetatietoon (tiedostoformaatti, koodaus, pituus, kieli) ja kontekstuaaliseen metatietoon (luomisaika, päivitysaika, suhteet muihin tietoihin) jaettava metadata liittyy tietovarantoihin, kuten tietovarastoon (Lehtinen et al. 2005, ss. 4, 8-9). Moss ja Atre (2003, s. 15) korostavat metatiedon merkitystä piirtämällä malliin vaiheet metadatan analysoinnista, suunnittelusta ja kehityksestä. Toisaalta metadataa kehittyvät muiden työvaiheiden ohessa, joten sitä ei välttämättä tarvitse suorittaa erillisenä osana. Metatietojen riittävyys on silti tärkeä suunnitella, jotta sitä voidaan tarvittaessa käyttää hyväksi. Metadata parantaa myös raporttien luotettavuutta (Hovi et al. 2009, s. 145). Metadataa voidaan käyttää hyväksi myös ETL-prosessin toteutuksessa kohdistettaessa lähdedatan kenttiä tietovaraston kenttiin.

Hovi et al. (2009, s. 175) esittävät mallissaan yhtenä ”putkena” tarpeiden määrittämisen ja käyttöönoton välille prototyypin suunnittelu- ja rakentamisvaihetta. Ohjelmistoihin liittyvä prototyyppi on kuitenkin epätarkka, useita merkityksiä sisältävä termi, joka tarkemmin ottaen on yleensä ohjelmiston vaatimusten malli (Forsberg et al. 2003, s. 55).

Hovin et al. (2009, s. 174) mukaan prototyypit voidaan jakaa kolmeen tyyppiin kevyeen prototyyppiin, proof-of-concept-prototyyppiin ja demoprototyyppiin. Näistä erityisesti kevyt prototyyppi ja demoprototyyppi ovat selkeästi malleja, joilla voidaan viestittää projektiryhmälle tai sidosryhmille ratkaisua, käyttöliittymää tai muuta sellaista. Proof-of-conceptia voidaan pitää prototyyppiä, joilla varmistetaan ratkaisun toimivuus. Näitä prototyyppisiä ja malleja on hyvä tehdä projektin eri vaiheissa ennen varsinaista tekemistä, niistä on varmasti hyötyä toteutuksessa. Kuitenkaan yhtä laajaa prototyyppiä, joka kestää määrittelystä käyttöönottoon, ei ole kovin mielekästä tehdä.

Tietovarastointiprojektissa voidaan käyttää elinkaarimallin kanssa spiraalimallia, sekä lisäksi voidaan toteuttaa inkrementaalisisissa osissa (Hovi et al. 2009, s. 131). Tietovarastointiprojektin elinkaaren malli toimisi yleiskuvana ja pienemmät kokonaisuudet voidaan toteuttaa valitun spiraalimallin mukaisesti. Lisäksi voidaan toteutus hoitaa inkrementaalisisissa osissa, esimerkiksi tietovarastoon kerätään dataa yhdestä järjestelmästä, ja vaikka yksi dimensio kerrallaan. Raportointia voidaan alkaa tekemään jo, kun tietovarastosta löytyy esimerkiksi asiakastiedot, myyntitiedot ja tuotetiedot. Tämän jälkeen tietovarastoon voidaan kerätä muutkin halutut tiedot, varastosaldoista asiakaskäynteihin ja ostotiedoista budjettilukuihin. Lopulta voidaan vielä kehittää erikseen historiointi valittujen tietojen osalta.

### 3.4. Ketterät menetelmät

Ketterät menetelmät ovat joukko ohjelmistokehitykseen suunniteltuja menetelmiä, jotka perustuvat lähinnä iteratiiviseen ja inkrementaaliseen kehitykseen. Ohjelmistokehityksessä ketterä ”liike” lähti kunnolla käyntiin vuonna 2001, kun ryhmä ammattilaisia ja ketterän kehityksen puolestapuhujia kokoontui yhteen, tunnistamaan yhteisiä näkökohtia, yhdistelemään vanhoja ja uusia ajatuksia. Tämän seurauksena syntyi ketterän ohjelmistokehityksen manifesti ja termi ”ketterä” (engl. agile) valittiin yhdistämään menetelmiä ja tekniikoita. (Salo 2006, ss. 24–25.) Manifestissa määriteltiin arvostamaan

- Yksilöitä ja vuorovaikutusta enemmän kuin prosesseja ja työkaluja
- Toimivaa sovellusta enemmän kuin kokonaisvaltaista dokumentaatiota
- Asiakasyhteistyötä enemmän kuin sopimusneuvotteluita
- Muutokseen reagoimista enemmän kuin suunnitelman noudattamista

Ja vaikka oikeanpuoleisilla asioilla on arvoa, vasemmanpuoleisia tulisi arvostaa enemmän. (Agile Alliance 2001.) Lisäksi manifestissa luotiin kaksitoista tärkeää periaatetta, joita tulisi noudattaa ketterissä menetelmissä. Ne ovat seuraavat:

- 1) Tärkein tehtävä on toimittaa asiakkaalle varhaisessa vaiheessa ja jatkuvasti arvoa tuottavaa ohjelmistoa
- 2) Muutokset vaatimuksiin ovat tervetulleita, myös myöhäisessä kehitysvaiheessa. Ketterillä prosesseilla muutokset saadaan muutettua asiakkaan kilpailueduksi.



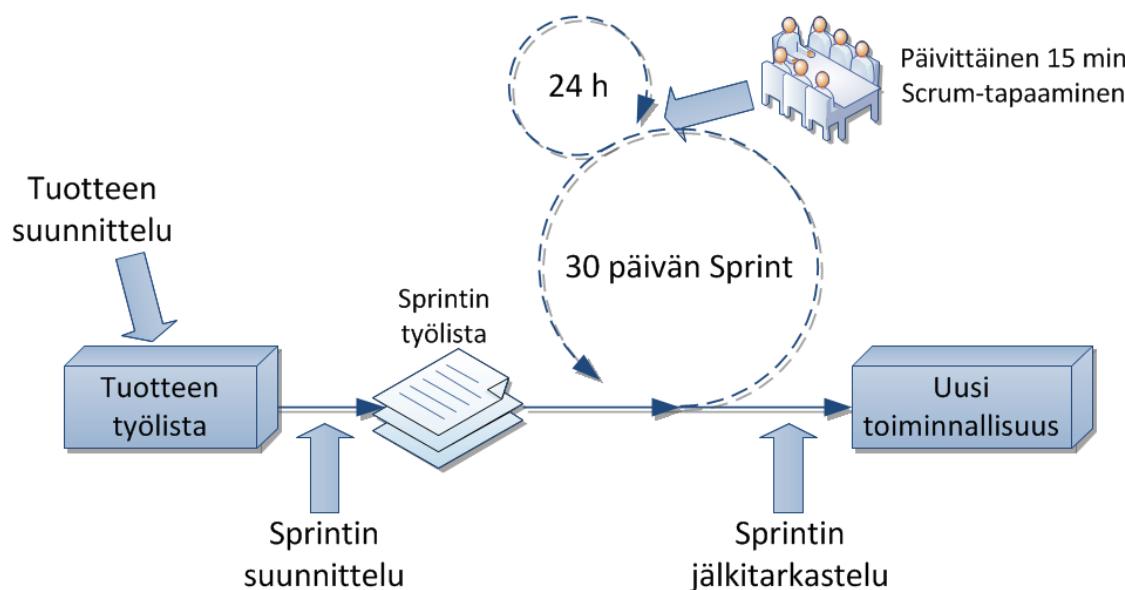
- 3) Toimitetaan toimiva ohjelmisto asiakkaalle mahdollisimman usein, muutamasta viikosta muutamaan kuukauteen.
- 4) Liiketoimintahenkilöiden ja kehittäjien on työskenneltävä yhdessä päivittäin koko projektin ajan.
- 5) Rakenna projektia motivoituneiden henkilöiden varaan. Rakenna heille tukea antava ympäristö ja luota heihin, että saavat työnsä tehtyä.
- 6) Tehokkain tapa välittää tietoa keskustelu kasvotusten.
- 7) Toimiva ohjelmisto on tärkein mittari prosessin edistymiselle.
- 8) Ketterät prosessit edistävät kestäväää ohjelmistokehitystä. Kehittäjien ja käyttäjien tulisi pitää kehityksessä tasainen vauhti jatkuvasti.
- 9) Jatkuva huomio tekniseen laatuun ja hyvään suunnitteluun edistää ketteryyttä.
- 10) Yksinkertaisuus – keskeistä on minimoida turhan työn määrä.
- 11) Parhaat arkkitehtuurit, vaatimukset ja suunnitelmat syntyvät itse-organisoiduista tiimeistä.
- 12) Säännöllisin väliajoin tiimin tulee pohtia, miten toimia tehokkaammin, ja sen jälkeen hienosäätää toimintaa sen mukaan. (Agile Alliance 2001.)

Näitä arvoja ja periaatteita tulee kaikkien eri ketterien menetelmien tukea. Periaatteiden ytimeen kuuluu itse-organisoituva tiimi, joka työskentelee joustavasti, tehokkaasti ja aina asiakkaan eduksi. Ketterän kehityksen menetelmät, kuten Scrum, Extreme Programming (XP) ja Adaptive Software Development (ASD) lupaavat parempaa asiakastyytyväisyyttä, alemmaa virhetasoa, nopeampaa kehitysaikaa ja ratkaisuja nopeasti muuttuviin vaatimuksiin. Suunnitteluohjautuvat lähestymistavat, kuten vesiputousmalli, puolestaan tarjoavat parempaa ennustettavuutta, vakautta ja korkeampaa varmuutta. (Boehm & Turner 2004, s. 718.) Edellä mainittujen lisäksi pääosin 1990-luvulla kehitettyjä ketteriä menetelmiä ovat myös mm. DSDM (Dynamic Systems Development Method), FDD (Feature-Driven Development), eXtreme testing ja Crystal-menetelmät, vaikkakin ketterien menetelmien ominaisuuksia löytyy jo Lean-ajattelusta (Salo 2006, s. 24).

Scrum on yksi käytetyimmistä ketteristä menetelmistä. Ensimmäisenä idean Scrum-kehityksestä kuvasivat Takeuchi ja Nonaka (1986) jo yli kaksi vuosikymmentä sitten (Abrahamsson et al. 2002, s. 27). He kuvasivat uudenlaisen, kokonaisvaltaisen lähestymistavan tuotekehitykseen, missä ryhmä ammattilaisia toteuttaa kehitysprosessin alusta loppuun (Takeuchi & Nonaka 1986, ss. 137–138). Tuotekehitystä on toteutettu perinteisesti liukuhihna-tyylisesti, jossa vastuu siirtyy tuotekehitysryhmältä toiselle. Tällaisessa tapauksessa toimisi parhaiten vesiputousmalli. Takeuchi ja Nonaka (1986, ss. 138–139) esittävät mallissaan lisäksi limittäiset toiminnot, jossa esimerkiksi suunnittelu voidaan aloittaa ennen kuin edellinen vaihe, toteutettavuustestaus on saatu kokonaan valmiiksi.

Scrumin varsinaisia kehittäjinä pidetään kuitenkin Sutherlandia ja Schwaberiä, jotka suunnittelivat 30-päiväisen iteraatioihin (ks. kuva 3.5) perustuvan mallin (Larman &

Basili 2003, s. 54). Näin Scrum on siis iteratiivinen malli, joka alkaa tuotteen suunnittelulla ja sen työlistan laatimisella. Tuotteen työlistan pohjalta tiimin jäsenet valitsevat seuraavaan Sprinttiin, eli iteraatiokierrokseen tehtävät, josta koostuu työlista. Sprintin aikana pidetään joka päivä lyhyt, noin 15 minuutin palaveri, missä käydään läpi mitä on tehty, mitä aiotaan tehdä ja mitä esteitä mahdollisesti on kohdattu (Scrum Alliance 2005, s. 48–49). Näin koko tiimi tietää mitä kukin jäsen on tekemässä ja minkälaisia haasteita on kohdattu. Kun esteet ja haasteet tuodaan ilmi välittömästi, voi ratkaisukin syntyä nopeammin.



Kuva 3.5 Scrum-prosessi (Scrum Alliance 2005, s. 5)

Scrum-menetelmään kuuluu erilaisia rooleja. Scrum-mestari (engl. Scrum Master) toimii projektipäällikkönä, joka valitsee asiakasnäkökulman omaavan ja tuotteen työlistasta vastaavan tuotteen omistajan (engl. Product Owner). Loput kuuluvat itseohjautuvaan Scrum-tiimiin. (Abrahamsson et al. 2002, ss. 30–31.) Tiimin suositeltava koko on noin 5-10 henkeä. He voivat menetelmän mukaan itse valita sprintin työlistalta tehtävät, jotka aikovat tehdä sen aikana. Toisin kuin perinteisissä projektimalleissa, Scrum-mestarilla ei ole määräysvaltaa tiimiin, vaan toimii enimmäkseen esteiden poistajana ja dokumentoijana.

Tuotteesta suunnitellaan työlista (engl. product backlog), joka sisältää toteutettavaksi aiotut vaatimukset. Seuraavaan julkaisuun valitaan tuotteen työlistalta vaatimukset julkaisun työlistaan (engl. release backlog). Tästä ryhmä valitsee jokaisen iteraatiokierroksen alussa vaatimukset toteutusvaiheen, eli sprintin työlistaan (engl. sprint backlog). Iteraatiokierroksen, eli sprintin lopuksi tiimi esittelee valmista tuotetta tuotteen omistajalle. Sen tulisi olla toteutettu, testattu ja dokumentoitu sekä käyttöön otettava. Lisäksi sprintin jälkitarkastelussa tiimin jäsenet kertovat mikä on sujunut hyvin ja missä olisi kehitettävää. Parannukset pyritään tekemään seuraavan iteraation aikana priorisoinnin mukaisesti.

## 4. KETTERÄ LIIKETOIMINTATIEDON HALLINTA

Agile business intelligence tarkoittaa suoraan suomennettuna ketterää liiketoimintatiedon hallintaa. Termin merkitys ei ole vielä vakiintunut ja sillä voidaan tarkoittaa monia eri asioita. Agile viittaa ketteriin menetelmiin, jolloin termillä voidaan tarkoittaa liiketoimintatiedon hallintaprojektin toteuttamista ketterillä menetelmillä. Tätä käsitellään seuraavassa alaluvussa. Ketterä viittaa myös ketterään liiketoimintatiedon hallintaympäristöön, joka on joustava ja ketterä muutoksille. Tätä tarkastellaan toisessa alaluvussa. Kolmannessa alaluvussa puolestaan pohditaan tietovaraston ketterää rakentamista.

### 4.1. Ketterät menetelmät liiketoimintatiedon hallinnassa

Pentaho on kaupallinen avoimen lähdekoodin BI-toimija, jonka tuotteita ovat ETL-työkalut, raportointi-, analysointi- ja tiedonlouhintatyökalut. Pentahon sponsoroimassa kilpailussa ihmisiltä kysyttiin mitä tarkoittaa ”Agile BI”. Suurin osa vastauksista, noin 34 % oli sitä mieltä, että se tarkoittaa iteratiivista ja inkrementaalista menetelmää kehittää tietovarastointiratkaisuja. (Moran 2010.) Toisin sanoen liiketoimintatiedon hallintaympäristöä rakennetaan ja kehitetään hyödyntämällä ketteriä menetelmiä. Erään vastaajan mielestä ketterä tarkoittaa suunnittelu-rakentamis-toimitus-tarkkailu-syklin nopeuttamista. Mitä nopeammin toimitetaan ja tarkkaillaan, sitä paremmin oppii suunnittelemaan ja tekemään. (Moran 2010.) Vastauksen mukaan ketterään liiketoimintatiedon hallintaan riittää tehokas iteratiivisen projektimallin käyttö, vaikka ketteriin menetelmiin kuuluu myös inkrementaalinen kehitys.

Ketterät menetelmät ovat toimineet hyvin muissa ympäristöissä, joten menetelmät kannattaisi ottaa myös BI-maailmaan (Virta 2010). Toisaalta ne eivät välttämättä aina sovellu täysin samanlaisina liiketoimintatiedon hallinnan projekteihin. Ketterässä tietovarastoinnissa (engl. agile data warehousing) ja ketterässä ohjelmistokehityksessä ei ole mitään yhteneväistä, vaikka termeissä on yksi sama sana (Swoyer 2010b). Toisaalta Wailgumin (2010) mukaan ketterässä liiketoimintatiedon hallinnassa toimintatavat ovat samanlaisia kuin ketterissä kehitysmenetelmissä niiltä osin, että tuotetta toimitetaan inkrementaalisissa osissa jatkuvasti eikä kertarykäyksellä, tehdään nopealla syklillä prototyyppejä tarkkojen määritelmien sijaan, pyritään reagoimaan eikä keskityä suunnittelemaan kaikkea, sekä panostetaan henkilökohtaiseen vuorovaikutukseen liiketoimintakäyttäjien kanssa enemmän kuin täsmälliseen dokumentaatioon. Nämä ovat pääosin samoja asioita, joita ketterien menetelmien manifestissa määriteltiin (ks. luku 3.4). Howsonin mukaan voidaan yleisesti todeta, että

ketterä kehitys on tärkeä tekijä onnistuneissa tietovarastointiprojekteissa (Powell & Howson 2010).

#### **4.1.1. Ketterien menetelmien hyödyt ja haitat liiketoimintatiedon hallinnan projekteissa**

Onnistuneen liiketoimintatiedon hallintaprojektin yksi kriittisimmistä elementeistä on vaatimusten määrittely, joka on ketterien menetelmien suurin vahvuus. Liiketoiminnan jatkuvien muutosten takia liiketoimintatiedon hallinnan vaatimukset muuttuvat myös jatkuvasti, jolloin ketterät menetelmät ovat vahvoilla. (Hammergren 2010a.) Vesiputousmallin mukaan vaatimukset määritellään aluksi ja projekti toteutetaan sen mukaisesti. Ketterillä menetelmillä voidaan jokaisen syklin alussa tarkastella sen hetkisiä vaatimuksia ja pyrkiä vastaamaan niihin. Vaatimukset muuttuvat liiketoimintatiedon hallinta- ja tietovarastointiprojekteissa vielä useammin kuin perinteisessä ohjelmistokehityksessä (Hammergren 2010b). Lisäksi kaikkia vaatimuksia on vaikea ymmärtää ja määritellä täsmällisesti etukäteen, joten tietovarasto on helpompi toteuttaa pienissä osissa (Swoyer 2010a). Tämä on ketterien menetelmien ehdoton vahvuus tietovarastointiprojekteissa.

Ketterässä manifestissa (ks. luku 3.4) määriteltiin kaksitoista tärkeää periaatetta, joita tulisi noudattaa. Näistä kuusi sopivat Howsonin (Powell & Howson 2010) mukaan liiketoimintatiedon hallintaprojekteihin:

- 1) Tärkein tehtävä on toimittaa asiakkaalle varhaisessa vaiheessa ja jatkuvasti arvoa tuottavaa ohjelmistoa
- 2) Muutokset vaatimuksiin ovat tervetulleita, myös myöhäisessä kehitysvaiheessa. Ketterillä prosesseilla muutokset saadaan muutettua asiakkaan kilpailueduksi.
- 3) Liiketoimintahenkilöiden ja kehittäjien on työskenneltävä yhdessä päivittäin koko projektin ajan.
- 4) Tehokkain tapa välittää tietoa keskustelu kasvojen välillä.
- 5) Yksinkertaisuus – keskeistä on minimoida turhan työn määrä.
- 6) Parhaat arkkitehtuurit, vaatimukset ja suunnitelmat syntyvät itse-organisoiduista tiimeistä.

Näistä neljä ensimmäistä viittaa asiakasrajapintaan, IT:n ja liiketoiminnan väliseen rajapintaan. Niillä pyritään parantamaan asiakastytyväisyyttä sekä tietovarastoinnista saatua hyötyä. Huonolla tietovarastointiratkaisulla ei ole mitään arvoa, sillä epäluotettavasta tai väärästä informaatiosta on vain haittaa. Itseään selvyyksien esittämisestäkään ei ole hyötyä. Hyötyjä voidaan saavuttaa usein heikollakin tietojärjestelmällä, kuten toiminnanohjaus- tai asiakashallintajärjestelmällä. Ne voivat silti ajaa asiansa ja olla hyödyllisiä, toisin kuin huono liiketoimintatiedon hallintajärjestelmä. Tämän takia asiakasrajapinta on erityisen tärkeä. Manifestissakin on määritelty arvostamaan vuorovaikutusta ja asiakasyhteistyötä enemmän kuin työkaluja ja sopimusneuvotteluita.

Hyvällä asiakasvuorovaikutuksella, eli tietovarastoinnin kehittäjien ja käyttäjien välisellä vuorovaikutuksella saadaan liiketoimintatiedon hallintaympäristöstä parempi. On keskityttävä toimivaan sovellukseen mieluummin kuin kattavaan dokumentaatioon. Ylimääräinen dokumentaatio on turhaa työtä, eikä dokumentaatio ole itseisarvo. Toimiva raportointisovellus tarjoaa tiedon helposti saataville ns. infoähkyä, eli liiallisen informaation esittämistä välttämällä. Käyttäjien on saatava juuri sitä tietoa, mitä he kulloinkin tarvitsevat. Parhaiten käyttäjän tietotarpeet tietävät käyttäjät itse. Tämän takia asiakasrajapinta korostuu myös toimivan sovelluksen näkökulmasta asiaa tarkasteltaessa. Liiketoimintatiedon hallintaympäristöä kehitetään tavallisesti jatkuvasti, jotta tiedosta saadaan mahdollisimman paljon irti. Tietovarastointijärjestelmät eivät ole myöskään niin stabiileja, kuin operatiiviset tietojärjestelmät. Tästä seuraa dokumentaation jatkuva vanheneminen. Ketterät menetelmät eivät vapauta dokumentoinnista kokonaan, mutta se voi vähentää muodollista ja ylimääräistä dokumentaatiota (Powell & Howson 2010).

#### **4.1.2. Milloin ketterät menetelmät sopivat liiketoimintatiedon hallinnan projekteihin**

Ketterät menetelmät sopivat niin innovatiivisiin yrityksiin kuin perinteisiin teollisuusyrityksiin, sekä yhtä hyvin myös tietovarastoinnin ulkoistaneille yrityksille (Powell & Howson 2010). Jotta voidaan saavuttaa ketteryyttä liiketoimintatiedon hallinnassa, on kuitenkin tarkasteltava miten muut osat BI-toimitusprosessissa ovat omaksuneet ketterät menetelmät. On oltava prosessit, jotka mahdollistavat viestinnän ja nopeat muutokset, sekä ihmisillä on oltava valtuudet tehdä päätöksiä. (Virta 2010.) Ketteriin menetelmiin kuuluu itseorganisoituva tiimi, jolloin vahvasti hierarkkisiin yrityksiin tai organisaatioihin ne eivät sovi.

Ketterän liiketoimintatiedon hallinnan menetelmät poikkeavat muista ketteristä menetelmistä siten, että se vaatii tuekseen uusia ja erilaisia teknologioita, kuten metadatan generoiva BI-sovellus (Wailgum 2010). Jos käytetyt työkalut ja arkkitehtuurit eivät sovi ketteriin toimintatapoihin, on parempi toteuttaa projektit esimerkiksi vesiputousmallin mukaisesti. Lisäksi jos ketteriä menetelmiä ei ole omaksuttu yrityksessä tai organisaatiossa ollenkaan tai niitä vastustetaan, on tietovarastointiprojektilla suurempi todennäköisyys onnistua perinteisemmän, vesiputousmallin mukaisesti. Yritykset, jotka haluavat siirtyä ketterään liiketoimintatiedon hallintastrategiaan, tulee aloittaa pienellä, kohdistetulla esimerkkitapauksella (Wailgum 2010).

Ketterät menetelmät eivät myöskään sovellu parhaiten projekteihin, joissa niiden tuoma hyöty jää vähäiseksi. Mikäli vaatimukset ovat hyvin selkeitä ja stabiileja, sekä projektilta vaaditaan hyvää ennustettavuutta, vakautta ja korkeaa varmuutta, on vesiputousmalli parempi vaihtoehto (Boehm & Turner 2004, s. 718). Lisäksi

itseorganisoidun tiimin tulee olla sopivan kokoinen, jotta on järkevää toimia itseorganisoidusti ja kokoontua päivittäiseen tapaamiseen, kuten Scrum-mallissa. Tiimin päivittäiset tapaamiset, keskustelut kasvotusten tiimin työntekijöiden ja asiakkaan kanssa voivat olla mahdottomia, jos tiimi on hajallaan. Toisaalta tällaisessakin tilanteissa ketteriä menetelmiä voisi käyttää soveltuvin osin tai soveltaa niistä itselle hyvä toimintatapa tehdä BI-projekteja.

## **4.2. Ketterä liiketoimintatiedon hallintaympäristö**

Liiketoimintatiedon hallintaympäristö voi olla ketterä kahdesta näkökulmasta. Se voi olla joustava ja tehokas lähdedatan muutoksille. Toisaalta se voi olla ketterä myös käyttäjän näkökulmasta, eli mukautua liiketoiminnan vaatimuksiin joustavasti. Jälkimmäisen mukaan se tarjoaisi ketterää raportointia, vastauksia monimutkaisiin ja ennalta määrittelemättömiin ad hoc -kyselyihin.

### **4.2.1. Muuttuva lähdedata**

Monet ihmiset ajattelevat että ketterä liiketoimintatiedon hallinta on yksinkertaisesti ketterien menetelmien soveltamista BI-hankkeissa, mutta se ei täysin pidä paikkansa. Ketterässä liiketoimintatiedon hallinnassa on kyse myös siitä kuinka työkaluja käytetään, mitä ne mahdollistavat ja miten työkalut on rakennettu. (Hammergren 2010b.) Hyvät työkalut mahdollistavat ketterämmän liiketoimintatiedon hallintaympäristön rakentamisen ja ketterämmän raportoinnin.

Lähdejärjestelmien data muuttuu ja lisääntyy jatkuvasti. Muutokset liiketoiminnassa vaatii edelleen uusia tai muuttuneita transformaatioita. (Biere 2003, s. 13.) Liiketoiminnassa tapahtuvat muutokset vaikuttavat lähdejärjestelmiin, kun operatiivisia järjestelmiä mukautetaan toimintaan. Koko lähdejärjestelmän rakenne saattaa myös muuttua esimerkiksi ohjelmiston versiovaihdon yhteydessä. ETL-prosessia on tällöin aina muutettava ja uusi tieto saatava tietovarastoon sekä raportointiin. Tämän ei voi, eikä tarvitsekaan olla täysin automaattista, mutta siihen on hyvä varautua. Operatiivisissa järjestelmissä tapahtuvista versiovaihdoista tulee kyllä signaalia ETL:n kehittäjille, mutta pienimmistä muutoksista harvemmin muistetaan ja ymmärretään tiedottaa.

ETL-prosessia rakennettaessa tulevat muutokset on otettava huomioon. Asia ei ole uusi, eikä edes aina vaikeasti toteutettavissa, mutta se vähentää huomattavasti ylläpidon työtä. Vikailmoitusjärjestelmä on Kimballin ETL-alisysteemi numero 30 (Casters et al. 2010, s. 124). Kun ajastettu ETL-prosessi epäonnistuu, sen tulee tehdä virheilmoitus esimerkiksi lähettämällä sähköpostiviesti. Näin ylläpitäjät voivat reagoida nopeammin muutoksiin.

#### **4.2.2. Ketterä raportointi**

Ketterällä liiketoimintatiedon hallinnalla on kaksi tarkoitusta: saada kehitys tapahtumaan nopeammin ja reagoida nopeammin liiketoiminnan muuttuviin vaatimuksiin (Wailgum 2010). Edellistä käsitellään seuraavassa alaluvussa, kun taas jälkimmäistä käsiteltiin jo aiemminkin. Ketterillä menetelmillä voidaan vastata paremmin muuttuviin vaatimuksiin, mutta liiketoiminnan muutokset pitää huomioida liiketoimintatiedon hallintaympäristön rakentamisessa. Ketterä liiketoimintatiedon hallintaympäristö mahdollistaa organisaation ihmiset ja prosessit sopeutumaan nopeasti uusiin ja muuttuviin vaatimuksiin, mieluiten reaktiivisesti (Moran 2010). Tällä voidaan tarkoittaa sitä, että tietovarastoinnin kehittäjät sopeutuvat uusiin vaatimuksiin ja kehittävät tietovarastoa ja raportointia vaatimusten mukaisesti tai puolestaan sitä, että tietovarasto on rakennettu joustavaksi ja helposti muutettavaksi.

ETL-prosessiin, tietojen integrointiin ja muokkaukseen kannattaa panostaa, sillä näin muodostettujen helppokäyttöisten rakenteiden ansiosta tehostetaan huomattavasti tietojen raportointia ja analysointia BI-välineillä (Hovi et al. 2009, s. 48). Liiketoimintatiedon hallintaprojektit ovat monimutkaisia ja tietotekniikka-arkkitehtuuri haastavaa, joten BI-sovellukset ovat vaikeakäyttöisiä ja joustamattomia (Wailgum 2010). Tietotarpeet muuttuvat jatkuvasti ja kun vastauksia saadaan joihinkin kysymyksiin, tulee uusia kysymyksiä mieleen. Tämän takia raportoinnin tulee olla ketterää ja mahdollisimman helppokäyttöistä. Analysoinnin tulee mahdollistaa monimutkaisia kyselyitä, se ei saa olla ainoastaan staattisia mittaristoja. Liiketoimintatiedon hallinnan hyödyt jäävät vähäiseksi, jos raportointi ei ole käyttäjäystävällistä ja monipuolista.

#### **4.3. Liiketoimintatiedon hallintaympäristön rakentaminen ketterästi**

Liiketoimintatiedon hallintaympäristön rakentaminen on perinteisesti paljon työtä vaativaa ja hyvin kallista. Pienilläkin muutoksilla ja automatisoinnilla voidaan saada aikaan suuria säästöjä. Tietovarastosta ei ole mitään arvoa loppukäyttäjälle, ennen kuin kaikki tarvittava data on tietokannassa ja siihen on lisätty raportointi- tai kyselymahdollisuus. Liiketoimintatiedon hallinnan projekteissa 80 % ajasta, budjetista ja resursseista kuluu tiedon siirtämiseen, muokkaamiseen ja lataamiseen tietovarastoon. (Biere 2003, s. 13.) Hovi et al. (2009, s. 48) arvioivat myös ETL-prosessin toteutuksen testausineen vievän noin 60–80 % tietovarastohankkeen toteutusajasta, vaikka nykyään ETL-työkalut sisältävät paljon valmista toiminnallisuutta helpottamaan ja nopeuttamaan työtä. Tämän takia ETL-prosessin rakentaminen sisältää paljon potentiaalia ja suuria mahdollisuuksia, kustannussäästöt ovat selviä.

Kun ETL-prosessiin käytettyjä resursseja minimoidaan, voidaan säästää huomattavasti sen aiheuttamissa kustannuksissa ja ylläpidossa. Se kuitenkin vaikuttaa dramaattisesti tietovarastoinnista saatuihin hyötyihin ja voi haitata oleellisesti raportoinnin toteuttamista. (Biere 2003, s. 14.) ETL-prosessi on toteutettava siis hyvin, jotta se ei pienennä raportoinnin mahdollisuuksia ja koko tietovarastoinnin hyötyjä. Esimerkiksi historiointi on haastavaa ja työlästä toteuttaa, mutta se voi olla oleellisen tärkeää tietovarastoinnin käyttäjille. Rutiinyyötä on pyrittävä vähentämään, mihin yksi ratkaisu on edellä mainitut tehokkaat ETL-työkalut. Ne helpottavat ja automatisoivat paljon pieniä asioita, yksittäisiä työvaiheita. Kaikkia perusasioita ei tarvitse enää erikseen toteuttaa, vaan nykyaikaisissa ETL-välineissä ne on valmiiksi toteutettu. Esimerkiksi erikseen ei tarvitse etsiä oikeaa riviä tietovarastosta, tarkastaa päivitetäänkö tietue vai lisätäänkö uusi ja luodaanko uusi surrogaattivain, vaan tämä kaikki on toteutettavissa yhdellä työvaiheella. ETL-välineet tulevat varmasti myös paljon kehittymään lähitulevaisuudessa. Useimmiten ne sisältävät myös paljon valmiita rajapintoja yleisimpiin tietojärjestelmiin, asiakashallintajärjestelmistä toiminnanohjausjärjestelmiin sekä erilaisiin tietokantoihin ja tiedostomuotoihin.

Suurempiakin kokonaisuuksia voitaisiin automatisoida, sillä eri prosessit ovat pitkälti samanlaisia. Esimerkiksi dimensioiden ja faktojen lataaminen tietovarastoon sisältää pitkälti vakiintuneita käytäntöjä. Automatisoimalla suurempia kokonaisuuksia ETL-prosesseista voitaisiin vähentää työtä huomattavasti. Tästä haittapuolena on joustamattomuus muutoksille, joita ei ole ennalta määritetty. Kun jokin prosessi automatisoidaan, se toimii aina samalla tavalla ja pieniäkin muutoksia voi olla hyvin vaikea toteuttaa. Toisaalta ennalta määritellyt muutokset voivat olla hyvinkin yksinkertaista toteuttaa, kun parametrisoidaan automatisoivaa rutiinia. Ideaalitalanne olisi, jos tarvitsisi vain määritellä lähdejärjestelmästä, mitä tietoja halutaan mitata ja mitkä ovat dimensiot. Toisaalta mittarit ja dimensiotkin ovat tyypillisesti pääosin samoja.

Tietovarasto voidaan standardoida kaikille samaksi, riittävän laajaksi ja monipuoliseksi. Näin teki ohjelmistotalo Oracle, joka julkaisi SAP:n toiminnanohjausjärjestelmien kanssa yhteensopivan BI-sovelluksen. Raporttipohjia siinä on valmiina jo kaksisataa suunniteltuna eri alojen johtajille. (Korhonen 2011.) Tietovarasto tarvitsee näin suunnitella vain kerran ja raportitkin ovat yhteensopivat kaikkiin. ETL:n tekeminen helpottuu ja nopeutuu, kun tietovarasto on valmiina, ellei koko ETL-prosessia automatisoida. Vaarana on liiallinen tiedon määrä, eli niin sanottu infoähky. Lisäksi käyttäjät eivät silti välttämättä saa BI:stä juuri niihin kysymyksiin vastauksia, joihin kulloinkin haluavat. Valmis tietovarasto voi olla myös hyvin joustamaton tapahtuville muutoksille. Liiketoimintatiedon hallinnassa onkin ollut tärkeää se, että raportointi määritellään ja räätälöidään asiakkaan tarpeisiin.



Ketterään tietovaraston rakentamiseen on myös muita vaihtoehtoja, kuten ”virtuaalinen” tietovarasto. Virtuaalinen tietokerros (engl. virtualized data layer) kertoo missä tieto sijaitsee ja näin mahdollistaa käyttäjille pääsyn tietoon, sen sijaan että se ladattaisiin tietovarastoon (Swoyer 2010a). Lisäksi se mahdollistaisi täysin reaaliaikaisen tiedon. Aivan ongelmaton tämäkään vaihtoehto ei ole. Vaikka ETL-prosessia ei tarvitsisi toteuttaa, menetetään sen tuomat hyödyt. Perinteiseen, fyysiseen tietovarastoon voidaan tallentaa historiatietoa ja summaustaulut tuovat nopeutta ja tehokkuutta, mitkä jäävät virtuaalisesta tietovarastosta puuttumaan. Nykytekniikka ei mahdollista täysin vielä virtuaalista tietovarastoa, sillä datan kerääminen suoraan operatiivisista järjestelmistä raporteille on hidasta.

Perinteinen yhtä kohdetta varten toteutettu ETL-rutiinin rakentaminen poikkeaa melko paljon edellä mainituista. Taulukkoon 4.1 on arvioitu karkeasti eri toimintatapojen eroja valittujen ominaisuuksien mukaan. Plus-merkkien määrä kertoo kuinka hyvä malli on kunkin ominaisuuden kohdalla. Monistettavuus kertoo kuinka helposti liiketoimintatiedon hallintaympäristön voi monistaa. Tämä on tärkeä ominaisuus, kun rakennetaan useita samankaltaisia tietovarastoja. Toinen ominaisuus kertoo kuinka nopeasti tietovarasto saataisiin rakennettua. Kolmas ominaisuus, reaaliaikaisuus, kuvaa kuinka reaaliaikaista tieto voisi olla. Neljäs kertoo kuinka joustava tietovarasto on muutoksille. Ensimmäinen toteutus tarkoittaa kuinka nopeasti ja vähillä resursseilla ensimmäinen tietovarasto ja raportointiympäristö saadaan rakennettua. Toiseksi viimeinen ominaisuus kuvaa kuinka hyvin ko. tavalla voidaan vastata liiketoiminnan vaatimuksiin ja viimeinen kuinka helposti tietovarasto on toteutettavissa kyseisellä tavalla.

	Monis- tetta- vuus	Käyt- töön noton nopeus	Reaali- aikaisuus	Muutok- sien tekeminen	Ensim- mäinen toteutus	Tukee vaati- muksia	Toteutet- tavuus
Perinteinen ETL	+	-	+	+++	+++	+++	+++
Autom. ETL	+++	+++	+	+	-	++	++
Valmis tietomalli	+++	+++	+	-	-	+	++
Virtuaalinen DW	-	+	+++	++	+	++	-

*Taulukko 4.1 Liiketoimintatiedon hallintaympäristön rakentamistapojen vertailu.*

Automatisoitu ETL ja valmiin tietomallin ratkaisu ovat parhaiten kopioitavissa ja hyvin nopeasti otettavissa käyttöön. Perinteinen ratkaisu, tiettyyn kohteeseen rakennettu ETL-prosessi on myös kopioitavissa tehdystä rakenteesta ja käytetyistä työkaluista riippuen. Virtuaalisen tietovaraston vahvuuksia ja heikkouksia on vaikea arvioida, sillä sitä ei taida olla toteutettu. Se voisi olla nopeasti rakennettavissa, sillä ETL:ää ei tarvitse

toteuttaa. Reaaliaikaisuus on selvä etu. Muutoksien tekeminen ja joustavuus on valttina perinteisessä ja virtuaalisen DW:n malleissa. Yhtä toteutusta varten perinteinen on nopeinta tehdä. Se tukee myös liiketoiminnan vaatimuksia parhaiten ja on helpoiten toteutettavissa.

Perinteinen ETL:n rakentaminen on vielä hyvin vahvoilla muihin ratkaisuihin verrattuna. Kun samanlaisia liiketoimintatiedon hallintaympäristöjä rakennetaan useita, alkavat automatisoidut ETL-prosessit ja valmiin tietomallin edut olemaan selkeät. Niillä voidaan vähentää huomattavasti työmäärää, mutta joustavuus saattaa ja vaatimustenmukaisuus saattaa heikentyä. Erityisesti ne vaativat paljon työtä ennen ensimmäistä toteutusta.

Ketterä liiketoimintatiedon hallinta on selvästi tullut jäädäkseen ja menestymään, sillä useimmat perinteiset, ei-ketterien BI-alustojen ja sovellusten on yhä vaikeampi pysyä mukana salamannopeasti muuttuvassa maailmassa (Walgun 2010). Aina riittää kuitenkin skeptikoita, sillä joidenkin mielestä ketterä liiketoimintatiedon hallinta, eli agile BI on uusi markkinointitermi, joka ei varsinaisesti tarkoita yhtään mitään (Moran 2010). Ketterän liiketoimintatiedon hallinnan tarkoituksena on saada kehitys tapahtumaan nopeammin ja reagoida nopeammin muutoksiin (Walgun 2010). Ketterät menetelmät ovat yksi keino vastata näihin haasteisiin, mutta tärkeintä on hyvät työkalut, prosessit, viestintä asiakkaan kanssa, avoin suhtautuminen uusiin ratkaisuihin ja ammattitaitoiset työntekijät.

## 5. TUTKIMUSAINEISTO JA MENETELMÄT

### 5.1. Kohdeyritys

Tämä diplomityö on tehty asiakashallintaan keskittyneelle konsulttiyritykselle IPSS Oy:lle (Intelligence Precision Solution and Services Oy). IPSS Oy on vuonna 1999 perustettu ja jatkuvasti kasvava pk-yritys. IPSS tarjoaa erilaisia SaaS-ratkaisuja (Software as a Service), ylläpitoa ja konsultointia. IPSS Oy:n sisaryhtiö on pari vuotta aikaisemmin perustettu ICMI Oy (Intelligent Customer Management International Oy), jonka kanssa tehdään paljon yhteistyötä.

IPSS Oy auttaa yrityksiä tehostamaan myyntiä ja markkinointia kokoamalla vaadittavat työkalut, toimintamallit sekä asiakastiedot. ICMI Oy puolestaan tarjoaa yrityksille asiantuntemusta asiakasjohtamisen ja asiakassuhteiden strategisiin ja toiminnallisiin kysymyksiin. IPSS Oy:n henkilöstö koostuu yli 30:stä CRM-, integraatio- ja tietovarastointialueen erikoisosajista. Yhteensä yritykset työllistävät noin 40 henkilöä.

IPSS Oy keskittyy uudenaisiin suhdehallinnan ja asiakkuuksien johtamisen tietojärjestelmäratkaisuihin. Yrityksen ratkaisut tuottavat lisäarvoa asiakasyritysten suhdehallintaan ja tehostavat näiden toimintaa. IPSS Oy:llä on pitkä kokemus asiakashallinnan ja järjestelmäratkaisujen kehittämisestä eri toimialoilla ja alustoilla.

IPSS Oy toimittaa omien asiakashallintajärjestelmien lisäksi yhteistyökumppaneiden järjestelmiä, joita tarvittaessa täydennetään omilla modulaarisilla komponenteilla. Omia tuotteita ja palveluita ovat CRM:n (Customer Relationship Management) lisäksi MDM:n (Master Data Management) ja tietovarastoinnin alueelta. Liiketoimintatiedon hallinta liittyy asiakashallintajärjestelmiin olennaisesti, sillä sitä käytetään muun muassa asiakaspotentiaalin tunnistamisessa ja aktiviteettien hallinnassa. Lisäksi tietovarastointiratkaisua on tuotteistettu ja raportointia on integroitu vahvasti muihin järjestelmiin.

Diplomityöntekijä on työskennellyt IPSS Oy:llä kesästä 2010 ja on toiminut Business Intelligence -konsulttina diplomityön ohella. Työnkuvaan kuuluu tietovarastopohjaisten BI-ratkaisujen toteutus, kuten ETL-rutiinien tekeminen, tietovaraston rakentaminen, kuutioiden ja analyysinäkymien tekeminen sekä mittaristojen ja raporttien laatiminen. Lisäksi toimenkuvaan kuuluu dataintegraatiot esimerkiksi asiakkaiden omien järjestelmien tai tietovaraston ja CRM-järjestelmän välillä.

## 5.2. Työn taustaa

Kellotaajuuden kasvu, eli globalisaation aiheuttama markkinoiden muutosnopeuden kasvu lyhentää päätöksentekoon käytettävää aikaa (Halonen & Hannula 2007, s. 3). Tämän seurauksena liiketoimintatiedon hallinta korostuu entisestään. Kellotaajuuden kasvu lisää muutosnopeutta myös liiketoiminnan tarpeissa, lähdejärjestelmissä ja tietosisällöissä. Tämä tuo paineita kehittää liiketoimintatiedon hallintaympäristöä entistä joustavammaksi ja ketterämmäksi jatkuville muutoksille.

Tietovarastointiprojektit ovat usein melko raskaita ja kuluttavat paljon resursseja. Joitakin vuosia sitten karkea budjetti-arvio tietovarastointiprojektille oli miljoona markkaa ohjelmistoon, miljoona laitteisiin ja miljoona työhön (Pulkkinen 2008). Kustannuksia riippuvat tietysti hyvin pitkälti yrityksen tai organisaation laajuudesta, tietovaraston laajuudesta, datan monimuotoisuudesta ja laadusta sekä käytetyistä ohjelmistoista ym. tekijöistä. Projektista riippuen pelkät vuotuiset ohjelmisto- ja ylläpitokustannukset voivat vaihdella sadoista tuhansista miljoonaan euroon (Biere 2003). Vuoren ja Hannulan (2009, s. 6) tutkimuksen mukaan suomalaisten suuryritysten liiketoimintatiedon hallinnan vuotuinen budjetti on keskimäärin 0,5-1 miljoonaa euroa.

Vaikka laitteiden hinnat ovat laskeneet ja markkinoille on tullut ilmaisia, avoimen lähdekoodin tuotteita, tietovarastointiprojektit vaativat paljon taloudellista pääomaa. Lisäksi tietovarastointiprojektit vaativat paljon osaamista ja kestävät usein melko pitkään. Näiden takia erityisesti pienemmillä yrityksillä ja organisaatioilla ei ole resursseja hankkia BI-järjestelmiä, rakentaa tai rakennuttaa tietovarastoa ja liiketoimintatiedon hallintaympäristöä.

Muutokset tietovarastoissa ovat kalliita toteuttaa ja usein hidastavat projektia. Erityisesti projektin valmistuttua pienetkin muutokset voivat aiheuttaa suuria kustannuksia. Pienistäkin muutoksista voi joutua muuttamaan ETL-prosesseja, välitietokantoja, tietovaraston tauluja, summaustauluja, kuutioita, raportteja ja mittaristoja sekä dokumentaatiota. Ylläpito on työlästä, jos muutoksia ei ole huomioitu rakentamisvaiheessa.

IPSS Oy:llä on omassa tuotekehityksessään tarpeita kehittää liiketoimintatiedon hallintaympäristön rakentamista. Aikaisempia projekteja halutaan hyödyntää löytämällä geneerisiä malleja ja tuotteistaa palveluja. Samoja asioita tehdään paljon jokaisessa tietovarastointiprojektissa ja asiakaskohtainen räätälöinti on vähäistä. Liiketoimintatiedon hallintaympäristön rakentamiseen vaaditaan paljon osaamista ja henkilötyötunteja. Ylläpidon aiheuttama resurssien kulutus halutaan myös minimoida. Dokumentaation tulisi olla automaattista siitä, mitä on tehty, jotta voitaisiin välttyä niiden jatkuvalta päivittämiseltä. Määrittely-, suunnittelu- ja toteutusvaiheen

dokumentointi ei ole täysin yhdenmukaista. Nämä asiat tuovat haasteita myös projektinhallintaan ja resurssien allokointiin.

Näihin haasteisiin pyritään löytämään ratkaisuja ja ratkaisuehdotuksia ketterällä liiketoimintatiedon hallinnalla ja tällä diplomityöllä. Ketterillä menetelmillä, ketterillä liiketoimintatiedon hallintaympäristöillä ja ketterästi rakennettavat, jopa automatisoidut tietovarastot voisivat tuoda halutun kilpailuedun tietovarastointiprojektien ja konsultoinnin markkinoilla.

### 5.3. Tutkimusmenetelmät

Tämän tutkimuksen empiirisen osan tutkimusmenetelmät ovat monimetodiset ja laadulliset eli kvalitatiiviset. Menetelmiin kuuluu haastatteluja ja omat kokemukset, osallistuva tekeminen. Tutkimusaineistoa kerätään puolistrukturoiduilla teemahaastatteluilla, joita tehdään kolme kappaletta. Haastattelujen kohderyhmä on liiketoimintatiedon hallintaprojekteihin osallistuvat henkilöt, otokseen kuuluu IPSS Oy:n BI-asiantuntijat kaikilta organisaatioitasoilta, projektinhallinnasta, konsultoinnista ja tietovarastojen rakentajista. Haastatteluilla pyritään saamaan tietoa erityisesti projektimalleista, menetelmistä, mutta myös liiketoimintatiedon hallintaympäristön rakentamisesta. Haastatteluilla pyritään löytämään vastauksia erityisesti työn toiseen alatutkimuskysymykseen, eli miten ketteriä menetelmiä voitaisiin hyödyntää liiketoimintatiedon hallintaprojekteissa.

Haastateltavien valinnassa käytetään ns. lumipallo-otantaa (engl. snowball sampling). Menetelmään kuuluu, että aluksi valitaan yksi haastateltava satunnaisesti ja häneltä kysytään seuraavaa mahdollista haastateltavaa (Saunders 2009, s. 240). Vanhat työntekijät tietävät hyvin, ketä kannattaisi haastatella, joten menetelmä sopii hyvin. Otos on melko pieni yrityksen BI-tiimin pienestä koosta johtuen, joten lopulliseen aineistoon päätynee iso osa koko otoksesta. Variaatiota kuitenkin pyritään laajentamaan sisällyttämällä otokseen henkilöitä eri tehtävistä, projektipäälliköistä ETL-rutiinien rakentajiin ja erilaisten työkalujen käyttäjiin.

Kaikki haastattelut nauhoitettiin haastateltavan suostumuksella, mutta nauhoja ei litteroida. Haastattelutilanteissa otetaan mahdollisimman kattavat muistiinpanot. Haastattelut pidetään IPSS Oy:n tiloissa sopivassa, rauhallisessa paikassa.

Haastattelujen lisäksi työn toinen tutkimusmenetelmä on omat kokemukset ja osallistuva tekeminen. Tietoa kerätään käytännön työstä jatkuvasti ja kerrytetään kokemuseräistä tietoa koko diplomityöprosessin ajan. Tietovarastoja rakennetaan ja kehitetään eri työkaluilla, lähinnä avoimen lähdekoodin PDI:llä (Pentaho Data Integrator) eli Kettlellä, mutta myös Microsoftin SSIS:llä (SQL Server Integration Services).

Työssä tehdään käytännön sovellus ETL-prosessista valituilla työkaluilla. Tiedonsiirtorutiini poimii haluttua dataa Salesforce-nimisestä CRM-järjestelmästä, tallentaa välivarastointitietokantaan, josta tiedot päivitetään tietovarastoon. Tämä pyritään tekemään mahdollisimman automaattiseksi, joustavaksi ja käyttöönotto helpoksi. IPSS Oy:n tietovarastointiprojekteissa Salesforce toimii tyypillisesti yhtenä tietolähteenä tietovarastoille asiakkaiden omien ERP-järjestelmien lisäksi. Salesforce on valittu työhön, koska se vaatii vähän asiakaskohtaista räätälöintiä tiedonlatausprosesseissa, kun taas toiminnanohjausjärjestelmät ovat hyvin asiakaskohtaisia. Järjestelmä tarjoaa hyvät rajapinnat tiedon lataukseen ja metatietojen kyselyyn sekä saatavilla on myös sen tietomallit. Salesforcen rakenne on samanlainen kaikissa ympäristöissä, vaikka tietosisältö ja kiinnostavat tietokentät vaihtelevat. Omat kokemukset vaikuttivat valintaan, vastaavanlaista ETL-prosessia on tehty jo useampaan ympäristöön. Sovelluksella pyritään löytämään vastauksia kolmanteen ja neljänteen alatutkimuskysymykseen, eli ketterään tietovaraston rakentamiseen ja ketterään liiketoimintatiedon hallintaympäristöön.

## **6. HAASTATTELUT JA NIIDEN TULOKSET**

Tässä luvussa käsitellään haastatteluja, niiden kulkua ja tuloksia. Aluksi käydään läpi haastatteluprosessi, esitellään haastateltavat henkilöt, heidän valinta, haastattelurunko ja haastattelujen kulku. Tämän jälkeen, toisessa alaluvussa pohditaan haastattelujen tuloksia osio kerrallaan ja lopuksi tiivistetään yhteenvetoon.

### **6.1. Haastatteluprosessi**

#### **6.1.1. Haastatellut henkilöt**

Haastattelujen ensimmäiseksi henkilöksi valittiin nuori BI-konsultti, Paavo Toiviainen. Hän on toiminut heti diplomi-insinööriksi valmistumisen jälkeen noin kaksi vuotta liiketoimintatiedon hallinnan projekteissa. Paavon tehtäviin kuuluu tietovarastoarkkitehtuurien suunnittelu, ETL-prosessien ja raportointijärjestelmien toteuttaminen sekä osallistuminen niiden ylläpitoon. Hän on toiminut pääasiassa yhdessä suuremmissa asiakasprojektissa, mutta myös monissa muissa pienemmissä projekteissa.

Toinen haastateltava henkilö oli IPSS Oy:n BI-tiimin vetäjä, konsultti Antti Paappanen. Hän toimii esimiehenä, tuotteen vastuuhenkilönä ja on toiminut lisäksi liiketoimintatiedon hallintaprojekteissa muun muassa toteuttavassa roolissa, projektipäällikkönä, asiantuntijana ja myyjänä. Antti on koulutukseltaan tradenomi ja hänellä on kymmenen vuoden työkokemus myynnin ja markkinoinnin tukisovelluksista ja tietovarastoinnista. Hänellä on myös pitkä ura IPSS Oy:llä BI-konsulttina.

Kolmantena haastateltavana oli IPSS Oy:n toimitusjohtaja, Terho Norja. Hänellä on pitkä kokemus asiakashallinnan ja asiakasjohtamisen alueilta ja on toteuttanut tietovarastoja jo niiden varhaisessa kehitysvaiheessa. Terho on opiskellut ohjelmistotekniikkaa ja on koulutukseltaan diplomi-insinööri. Hän on ollut kehittämässä tietovarastointiratkaisuja, kun asiakashallinnassa on pitänyt saada asiakastiedosta kaikki mahdollinen irti. Terholla on kokemusta tiedon integroinnista sekä tietojärjestelmien suunnittelu- ja projektinjohtotehtävistä. Terho on IPSS Oy:n perustajajäsen ja seniorikonsultti ICMI Oy:ssä.

Haastateltavat koostuivat siis organisaation eri tasoilta. Paavolla on uusinta tietoa, Antilla pitkä kokemus erilaisista projekteista tähän päivään asti ja Terholla puolestaan on varhaisinta kokemusta ajalta, jolloin tietovarastot olivat vasta kehityksen ensi

askeleilla, sekä kokemusta projektien vetämisestä, liiketoimintatiedon hallinnan hyödyntämisestä ja asiantuntijayrityksen johtamisesta.

### **6.1.2. Haastattelukysymykset**

Haastattelurunko (liite 1) koostuu taustatietoa kartoittavista alkukysymyksistä, liiketoimintatiedon hallinnan projektiin ja ympäristöön sekä dokumentointiin liittyvistä kysymyksistä sekä loppukysymyksistä. Aluksi kysytään voidaanko haastattelu nauhoittaa, esitellään haastattelurunko ja kysytään lämmittelykysymyksiä haastateltavan taustoista. Tämän jälkeen mennään ensimmäiseen varsinaiseen haastatteluosioon, liiketoimintatiedon hallintaprojektiin. Projektimenetelmistä mielenkiintoisimmat ovat ketterät menetelmät, niiden toimivuus liiketoimintatiedon hallintaprojekteissa ja tyytyväisyys niihin.

Toinen osio on liiketoimintatiedon hallintaympäristöt ja niiden rakentaminen. Tietovarastojen rakentamisessa halutuinta tietoa on kokemus erilaisista työkaluista, niiden soveltuvuus, ETL-rutiinien kopioitavuus ja tyytyväisyys työkaluihin. Lisäksi kysytään tietoa hyväksi havaituista käytännöistä, liiketoimintatiedon hallintaympäristöjen joustavuudesta ja ylläpidosta.

Kolmannessa, hieman pienemmässä osiossa kysytään dokumentaatiosta. Haastateltavilta kysytään, minkälaisia dokumentteja syntyy, onko dokumentointi riittävää ja hyödynnetäänkö sitä. Lopuksi kysytään yleisesti projektien ja tietovarastoinnin hyvin sujuvia asioita, kehityskohteita ja tulevaisuudennäkymiä. Kun on varmistettu, ettei haastateltavalla ole muuta sanottavaa, kysytään mielipidettä itse haastattelusta ja seuraavan mahdollisen haastateltavan nimeä lumipallo-otannan mukaisesti.

### **6.1.3. Haastattelujen kulku**

Ennen varsinaisia haastatteluja suunniteltiin haastattelurunko ja kirjoitettiin haastattelukysymykset. Tätä varten oli toki mietitty mitä tietoa haastatteluilla halutaan kerätä. Haastatteluilla haluttiin löytää yrityksestä kehittämiskohtia, saada kattavampi kuva liiketoimintatiedon hallinnan projekteista ja ketterien menetelmien käytöstä.

Haastattelurunkoa ei testattu etukäteen ulkopuolisella henkilöllä, vaan käytettiin suoraan ensimmäisessä haastattelussa. Haastateltavat olivat kaikki yrityksen sisältä, joten oli mahdollista kysyä puuttuvia kysymyksiä jälkeensä. Haastatteluista pyrittiin tekemään mahdollisimman kattavia niiden vähäisestä määrästä johtuen. Kaikki haastattelut olivat kahdenkeskisiä keskusteluita, jotka pidettiin yrityksen tiloissa rauhallisessa paikassa lokakuun 2010 aikana.

Ensimmäinen haastattelu kesti noin yhden tunnin. Haastateltavalta, Paavolta kysyttiin kysymyksiä kaikista osioista. Tämän jälkeen haastattelurunkoa mietittiin uudestaan ja



joitakin pieniä korjauksia tehtiin. Toinen haastattelu pidettiin Antille. Noin puolitoistatuntia kestäneessä haastattelussa kysyttiin myös kaikkien osioiden kysymyksiä. Tässä haastattelussa tuli ilmi paljon samoja asioita kuin ensimmäisessäkin, etenkin tietovarastojen rakentamisen osalta. Tämän takia viimeisessä, Terhon haastattelussa kysyttiin alku- ja loppukysymysten lisäksi vain ensimmäisestä varsinaisesta osiosta, liiketoimintatiedon hallintaprojektista. Ennen haastattelua haastattelurunko tarkastettiin, mutta tehtiin vain pieniä muutoksia. Reilun tunnin kestänyt haastattelu keskittyikin tietovarastoinnin projekteihin ja ketterien menetelmien käyttämiseen.

Kukaan ei kieltäytynyt nauhoittamisesta, joten kaikki haastattelut saatiin äänitettyä. Jokaisesta haastattelusta otettiin myös muistiinpanot kannettavalle tietokoneelle. Haastattelurungon dokumentaatiota käsittelevä kolmas osio jäi haastatteluissa vähän vähemmälle, sillä asiat olivat jo melko pitkälti käsitelty ensimmäisessä ja toisessa osiossa. Tämän takia dokumentoinnin tulokset esitetään liiketoimintatiedon hallintaprojektin alla. Kysymyksissä oli myös hieman toistoa, erityisesti ensimmäisen ja toisen osion kohdalla. Tästä ei kuitenkaan aiheutunut ongelmia, sillä haastattelurunkoa käytettiin vain apuna. Siitä valittiin sopivat kysymykset, joita ei ollut käsitelty. Saadun palautteen mukaan haastattelut olivat kattavia ja sujuivat hyvin.

## **6.2. Haastattelujen tulokset**

### **6.2.1. Liiketoimintatiedon hallintaprojektit kohdeyrityksessä**

Liiketoimintatiedon hallinnan projekteista kysyttiin kaikilta haastateltavilta. Formaalit projektimallit eivät ole kovin paljon käytössä ainakaan tietovaraston rakentajan näkökulmasta. Paavo on toiminut paljon yhden asiakkaan projektissa heidän omissa tiloissaan, jolloin hän on käyttänyt asiakkaan työkaluja ja projektimalleja, tässä tapauksessa Scrum-menetelmää. Antin mukaan pelkät tietovarastointiprojektit toteutetaan yleensä vesiputousmallin mukaisesti, mutta kun tietovarasto toteutetaan CRM-projektin yhteydessä, käytetään CRM-projektin toimintatapoja, usein ketteriä menetelmiä ainakin jollakin tasolla. CRM-projekteissa käytetään tyypillisesti Scrumia. Terhonkin mukaan projekteissa käytetään vesiputousmallia, mutta sen lisäksi on käytetty myös syklisiä Hadden-Kelly-mallia. Tyypillistä on, että toteutetaan ensin perustiedot sisältävä tietovarasto ja sen jälkeen kehitetään sitä eteenpäin lataamalla lisää tietoa ja kehittämällä raportointia, eli kehitetään inkrementaalisissa osissa.

IPSS Oy:ssä koetaan, että ketterien menetelmien Scrum-malli ei ole vielä kovin hyvin hallinnassa liiketoimintatiedon hallinnan projekteissa. Scrumin toimivuudesta ja sopivuudesta tietovarastoinnin projekteihin ei ole täyttä varmuutta. Ketterien menetelmien pitäisi tuottaa asiakkaalle arvoa jokaisessa iteraatiossa, mutta varsinainen hyöty saadaan vasta valmiin tietovaraston ja raportoinnin valmistumisen jälkeen.

Terhon mukaan projekteissa tehdään kyllä oikeita asioita, mutta niitä ei saada valmiiksi suunnitellussa ajassa, kun matkan varrella tulee esiin uusia asioita ja hidasteita. Tietovarastoinnin projekteissa paljon ketteriä menetelmiä käyttänyt Paavo on sitä mieltä, että ketteristä menetelmistä on hyötyä, mutta niihin sisältyy myös haasteita. Hyötyinä on muun muassa se että tehdään vain lisäarvoa tuottavia asioita, tehtävien prioriteetti huomioidaan paremmin ja muutoksiin on helpompi reagoida. Toisaalta haasteita tulee, kun integraatorajapintoja on paljon tai tiimi on hajautunut. Tietojärjestelmähankkeet, kuten tietovarastointiprojektit riippuvat usein myös muista, projektin ulkopuolisista henkilöistä, joten haasteita seuraa tästäkin. Ohjelmistoprojektit, joihin ketterät menetelmät on alun perin suunniteltu, ovat modulaarisempia. Antti tuo esiin ketterien menetelmien hyötyinä tavoitteellisen tekemisen, tehdyn työn seuraamisen ja paremman kommunikoinnin liiketoiminnan työntekijöille. Scrum-menetelmä on myös melko kevyt käyttää. Dokumentaatiota syntyy jo luonnostaan ja sen päivitystarve on pienempi. Esimerkiksi Gantt-projektikaavion päivittäminen on hyvin työlästä, verrattuna Scrum-mallista syntyvään dokumentaatioon. Siinä dokumentoidaan aina mitä on tehty ja mitä aiotaan tehdä kussakin iteraatiossa. Näitä ei tarvitse päivittää, toisin kun ajankohtaista tilannekuvaa.

Asiakkaiden mielenkiinnon kohteina tietovarastointiprojekteissa on Terhon mukaan tuottava analysointikyky, mutta toisaalta halutaan vain valmiit mittarit ja raportit. Toive olisi usein enemmän, mutta kustannukset karsivat vain olennaiseen. Asiakkaat haluavat usein myös visualisointia mahdollisimman varhaisessa vaiheessa, mutta tietovaraston rakenne tulisi olla kuitenkin hyvin suunniteltu ja valmiiksi tehty. Tietovarastointiprojektit kestävätkin tyypillisesti muutamista viikoista useisiin kuukausiin, vaikka jokainen projekti on eripituinen. Konsulttiyhtiön puolesta toki toivotaan pitkiä projekteja, mutta asiakkaan liiketoiminnasta vastaavat vaatii nopeampia ja tulosta tuottavia projekteja.

Eniten resursseja kuluu selvästi ETL-rutiinin tekemiseen. Antin mukaan tiedon latauksiin, tiedon laatuun ja lähdedatan käsittelyyn kuluu noin 60–70 %, loppujen kuluessa muun muassa tietomallin tekemiseen, projektinhallintaan ja visualisointiin. Paavo on myös samoilla linjoilla; noin 60 % käytetyistä tunneista kuluu ETL:n suunnitteluun ja toteutukseen, loput 40 % vaatimusmäärittelyyn, arkkitehtuuriin, sovelluksen valintaan, tietomallinnukseen ja testaukseen.

Projekteissa toimii projektipäällikön lisäksi arkkitehti, toteuttajia, testaaja sekä ohjausryhmä, jossa vähintään yksi kokeneempi työntekijä sekä asiakkaan että toimittajan puolelta. Ohjausryhmässä tulisi olla määrittelykykyä, sillä he laativat vaatimukset. Projektin koosta riippuen toteuttajia voi olla hyvinkin monta ja testaajia useampi. Projektipäällikkö voi toisinaan olla toteuttamassa, vaikka se ei olekaan hyvä käytäntö. Toisinaan resurssien puutteen takia voi myyjä, projektipäällikkö ja toteuttaja olla jopa sama henkilö. Jo ylläpidon kannalta on tärkeää, että projektista on hyvin

perillä vähintään kaksi henkilöä, mielellään jopa toteuttamassa. Silloin ylläpito ja jatkokehitys ei ole täysin yhdestä työntekijästä riippuvainen.

Projektin alkaessa konsultit toteuttavat asiakkaan liiketoiminnan kanssa määrittelydokumentit. Siihen tehdään muun muassa dimensiomallit, kirjataan lähdejärjestelmät ja kuvataan yleisen tason arkkitehtuuri. Tämän jälkeen siirrytään teknisempään rooliin ja kuvataan välivarastotietokannat ja tietovarasto sekä tiedonlataukset. Toteutuksen ohessa syntyy dokumentaatiota, esimerkiksi visuaaliset ETL-välineet ovat jo sinällään dokumentteja. Lisäksi projekteissa syntyy ylläpitodokumentteja, liiketoimintatiedon hallintaympäristön konfigurointidokumentteja, toipumissuunnitelma, tietomallien dokumentointia, käyttöohjeita ja koulutusmateriaaleja. Paavo painottaa itsedokumentoivaa ETL:ää ja dokumentaation automatisointia, sillä dokumentointi on usein työlästä ja vaatii lähes jatkuvaa päivitystä. Antin mielestä asiakastuki- ja ylläpitodokumentit ovat tärkeimpiä. Hän sanoo hieman karrikoiden, että dokumentointi ei ole ikinä riittävällä tasolla, eikä koskaan ajan tasalla. Kuitenkaan dokumentoinnissakaan ei saa tehdä ylilyöntejä, sillä se ei ole pääasia.

### **6.2.2. Tietovaraston rakentaminen kohdeyrityksessä**

Tietovaraston ja raportointiympäristöjen rakentamisesta kysyttiin vain Paavolta ja Antilta. Tietovaraston rakentamisessa on käytetty paljon erilaisia työkaluja, Paavo on käyttänyt pääasiassa Oraclen tuotteita, Warehouse Builderia ja Data Integratoria, mutta myös Microsoftin SSIS-paketteja. Antti on käyttänyt muun muassa Microsoftin DTS (Data Transform Services), SSIS:ää, Oraclea, Cognosta ja nykyään enemmän Pentahoa. Asiakkaan toiveita kuunnellaan ja tietovarasto voidaan rakentaa heidän omilla työkaluillaan. Tiettyihin tuotteisiin onkin IPSS:llä sitouduttu kehittämään osaamista. Pentaho on kuitenkin vakiintunut omassa käytössä, ja jos työkalulla ei ole asiakkaalle merkitystä. Ohjelmistot ovat myös melko erilaisia ja oikeilla välineillä tietovarasto syntyy nopeammin. Työkalujen valintaan vaikuttaa erityisesti oma osaaminen ja työkalun käytettävyys, mutta voi vaikuttaa myös projektin koko. Paavo käyttäisi mieluiten Oraclea, sillä hänellä on siitä paljon kokemusta, kun taas Antti Pentahoa tai Microsoftin tuotteita riippuen projektista.

Tietovarastointiprojekteissa on paljon vakiintuneita käytäntöjä muun muassa tiedostorakenteet ja tiedostojen, tietokannan taulujen ja sarakkeiden nimeämiset. Lippuja käytetään esimerkiksi rivien tunnistamisessa ja merkitsemisessä. Välivarastotietokannan rivejä merkitään lipuilla käsitellyiksi tai virheellisiksi. Yksi vakiintunut käytäntö on myös se, että data ladataan lähdejärjestelmistä sellaisenaan välivarastokantaan, jonka jälkeen tehdään muutokset ennen päivittämistä tietovarastoon. Tämä parantaa Antin mukaan tietovaraston eheyttä, sillä usein datasiirtojen kaatumiset aiheutuvat verkkoyhteysvirheistä. Dimensiomallinnuksesta on tullut myös vahva käytäntö, vaikkakin mallinnus suunnitellaan tapauskohtaisesti.

ETL-rutiineja ei yleensä voi kopioida sellaisenaan toiseen projektiin, ainakaan useimmissa työkaluissa. Työtapoja, kokemusta ja parhaita käytäntöjä hyödynnetään kuitenkin tulevissa projekteissa. Tuotelähtöisessä BI:ssä tulisi kuitenkin hyödyntää paljonkin, mutta lähdejärjestelmien tulisi pysyä samana. Silti siirtorutiinit tulisi käydä hyvin läpi halutun lopputuloksen saamiseksi.

Tietovarastot eivät perinteisesti ole kovin joustavia muutoksille. Riippuu kuitenkin muutoksesta: tiedon lisäämisestä aiheutuvat muutokset ovat nopeita toteuttaa, mutta tietovaraston ytimeen, faktatauluihin tai rakenteeseen kohdistuvat muutokset tuovat paljon lisätyötä siirtorutiineihin, tietokantoihin ja raportointiin. Hyvin mallinnettu ja suunniteltu rakenne vähentää muutoksista aiheutuvaa lisätyötä. Muutoksia tulee kuitenkin verrattain harvoin, kun lähdejärjestelmät ovat samoja. Lähdejärjestelmien muuttuessa siirtorutiinit tulee rakentaa yleensä kokonaan uudestaan, versiovaihdoksissa riittää usein siirtorutiinien päivitys. Automatisointia tehdään yleensä vain kun on välttämätöntä, kuten ennalta määritellyille kuukausittaisille tiedon lisäyksille. Muuttuvista lähdejärjestelmistä tai tietosisällöistä kommunikoidaan, jolloin muutokset tehdään vasta tarvittaessa.

Ylläpito koostuu muun muassa tiedonsiirron valvonnasta, pääkäyttäjän roolin tehtävistä ja levytilan valvonnasta. Tietovaraston päivitys epäonnistuu toisinaan johtuen virheellisistä siirtotiedostoista, tiedonsiirtorutiinin virheistä, tai verkkoyhteysvirheistä. Pääkäyttäjän rooliin kuuluu tyypillisesti käyttäjätunnusten ja salasanojen hallinnointi ja pääsyoikeudet. Automaattiset valvontaviestit mm. levytilan valvonnasta ja tiedonsiirtorutiinien tuloksista sekä lokitiedostot helpottavat ja vähentävät ylläpitoon liittyvää työtä. Ylläpitoon kuluu vaihtelevasti aikaa, mutta arviolta noin yksi työpäivä kuukaudessa projektista riippuen kuluu aikaa ylläpitotehtäviin.

### **6.2.3. Ketterä liiketoimintatiedon hallinta ja tulevaisuudennäkymät**

Agile BI, eli ketterä liiketoimintatiedon hallinta nähdään toisaalta ketterien menetelmien hyödyntämisenä liiketoimintatiedon hallintaprojekteissa, mutta toisaalta BI-ympäristön joustavuutena muutoksille. Ketterän BI-ympäristön tulee reagoida muutoksiin nopeasti. Muutokset voivat olla muuttuvat lähdedata tai muuttuvat liiketoiminnan, eli BI-ympäristön käyttäjien vaatimukset. Terhon mielestä ketterä liiketoimintatiedon hallinta on tapa tehdä projekteja, ja erityisesti ketterä tapa tehdä töitä. Siihen liittyy tietynlainen mittarointi, tekemisen jatkuva seuraaminen. Hänen mielestään ketteryys ei tarkoita sitä, että voidaan reagoida mihin vain muutoksiin, vaan tiimi valitsee mihin muutoksiin reagoidaan milläkin aikavälillä. Liiketoiminnan näkökulmasta ketterä liiketoimintatiedon hallinta tarjoaa ketterästi vastauksia mitä saadaan tietovarastosta. Ketterä tietovarastointi kuvaa kuinka joustava se on ad hoc -kyselyille ja kuinka hyvin se antaa vastauksia etsittyihin kysymyksiin, ei niinkään datan muuttumiseen reagoiminen.

Antin mielestä IPSS Oy:n vahvuudet on hyvät työkalut, stabiili ja avoin CRM, mutta toimintaa tulisi kehittää ottamalla käyttöön entistä enemmän ketteriä menetelmiä ja vakiinnuttaa omia projektimenetelmiä. Paavo kehittäisi erityisesti dokumentointia, sen tulisi olla mahdollisimman helposti päivitettävää, yksi asia yhdessä paikassa ja pitkälti automatisoitua. IPSS Oy:n tulevaisuudennäkymät nähdään silti oikein hyvänä, sillä asiakasjohtamisen malli nojaa vahvasti liiketoimintatiedon hallintaan.

Liiketoimintatiedon hallinta nähdään tulevaisuudessa muuttuvan yhä yleisemmäksi, helpommaksi toteuttaa ja tehokkaammaksi. Paavo ennustaa isojen ohjelmistotoimittajien tuottavan yhä enemmän valmiita ratkaisuja, mikä vähentää tarvetta räätälöintiin. Kehityksessä on pysyttävä mukana, jotta pärjää kiristyvässä kilpailussa. Automatisointi tulee vähentämään työtä ja reaaliaikaisuus tulee jo muutamassa vuodessa yleistymään. Tietovaraston takaisinkytkentä muihin järjestelmiin tulee myös yleistymään, tiedonsiirrosta tulee kaksisuuntaista.

Antti näkee erillisten tietovarastojen, data marttien eli tietokomerojen mahdollisuudet ja uskoo niiden yleistyvän. Yhden, kaiken datan sisältävän tietovaraston rooli ja välttämättömyys jatkossa on kyseenalaista. Operatiivinen ja analyttinen tieto tulee olemaan nivottu toisiinsa. Muistinvaraisten tietojen analysointi, tai suoraan datavirrasta analysointi tulee mahdolliseksi, jolloin suuria tietokantoja ei enää tarvita. Tämän seurauksena raportoinnin nopeus paranee huomattavasti.

Terhon mielestä yritykset eivät vielä saa tiedostaan kaikkea irti, mikä olisi mahdollista saada. Tietovarastojen hyödyntäminen on heikkoa ja asioihin reagoidaan viiveellä. Toiminta tulee muuttumaan reaktiivisesta proaktiiviseen, reagoivasta ennakoivaan. Mittaristojen seuraamisessa tullaan hyödyntämään koneita enemmän. Hälytysrajat tehdään automaattiseksi, säännöstä kehitetään ja ennustettavuus paranee. Välineitä erityisesti liiketoimintaan tulee olemaan, joilla saadaan parempaa ymmärrystä suurista tietomääristä. Tämä vaatii tietotekniikalta paljon yhteistyötä liiketoiminnan kanssa.

### **6.3. Yhteenveto haastatteluista**

Haastateltavat olivat tyytyväisiä haastatteluihin ja haastattelujen määrän vähäisyydestä huolimatta niistä saatiin paljon tietoa irti. Haastateltavien näkemykset olivat pääosin hyvin yhdenmukaisia. Haastatteluista syntyi melko hyvä näkemys liiketoimintatiedon hallinnan tilasta erityisesti kohdeyrityksessä, mutta myös alalla yleisesti. Kiinnostus ketterien menetelmien soveltamiseen alalla on kasvussa niiden suomien mahdollisuuksien ansiosta. Lisäksi toimintaa halutaan tehostaa ja vastata liiketoiminnan yhä nopeammin muuttuviin vaatimuksiin.

IPSS Oy:llä on käytetty tietovarastointiprojekteissa erilaisia malleja, mikä on sinänsä hyvä asia. Näin eri toimintatapojen hyötyjä ja haittoja on päässyt näkemään. Toimintatapoja olisi hyvä kuitenkin vakiinnuttaa ja valita jokin projektimalli, jonka pohjalta toimintaa suunnitellaan. Ketterien menetelmien hyötyjä ja mahdollisuuksia on pohdittu ja Scrum-menetelmä on tullut tutuksi. Scrumia voisi soveltaa täysipainoisesti johonkin tietovarastointiprojektiin, ja mahdollisesti sen jälkeen vakiinnuttaa toimintatavaksi kaikkiin projekteihin, jos se koetaan hyväksi.

Dokumentointia tulisi kehittää riittävän kattavaksi kaikkiin tarkoituksiin, mutta silti mahdollisimman pieneksi. Ylimääräinen dokumentointi on turhaa työtä, vaikka voi olla vaikea sanoa missä menee raja. Dokumentointi on tehtävä sellaiseksi, että sen päivitystarve on mahdollisimman pieni. Dokumentoinnin automatisointi ja itsedokumentoiva ETL vähentää dokumentointiin kohdistuvan työn tarvetta. Dokumentaatio on myös suunniteltava niin, että sitä ei tarvitse jatkuvasti päivittää manuaalisesti. Dokumentoinnin tulisi olla myös yhtenäistä projektien välillä ja sen on oltava myös helposti saatavilla.

Suurin potentiaali toiminnan tehostamisessa on pienentää ETL-rutiineihin käytettyä aikaa, sillä se kuluttaa projekteissa eniten resursseja. Hyvät ja tehokkaat ETL- ja raportointityökalut ovat tärkeässä asemassa tietovarastointiprojekteissa ja niiden monipuolinen kokemus ja osaaminen ovat hyödyksi. Valitsemalla oikeat työkalut rakentamiseen ja ylläpitoon kuluva työmäärä saadaan selvästi pienennettyä. Parhaita käytäntöjä tulisi vielä etsiä ja vakiinnuttaa. ETL-rutiinien automatisointi on kohdeyrityksessä ja alalla yleisemmin kiinnostuksen kohteena, joten sitä tulisi kehittää mahdollisuuksien mukaan.

Liiketoimintatiedon hallintaympäristöistä tulisi rakentaa ketterämpiä muutoksille. Muutosten mahdollisuus tulisi ottaa huomioon kaikessa tekemisessä. Muutosvalmiutta tulisi kohottaa sekä datan ja lähdejärjestelmien osalta, mutta myös liiketoiminnan vaatimusten osalta. Ylläpito tulisi huomioida tekemisessä, jotta sen vaatimaa työmäärää voidaan vähentää. Tutun sanonnan mukaan hyvin suunniteltu on puoliksi tehty.

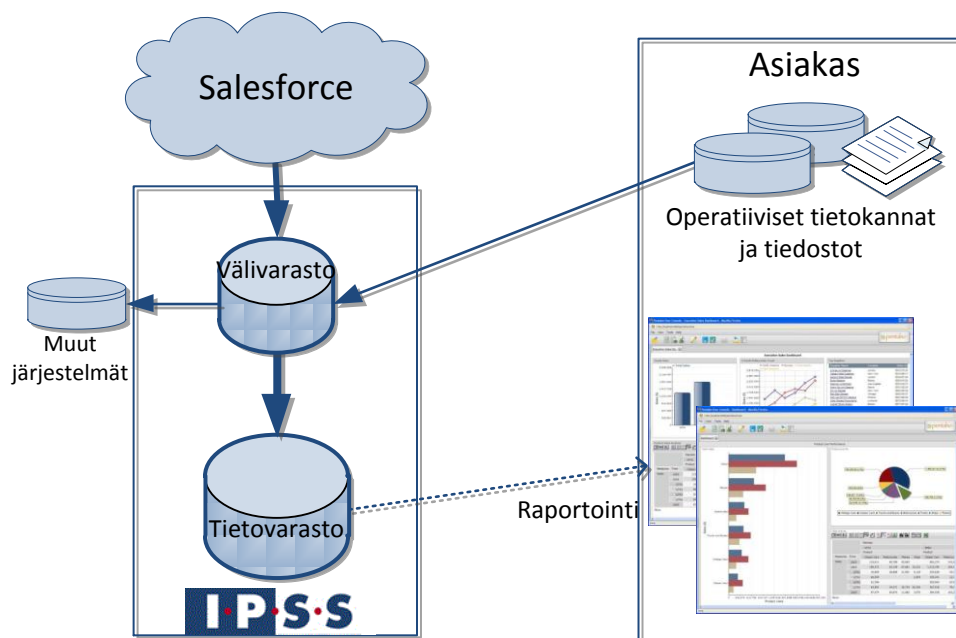
IPSS Oy:llä on hyvät tulevaisuudennäkymät, vaikka kehitettävää on aina. Koko ala on nopeasti kehittynyt ja tulee kehittymään edelleen. IPSS Oy:n on jatkossakin pysyttävä kehityksen kärjessä mukana, jotta pysyy kilpailukykyisenä.

## 7. KÄYTÄNNÖN SOVELLUS

Tässä luvussa esitellään työssä toteutettu käytännön sovellus, ketterä tiedonlataus. Luvun aluksi esitellään kohdeympäristö, johon käytännön sovellus on toteutettu. Toisessa alaluvussa kerrotaan toteutetusta ohjelmasta ja sen toiminnasta. Lopuksi sovelluksesta saatuja hyötyjä ja haasteita pohditaan viimeisessä alaluvussa.

### 7.1. Kohdeympäristön kuvaus

Kohdeympäristöön kuuluu IPSS Oy:n hallinnoima tietovarasto, johon ladataan tietoa ETL-prosessilla asiakashallintajärjestelmästä Salesforcesta (kuva 7.1). Tietovarastoon voidaan ladata dataa myös asiakkaan omista operatiivisista järjestelmistä, kuten toiminnanohjausjärjestelmästä, laskituksen, varaston tai henkilöstöhallinnon järjestelmistä sekä yksittäisistä tiedostoista. Tiedot ladataan Salesforcesta ja asiakkaan operatiivisten järjestelmien tietokannoista ensin sellaisenaan välivarastotietokantaan, joka sijaitsee tyypillisesti samalla palvelimella kuin tietovarasto. Tietovarastoa päivitetään ETL-rutiinilla välivarastosta.



Kuva 7.1 Yleiskuva kohdeympäristöstä

Välivarastotietokanta toimii solmukohtana eri järjestelmien välillä. Tietovarastoon ladataan kaikki data sen kautta, mutta sieltä voidaan ladata dataa myös asiakkaan muiden järjestelmien käyttöön. Tietoa voidaan siirtää myös molempiin suuntiin, vaikkakin pääosin tieto ladataan yksisuuntaisesti eri järjestelmistä tietovarastoon. IPSS

Oy tarjoaa asiakkaalle raportointia ja analysointia tietovarastosta palveluna. Mittaristoja ja raportteja voidaan asiakkaan halutessa upottaa myös Salesforceen, tai muihin web-pohjaisiin järjestelmiin. Raportointi on mahdollista toteuttaa myös osin välivaraston tietoon perustuen. Tällöin raportointi kohdistuu vain yhden järjestelmän dataan, mutta on mahdollista toteuttaa jopa reaaliaikaisesti.

### **7.1.1. Salesforce**

Salesforce tarjoaa asiakashallintapalvelua (CRM) yrityksille ja yhteisöille maailmanlaajuisesti. Salesforce on perustettu vuonna 1999 ja yrityksen pääkonttori sijaitsee San Franciscossa, Kaliforniassa. Salesforce tarjoaa asiakkuudenhallintajärjestelmää palveluna (Software as a Service, SaaS) sekä teknologia-alustana (Platform as a Service, PaaS). (Salesforce 2010, ss. 1–2.) Asiakkaat ja sovelluskehittäjät voivat rakentaa ja käyttää myös omia sovelluksiaan teknologia-alustan päällä.

Salesforce on siitä erikoinen yritys, että sillä on asiakkainaan paljon jopa yhden hengen yrityksiä. Tämä johtuu siitä, että heidän ei tarvitse ostaa kalliita palvelimia, järjestelmiä ja tietokantoja. (Ihanus 1999.) Salesforcen asiakkaat maksavat lisenssimaksua käyttäjämäärän mukaan, yhdestä käyttäjästä kymmeneen tuhansiin (Salesforce 2010, s. 9). Tämä mahdollistaa kaikenkokoiset asiakkaat käyttämään Salesforcea. Salesforce valittiin vuoden 2010 kymmenen parhaan pilvipalvelun (engl. cloud computing) joukkoon (Williams 2010). Tämä kertoo sen hurjasta suosioista ja menestyksestä.

### **7.1.2. Pentaho Data Integrator – PDI**

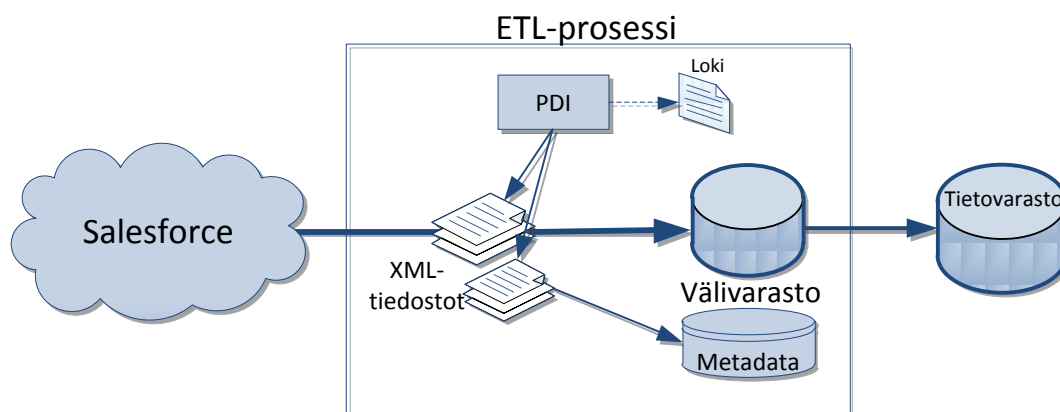
Pentahon tarjoama BI-tuotepaletti on maailmanlaajuisesti käytetyin avoimen lähdekoodin business intelligence -ratkaisu (Kauppi 2010). Pentaho Data Integrator (PDI), toiselta nimeltään Kettle on graafinen ETL-työkalu. PDI on kehitetty vastaamaan paremmin liiketoimintakäyttäjien muuttuvia tarpeita. Lisäksi se vähentää eri työkalujen määrää sisältäen kaiken tarpeellisen ETL:n, visualisoinnin, sisäänrakennetun raportointi- ja analysointimoottorin. (Worthington 2010.)

IPSS Oy:ssä käytetään muun muassa PDI:tä ETL-prosessien toteuttamisessa. Se on avoimen lähdekoodin ansiosta ilmainen sekä lisäksi koettu tehokkaaksi ja monipuoliseksi. PDI:n sovelluksella Spoon:illa luodaan graafisesti datan käsittelyrutiinit, joista syntyy KTR- ja KJB-tiedostoja. Tiedostot ovat XML-koodia (eXtensible Markup Language), joita ajetaan ajastetusti Kitchen- ja Pan-sovelluksilla. KTR-päätteiset tiedostot ovat Kettlen transformaatioita, jotka lukevat tiedostoja ja tietokantoja (esim. operatiivisten järjestelmien tietokantoja), käsittelevät datavirtaa ja suorittavat tarvittavia muutoksia ja tallentavat tietoa muun muassa tietokantoihin ja tiedostoihin. KJB-päätteiset tiedostot ovat Kettlen jobeja, joiden avulla päästään käsiksi toisiin jobeihin tai transformaatioihin, ja määräävät niiden järjestyksen.



### 7.1.3. ETL-prosessi kohdeympäristössä

Kehityksen kohteeksi valittu ETL-prosessi lataa tietoa Salesforcesta ja tallentaa välivarastoon kuvan 7.2 mukaisesti. Lisäksi se tallentaa metadataa tietokantaan, josta se on tarvittaessa helposti saatavilla. ETL-rutiinia ajetaan PDI:llä, josta syntyy lokia. Lisäksi PDI lähettää sähköpostia prosessin tuloksista. Näin on helpompi seurata, miten ETL-prosessi käyttäytyy vähentäen ylläpidon työtä.



Kuva 7.2 ETL-prosessi Salesforcesta välivaraston kautta tietovarastoon.

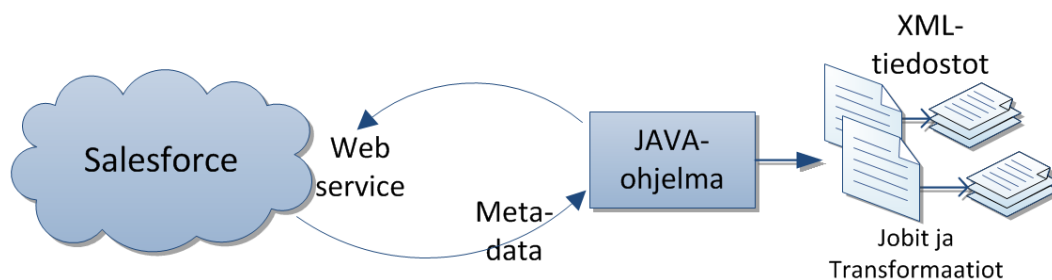
Kyseinen ETL-prosessi on IPSS Oy:lle tärkeä. Samanlaista prosessia voidaan käyttää useiden asiakkaiden ympäristöissä sekä omassa käytössä, sillä Salesforce on paljolti käytössä. Salesforceen rakenne on samanlainen kaikissa ympäristöissä, mutta pieniä muutoksia tulee silti jatkuvasti joustavan räätälöinnin ansiosta. ETL-prosessi on silti yksinkertainen, kantaan halutaan samat tiedot samanlaisena kuin Salesforcesta. Se ei myöskään sisällä asiakaskohtaista räätälöintiä niin paljon kuin tietovaraston päivitys välivarastosta. Siksi tätä ETL-prosessia, Salesforcesta välivarastoon halutaan automatisoida ja tehdä joustavammaksi muutoksille.

## 7.2. Tiedon latauksen automatisointi

Käytännön sovelluksen tarkoituksena oli automatisoida ETL-prosessin rakentaminen, ja tehdä siitä erittäin joustava muutoksille. Tavoitteena oli aikaansaada parametrisoituvaa sovellusta, joka vähentäisi huomattavasti BI-ympäristön rakentamiseen ja ylläpitoon kuluva työtä sekä toisi muutokset läpinäkyväksi metatiedon esittämisen ansiosta.

### 7.2.1. Ohjelman toiminta

Java-ohjelma (lähdekoodi luottamuksellinen) käyttää Salesforceen web service -rajapintaa (ks. kuva 7.3). Ohjelma kutsuu Salesforceen funktioita, jotka palauttaa muun muassa metatietoja Salesforceen tietokannasta. Metatietojen ovat muun muassa Salesforceen objektien nimet ja kentät, joiden perusteella tietokantaan luodaan taulut ja niiden kentät.



Kuva 7.3 Java-ohjelman peruseräite.

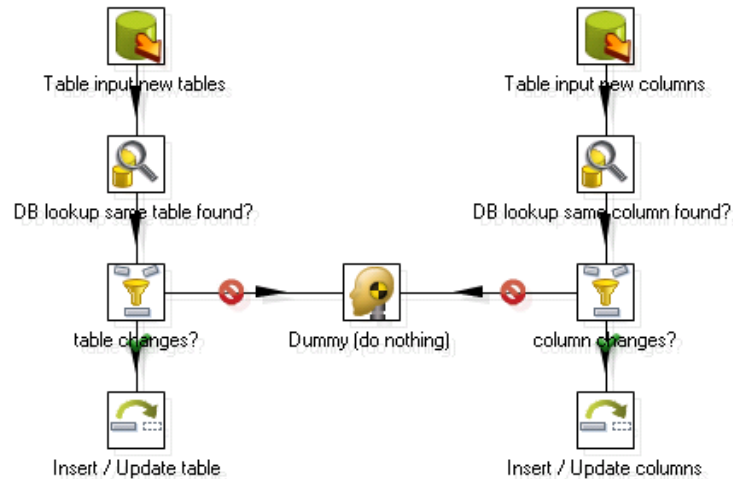
Aluksi ohjelma lukee tarvittavat parametrin properties-tiedostosta, kuten Salesforceen tunnuksen ja salasanan, projektin nimen tiedostojen nimeämistä varten sekä halutut objektit perusobjektien lisäksi. Seuraavaksi annetuilla tunnuksilla kirjaututaan Salesforceen ja haetaan kaikkien objektien nimet. Ohjelma käy läpi jokaisen objektin, tarkastetaan onko se luettavissa ja valitsee halutut objektit jatkokäsittelyyn. Tämän jälkeen käydään läpi kunkin halutun objektin kaikki kentät, kysytään niiden tietokantanimet, nimet käyttöliittymässä, tietotyypit ja pituudet. Lopuksi ohjelma tulostaa kaksi KTR-tiedostoa kustakin objektista, metadatan käsittelystä muutama KTR-tiedosto sekä kaksi KJB-tiedostoa, jotka kokoavat yhteen aiemmat transformaatiotiedostot. Kaikki tiedostot ovat tulostettu XML-kielellä.

### 7.2.2. Generoituvat tiedostot ja niiden toiminta

Syntyneistä KJB-tiedostoista ensinnä ajetaan SQL-skriptejä sisältävä job. Sen ensimmäisessä transformaatioissa luodaan tietokantaan varsinaiset ja väliaikaiset metadatataulut sekä objekteille että kentille, mikäli tauluja ei ole olemassa. Tämän jälkeen tyhjennetään väliaikaiset metadatataulut. Seuraavaksi job sisältää transformaatiot jokaiselle Salesforceen objektille. Kustakin löytyy SQL-skripti, joka luo väliarastotietokantaan taulun objektille, mikäli sitä ei vielä ole sekä lisää objektin tiedot väliaikaiseen metadatatauluun. Tämän lisäksi objektin jokainen kenttä käydään läpi, lisätään kyseisen objektin tauluun uusi sarake sekä tallennetaan metatietoja väliaikaiseen metadatatauluun. Objektien kaikki merkkijonoja sisältävät kentät ovat tietotyypiltään vaihtuvamittaisia merkkijonoja (varchar), joiden maksimipituudeksi on asetettu 512 merkkiä. Pidemmät kentät leikataan aina maksimipituuteen. Kyseinen pituus on todettu riittäväksi, sillä harvoin raportoinnissa halutaan pitkiä tekstikenttiä näkyville. Maksimipituutta voidaan kuitenkin tarvittaessa kasvattaa, jos se koetaan tarpeelliseksi. Tietotyyppiä ei voi muuttaa lähdejärjestelmässä, joten sen muuttamista ei tarvitse tarkastaa. Tietotyypin pituuteenkaan ei tarvitse ottaa kantaa sillä, ne leikataan maksimipituuteen.

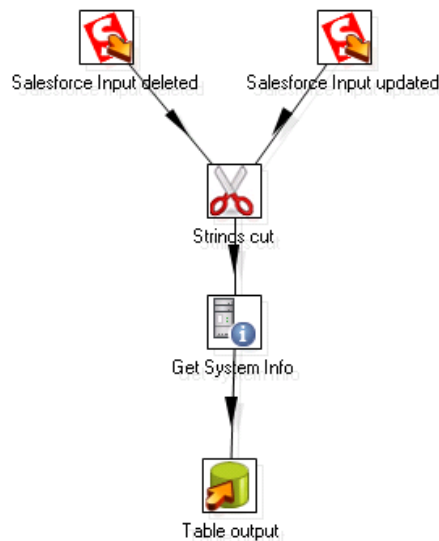
Jobin kaksi viimeistä käsittelee varsinaisia ja väliaikaisia metadatatauluja. Ensinnä tarkastetaan onko varsinaisessa taulussa sellaista tietoa, mitä väliaikaisessa ei ole. Jos tällaista tietoa on, ne ovat poistuneet joten niihin riveihin asetetaan poistopäivämäärä. Rivejä ei poisteta kuitenkaan kokonaan tietokannasta, jotta nähdään mitä on poistettu ja

milloin. Tämän jälkeen tuodaan uusi metadata väliaikaisesta metadatataulusta varsinaiseen metadatatauluun. Aluksi luetaan rivit väliaikaisista tauluista (kuva 7.4), tarkastetaan löytyykö täsmälleen samanlaiset tiedot varsinaisista metadatatauluista. Jos löytyy, tieto on ajan tasalla eikä vaadittavia toimenpiteitä ole. Mikäli jokin on muuttunut, tai tietoa ei ennestään ole, päivitetään se varsinaiseen metadatatauluun. Tässä kohtaa päivitetään myös päivitysaika, mutta ei kuitenkaan luontiaikaa. Uusien rivien kohdalla luontiaika merkitään samaksi kuin päivitysaika.



Kuva 7.4 Metadatan käsittely -transformaatio.

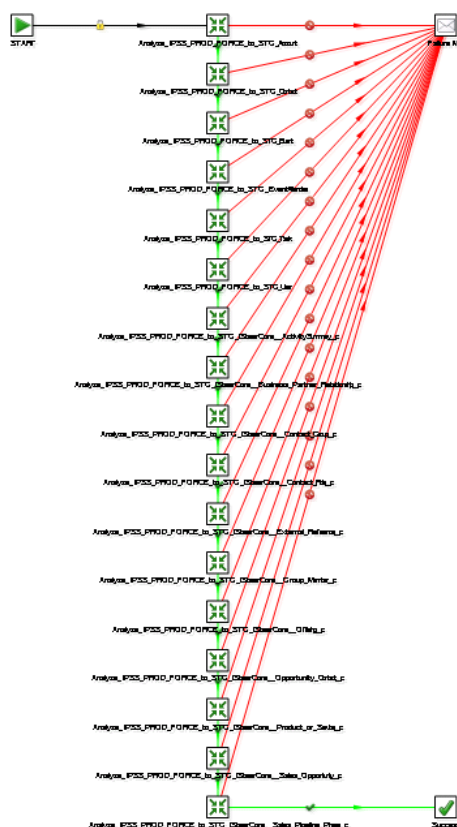
Toinen job sisältää varsinaisen datan lataamisen Salesforcesta paikalliseen tietokantaan. Se sisältää yhden transformaation jokaiselle objektille. Transformaation ensimmäinen askel on tiedon lataus Salesforcesta kuvan 7.5 mukaisesti. Mikäli halutaan ladata myös poistetut tiedot, sekä objekti on poistettavissa, Salesforce input -askelmia on kaksi. Toinen hakee poistetut rivit tietyltä aikajaksolta ja toinen hakee joko päivitettyt rivit samalta aikajaksolta tai kaikki.



Kuva 7.5 Tiedon lataus -transformaatio.

Java-ohjelmalle on annettu parametreissa tieto myös siitä, halutaanko ladata poistettuja ja halutaanko ladata kaikki tiedot vai päivitettyt. Näihin molempiin askeliin on määritelty käyttäjätunnusten lisäksi haettava objekti sekä kaikki sen kentät ja niiden tietotyypit. Transformaation seuraava askel leikkaa kaikkien merkkijonokenttien pituudet määriteltyyn maksimipituuteen, jotta lataus tietokantaan onnistuu. Toiseksi viimeisessä kohdassa pyydetään järjestelmän kellonaikaa ja päivämäärää, jotta tietokantaan saadaan lisäyspäivämäärä. Lopuksi rivit tallennetaan tietokannan tauluun.

Tiedon lataus -job sisältää kaikki yllä mainitut tiedonlataus-transformaatiot (ks. kuva 7.6). Transformaatiot käydään läpi yksi kerrallaan, ja virheen sattuessa lataus keskeytetään ja lähetetään virheilmoitus. Job tarvitsee parametreikseen JNDI-yhteyden (Java Naming and Directory Interface) nimen, alku- ja loppupäivämäärän sekä sähköpostitiedot. Myös toinen job, joka käsittelee metatietoja ja luo tietokantaan tarvittavat taulut, tarvitsee JNDI:n nimen ja sähköpostitiedot parametreiksi.



Kuva 7.6 Tiedon lataus -job, joka sisältää useita transformaatioita.

JNDI-yhteys on määritelty PDI:n tietokantayhteyksiin, joka muutetaan ODBC-yhteydeksi (Open Database Connectivity). Alku- ja loppupäivämääriä tarvitaan puolestaan ladattaessa poistettuja ja päivitettyjä tietoja Salesforcesta. Sähköpostitietoihin kuuluu sähköpostipalvelimen nimi, lähettäjän nimi ja osoite sekä vastaanottajan sähköpostiosoite. Näitä tietoja tarvitaan sähköpostin lähetyksessä, jos jokin epäonnistuu. Sähköpostia voidaan lähettää myös, kun koko ETL-rutiini on suoritettu onnistuneesti. Tällöin sähköpostin lähetyksessä tulee ns. master jobiin, joka sisältää

generoidut jobit ja parametrien asettamisen. Master jobia voidaan sitten ajaa ajastetusti, esimerkiksi joka yö.

### 7.3. Tulokset

Java-ohjelmasta tuli tehokas. Ohjelman suorittaminen kesti vain puoli minuuttia ja se tulosti vajaa neljäkymmentä tiedostoa testiympäristöstä. Ohjelma voidaan käynnistää ajastetusti esimerkiksi joka yö, jolloin se kirjoittaa uudet tiedostot entisten päälle. Näin vältetään turhilta tiedostoilta, mutta saadaan automaattisesti muutokset tietokantaan asti sekä poisto- ja lisäyspäivämäärät metadatatauluihin.

SQL-koodia sisältävät metadatan käsittelyt sekä tietokannan taulujen luomiset ja muutokset sujuivat myös tehokkaasti. Jobin suoritus kesti vain kymmeniä sekunteja, huolimatta siitä oliko taulut luotuja valmiiksi vai ei. Metadatatauluihin syntyi rivi jokaista objektia ja jokaista kenttää kohti. Lisäys-, päivitys- ja poistopäivämäärät kertovat tärkeää informaatiota muutoksista. Päivityksestä puuttuu vain tieto, mikä on päivittynyt, mutta kenttiä ei ole paljon. Käytännössä päivityksiä ei juuri tule, mutta uusia kenttiä kyllä tulee.

Datan lataamisen kesto riippuu täysin datan määrästä. Testiympäristössä on useimmissa objekteissa useita tuhansia rivejä dataa, mikä vastaa tavallista tuotantoympäristöä, kesti kaiken datan lataamisessa noin viisitoista minuuttia. Kun taas ladataan ainoastaan päivitettyt tiedot, kesto putoaa jo vajaaseen pariin minuuttiin.

Java-ohjelma vaatii toimiakseen muun muassa Salesforcen JAR-tiedoston (Java ARchive), jonka luokat otetaan käyttöön import-komennolla lähdekoodissa. Kyseinen tiedosto löytyy myös PDI:stä, sillä siihen on sisäänrakennettu Salesforce-rajapinta. Kettlen transformaatiot ja jobit generoivassa Java-ohjelmassa tulee olla sama versio kyseisestä Salesforce.jar-tiedostosta, kuin Pentaho Data Integratorissa. Muutoin metadata saattaa näyttää hieman erilaiselta, eikä datan lataaminen toimi. Salesforcen JAR-tiedostosta tulee aika ajoin uusia versioita, jotka otetaan käyttöön myös PDI:hin. Tällöin Java-ohjelmalle pitää antaa sama JAR-tiedosto sekä tarkastaa yhteensopivuus lähdekoodissa. PDI:sta tulee myös uusia versioita, joten yhteensopivuus vanhojen transformaatioiden kanssa tulee tarkastaa ennen niiden käyttöönottoa. Tämä tuo hieman lisätyötä, mutta useimmiten ohjelmistoja päivitetään niin, että yhteensopivuus aiempien versioiden kanssa säilyy.

Hidasteita tuli myös Java-ohjelman testaamisessa, sillä PDI:stä löytyi ohjelmistovirheitä, niin sanottuja bugeja. PDI on avoimen lähdekoodin ohjelmisto, joten mitään takeita sen toimivuudesta ei ole. Lähdekoodi on kuitenkin kaikkien saatavilla, joten ohjelman toiminnallisuutta voi sieltä tarkastaa. Pentahon Internet-sivuilla otetaan

vastaan virheilmoituksia ja kehitysideoita, joten viestiä välittämällä virheetkin useimmiten korjataan nopeasti.

ETL-rutiinista ei tullut täysin automaattista, sillä syntynyt rutiini täytyy sisällyttää suurempaan kokonaisuuteen, master jobiin. Automaattisesti syntyvät jobit vaativat parametreja, jotka välitetään toisesta jobista. Tämä master job olisi voitu myös generoida automaattisesti Java-ohjelmalla, mutta toisaalta se täytyy luoda vain kerran ja se on erittäin helppo ja nopea tehdä. Toisinaan jobit halutaan sisällyttää johonkin suurempaan kokonaisuuteen joka tapauksessa.

Tiedot ladataan Salesforcesta välivarastotietokantaan sellaisenaan, joten tietokannan koko kasvaa jatkuvasti. Erityisesti, jos aina haetaan kaikki tiedot ja ETL-prosessia ajetaan usein. Historiatietoa on hyvä säilyttää, ja se on todettu paremmaksi vaihtoehdoksi kuin päivittää vanhoja rivejä. Kuitenkin tietomassan kasvua halutaan hillitä, joten sitä varten on tehty transformaatio, joka poistaa valittua aikaväliä vanhemmat rivit. Tämäkin täytyy tehdä vain kerran, ja se voidaan aiemmista projekteista kopioida sellaisenaan, mutta se olisi voitu generoida myös automaattisesti. Toisaalta, mitä enemmän tiedostoja Java-ohjelma generoi, sitä työläämpää on niiden muuttaminen ja ylläpito esimerkiksi ohjelmistopäivitysten yhteydessä.

Heikkoutena tiedostogeneraattorille on toistaiseksi myös se, että se sopii vain Microsoftin SQL Server -tietokantaan. Kun käytössä on esimerkiksi MySQL-kanta, transformaatiot eivät enää toimikaan. Tämä on myös yksi jatkokehitysidea, mutta toistaiseksi ohjelma riittää sellaisenaan. Muutos ei ole kovin suuri, ja tietokannan tyyppi voitaisiin antaa ohjelmalle parametreissa. Ohjelman toiminta on kuitenkin pyritty pitämään yksinkertaisena, sillä turhia ja ylimääräisiä vaatimuksia ei kannata toteuttaa. Vasta, kun ohjelma on ollut käytössä jonkin aikaa ja uusia parannusehdotuksia on tullut, ne kannattaa toteuttaa kerralla.

## 8. PÄÄTELMÄT

Tässä luvussa käsitellään työn loppupäätelmät. Ensimmäinen alaluku keskittyy työn yhteenvedoon. Toisessa arvioidaan tutkimusta ja kolmannessa alaluvussa esitellään suosituksia kohdeyritykselle ja tiedeyhteisölle.

### 8.1. Yhteenveto

Agile on ollut puheenaiheena ohjelmistokehityksessä jo vuosia. Nyt Agile on jalkautumassa liiketoimintatiedon hallintaan. (Henschen 2010.) Ketterällä liiketoimintatiedon hallinnalla on kaksi tarkoitusta: saada kehitys tapahtumaan nopeammin ja reagoida nopeammin liiketoiminnan muuttuviin vaatimuksiin (Wailgum 2010). Suurimmat esteet tietovarastoinnin käyttöönotossa ovat kalliit lisenssimaksut, kallis henkilöstön koulutus ja BI-analyttikkojen kalleus. Lisäksi sen käyttöönottoa hankaloittavat BI-ohjelmistojen tekninen luonne ja vaikeakäyttöisyys. (Henschen 2010.) Kustannusten suuruus johtuu tietovarastoinnin monimutkaisuudesta ja sen vaatimasta työmäärästä. Tämän takia tietovaraston tulee olla nopeammin rakennettavissa sekä raportoinnin ja analysoinnin helppokäyttöisempää. Juuri näihin haasteisiin pyritään ketterällä liiketoimintatiedon hallinnalla vastaamaan. Ketterien menetelmien avulla voidaan paremmin keskittyä hyödylliseen ja tuottavaan työhön, vastata paremmin muuttuviin tarpeisiin sekä parantaa asiakastytyväisyyttä. ETL-rutiinien automatisoinnilla saadaan nopeutettua tietovaraston rakentamista ja laskettua kustannuksia. ETL:n kopioitavuus, tietovaraston standardoiminen ja valmiit raportit tehostavat työtä tuotteistamalla BI:tä. Virtuaalinen tietovarasto mahdollistaisi kokonaan ETL-prosessista luopumisen. Uusia ja ainutlaatuisia ns. ”out of the box”-ratkaisuja esitetään jatkuvasti lisää.

Haastattelujen perusteella IPSS Oy:ssä pohditaan tietovarastoinnin tehostamista ja tuotteistamista. Tietovarastointiprojekteissa on käytetty erilaisia projektimalleja, myös ketteriä menetelmiä on ollut käytössä. IPSS Oy:ssä tulisi kuitenkin yhtenäistää käytäntöjä ja menetelmiä sekä harkita ketterien menetelmien käyttöönottoa jokaiseen projektiin. Niitä voitaisiin soveltaa halutun mukaisesti, jotta saataisiin paras mahdollinen hyöty. Kohdeyrityksessä pidetään tärkeänä eri ETL-työkalujen käytön hallitsemista ja osaamista. Työntekijöiden osaamista eri välineistä tulisi jatkossakin ylläpitää ja kehittää. On tärkeää osata myös valita aina parhaimmat ja sopivimmat työkalut jokaiseen projektiin. IPSS Oy:n tulevaisuudennäkymät näyttävät hyvältä, kun jatkossakin pysytään mukana tekniikan kehityksen kärjessä.

Empiriaosaan kuului käytännön sovellus, jolla automatisoidaan ETL-rutiinia ja tehdään siitä joustavamman muutoksille. Se on pieni osa koko liiketoimintatiedon hallintaympäristön rakentamisessa, mutta tehokas. Se vähentää rutiinyyttä, jolloin voidaan keskittyä hankalampaan ja enemmän asiakaskohtaista räätälöintiä vaativampaan osaan, tietovaraston päivittämiseen. Lisäksi ohjelma on erittäin hyvin kopioitavissa eri projekteihin ja se reagoi muutoksiin välittömästi. Metadatataulujen ansiosta muutokset ovat myös helpommin havaittavissa. Muutoksia itse ETL-rutiineihin on hankalampaa ja työläämpää tehdä, pois lukien parametrisoitavat osat. Tällaiset muutokset ovat kuitenkin verrattain harvinaisia, sillä rutiini on melko yksinkertainen ja sitä on käytetty jo kauan. Lisää työtä saattaa aiheuttaa myös Pentaho Data Integratorin versiovaihdot, mikäli uuteen versioon tai Salesforcen rajapintaan tulee yhteensopimattomuutta aiheuttavia muutoksia. Empiriaosan ohjelma on toteutettu jo kahteen eri PDI:n versioon, joten työmäärä ja tarvittavat toimenpiteet ovat hyvin tiedossa.

Koko tietovarastoinnin näkökulmasta empirian sovellus automatisoi vain yhden, ensimmäisen osan. Toinen osa ETL-prosessista on välivarastotietokannasta tietovarastoon. Tietovaraston rakentaminen vaatii edelleen paljon työtä. Muutokset tulevat vain välivarastotietokantaan, eikä sen pidemmälle. Tiedon läpinäkyvyys paranee joka tapauksessa, kun data on paikallisessa tietokannassa helposti saatavissa. Muutoksia on myös nopeampi viedä loppuun asti, mutta se vaatii ensin herätteen tekemiselle, muutokset tietovaraston päivittämisen ETL-rutiiniin, tietomalliin ja tarvittaessa raportointiin asti.

Lisäksi raportteja voidaan toteuttaa suoraan välivarastotietokannasta. Tätä varten tulisi kantaan lisätä näkymät, joiden avulla päästää käsiksi uusimpaan tietoon. Tietokannasta tulisi samanlainen kuin operatiivisesta kannasta ja se mahdollistaisi reaaliaikaisen raportoinnin. Välivarastotietokanta on tärkeä osa tietovarastoinnissa, sillä sen avulla integroidutaan muihin järjestelmiin. Kanta on risteyskohta tietovarastolle, Salesforcele, asiakkaan omiin tietokantoihin ja järjestelmiin, kuten toiminnanohjausjärjestelmälle ja muille sovelluksille, joilla esimerkiksi muodostetaan kohderyhmiä ja asetetaan tavoitteita. Jotta järjestelmät keskustelisivat tehokkaasti keskenään, ne vaativat mahdollisimman reaaliaikaista dataa.

## **8.2. Tutkimuksen tarkastelu**

Työn tutkimuskysymys oli miten kehittää liiketoimintatiedon hallintaa ketterämmäksi. Tarkastellaan vastausta alakysymysten kautta. Ensimmäinen alatutkimuskysymys liittyi termiin ketterä liiketoimintatiedon hallinta, mitä se tarkoittaa ja miten se määritellään. Teoriaosan viimeisessä luvussa ketterän liiketoimintatiedon hallinnan tarkoitusta pohdittiin eri näkökulmista melko monipuolisesti. Sen tarkoituksena on nopeuttaa kehitystä ja muutoksiin reagoimista. Keinot olivat muun muassa ketterien menetelmien



käyttö, muutoksiin valmistautuminen ja ETL:n automatisointi. Agile BI eli ketterä liiketoimintatiedon hallinta on aiheena uusi, joten suomenkielistä lähdemateriaalia ei löytynyt ollenkaan ja englanninkielisiäkin lähteitä löytyi niukasti. Lähteissäkään asiat eivät olleet täysin yksiselitteisiä. Kaikki tekijät huomioon ottaen ensimmäiseen tutkimuskysymykseen saatiin hyvin vastauksia. Se herätti lisäksi uusia kysymyksiä ja ajatuksia aiheesta.

Toinen alatutkimuskysymys liittyi ketterien menetelmien hyödyntämiseen tietovarastointiprojekteissa. Tätä aihetta käsiteltiin työn teoriaosassa. Ketterien menetelmien käyttö tietovarastointiprojekteissa tuo suuria hyötyjä, mutta siinä on myös heikkouksia. Empiriaosassa aihetta käsiteltiin haastatteluiden kautta, jolloin aiheesta saatiin hieman käytännön kokemuksia. Scrum on ollut käytössä muutamissa projekteissa, mutta sen toimivuudesta tietovarastointiprojekteihin ei olla varmoja. Sen potentiaali ja hyödyt tiedetään, mutta myös haasteita on löytynyt. Ketteriä menetelmiä tulisikin soveltaa tietovarastointiprojekteihin sopivaksi. Rajallisten resurssien takia tässä työssä ei sen tarkemmin koeistettu ketterien menetelmien toimivuutta käytännössä, mutta aihetta tuli käsiteltyä tyydyttävästi.

Kolmanteen tutkimuskysymykseen, tietovaraston rakentamisen nopeuttamiseen keskityttiin myös sekä teoria- että empiriaosassa. Toisen luvun lopussa käsiteltiin tietovarastointia ja sen rakentamista. Kolmannessa luvussa asiaa tarkasteltiin koko projektin näkökulmasta ja viimeisessä teorialuvussa varsinaisesti tietovaraston rakentamisen nopeuttamista. Ketterien menetelmien lisäksi kehitystä voidaan nopeuttaa automatisoinneilla, standardoimalla sekä mahdollisesti myös uusilla ja ainutlaatuisilla ratkaisuilla. Empiriaosan sovellus, ETL-rutiinin automatisointi keskittyi myös vastaamaan tähän alatutkimuskysymykseen. Tämän osalta tavoitteisiin päästiin hyvin.

Neljännessä alatutkimuskysymyksessä kysyttiin, miten liiketoimintatiedon hallintaympäristöä voidaan kehittää ketterämmäksi. Työn teoriaosa jäi toivottua suppeammaksi lähdemateriaalin puutteessa, eikä siihen tuntunut löytyvän kovin konkreettisia ratkaisuja. Haastatteluissa tuli ilmi, että muutokset eivät ole jatkuvia, mutta tiedon kulku estää usein muutosten havaitsemisen. Käytännön sovelluksella pyrittiin tuomaan ketteryyttä välivarastointiin asti sekä metatiedon avulla tuomaan muutoksia läpinäkyvämmäksi. Viimeinen tutkimuskysymys ei ollut aikaisempiin verrattuna kovin tärkeä, mutta sitä käsiteltiin resurssien puitteissa tyydyttävästi.

Työ vastasi hyvin kohdeyrityksen tarpeisiin. Vaikka haastatteluja oli vähän, yrityksessäkin alettiin pohtia tulevaisuutta, mahdollisuuksia ja haasteita sekä muita asioita yleisemmästä näkökulmasta. Haastatteluja olisi voinut olla myös kohdeyrityksen ulkopuolelta, jotta näkemys asiasta olisi laajempi, eikä olisi keskittynyt yhteen yritykseen. Haastattelukysymyksiä olisi pitänyt miettiä enemmän etukäteen selkeämmän kokonaiskuvan saamiseksi, mutta ne olivat kuitenkin kattavia ja

onnistuivat kohtuullisen hyvin. Kehityskohteita, uusia ideoita ja huomiota vaativia asioita löytyi, mutta konkreettisiin parannusehdotuksiin ei päästy. Työn yhtenä tavoitteena oli herättää ajatuksia ja löytää kehityskohteita.

Empiriaosan käytännön sovellus oli tarpeellinen kohdeyritykselle. Se vähensi työtä ja reagoi muutoksiin odotusten mukaisesti. Haasteitakin löytyi matkalla, versiovaihdot aiheuttavat lisätyötä ja ohjelmistovirheitä löytyi myös ETL-työkalusta. Sovelluksesta ei täysin automaattista tullut, vaan se tulee sisällyttää suurempaan kokonaisuuteen ETL-prosessia. Vähintäänkin sille on luotava tarvittavat parametrit, jotta tietoa voidaan ladata. Tämä on pieni työvaihe, mutta se olisi voitu myös generoida automaattisesti. Vaihe jätettiin automatisoimatta, sillä se olisi vaatinut paljon lisätyötä, eikä sitä tarvitse toteuttaa kuin kerran. Ohjelma olisi lisäksi voinut luoda tietokantaan näkymät, joilla päästäisiin käsiksi aina viimeisimpään tietoon. Lisäksi ohjelmakoodista olisi voinut tehdä vieläkin yleiskäyttöisemmän, jotta se olisi yhteensopiva muiden tietokantojen kanssa ja mahdollisesti muiden tietojärjestelmien. Toisaalta automatisointi on hidasta toteuttaa, joten siihen kannattaa sisällyttää vain välttämättömimmät osat. Ohjelma toteutettiin suunnitelmien mukaisesti ja se vastasi siihen kohdistettuihin odotuksiin. Sovellus on otettu käyttöön ja se on toiminut moitteettomasti. Erityisesti metadatan esille tuominen oli tärkeä parannus aiempaan ETL-prosessiin.

### **8.3. Työn reliabiliteetin ja validiteetin arviointi**

Ymmärtämiseen pyrkivässä lähestymistavassa käytetään tavallisesti kvalitatiivista analyysia ja päätelmien tekoa (Hirsijärvi et al. 1997, s. 220). Tämä tutkimus tavoitteena oli kasvattaa ymmärrystä aiheesta ja etsiä parempia keinoja projektien läpiviemiseksi, joten työ on laadullinen. Laadullisissa, kvalitatiivisissa tutkimuksissa ei reliabiliteetin ja validiteetin pohtiminen määrällisen tutkimuksen muodossa ole tarkoituksenmukaista. Niitä voidaan tarkastella silti yleisesti myös laadullisissa tutkimuksissa.

Reliabiliteetti tarkoittaa mittauksen uskottavuutta (Hirsijärvi et al. 1997, s. 249). Toisin sanoen reliabiliteetti kertoo kuinka luotettava mittausta on, eli kuinka hyvin mittausta on toteutettu. Vaikka tämä tutkimus olisi toistettavissa, tulokset on sidottu kuitenkin yhteen yritykseen ja tiettyyn hetkeen. Näin tämä on ainutkertainen tutkimus. Samankaltainen tutkimus voitaisiin tehdä samassa tai toisessa yrityksessä myöhemmin, olisi tulokset sidoksissa siihen yritykseen, sen henkilöihin ja ajankohtaan.

Tämän tutkimuksen reliabiliteettia on pyritty parantamaan valitsemalla haastatteluihin erilaisia työntekijöitä. He ovat kaikki erilaisessa työtehtävässä, omaavat erilaiset taustat ja mielipiteet, vaikka kaikilla on vuosien kokemus liiketoimintatiedon hallinnasta. Haastatteluita suoritettiin vain kolme, joten jokaisen vaikutus tuloksiin on merkittävä. Haastatteluiden pieni määrä johtuu rajallisen ajan lisäksi sopivien haastateltavien vähäisestä määrästä sekä siitä, että työssä on keskitetty enemmän käytännön

sovellukseen. Haastatteluista pyrittiin tekemään mahdollisimman kattavia ja niissä tulikin hyvin paljon asiaa aiheeseen liittyen. Tämäkin osaltaan kasvattaa reliabiliteettia, sillä kiire ei päässyt vaikuttamaan saatuihin vastauksiin.

Käytännön sovelluksena toteutettiin automatisoitu ETL-prosessi. Sen tulosten reliabiliteetin arviointi on melko kyseenalaista. Tulosten analysoinnissa ei juuri ole käytetty kvantitatiivisia menetelmiä. Ainoastaan ohjelman suorittamisen kesto on mitattu kvantitatiivisesti tuomaan jonkinlaista käsitystä sen raskaudesta. Työntekijä on arvioinut subjektiivisesti sovellusta ja verrannut omiin aikaisemmista projekteista saatuihin kokemuksiin. ETL-rutiini on vakiintunut ja ollut paljon käytössä, sitä ei ollut edes tarvetta arvioida. Tulokset ovat kuitenkin selkeitä, ETL-transformaatioiden käsin tekeminen on huomattavasti hitaampaa kuin niiden automaattinen generoiminen. Yleisesti työn reliabiliteetin voitaisiin sanoa olevan kohtuullinen työhön kohdistettujen resurssien määrään suhteutettuna.

Tutkimuksen validius tarkoittaa mittaako tutkimus sitä, mitä sen on tarkoitus mitata (Hirsijärvi et al. 1997, s. 222). Validi mittari siis mittaa mahdollisimman tarkasti mitattavaa ilmiötä. Tässä työssä validius voidaan ajatella tulkinnoilla. Haastattelijalla, haastateltavilla ja lukijoilla on omat tulkintansa asioista. Työn validiteetin parantamiseksi haastattelukysymykset on pyritty tekemään mahdollisimman selkeiksi ja tarvittaessa niitä on selitetty tarkemmin haastattelutilanteessa, jotta haastattelijat ja haastateltavat ymmärtäisivät toisiaan. Jotta myös lukija ymmärtäisi tutkijaa, tuloksia ja haastateltavia, on haastattelukysymykset laitettu liitteeksi (liite 1), haastatteluiden tulokset kerrottu mahdollisimman selkeästi erikseen sekä niistä koostettu yhteenveto. Käytännön sovelluksessa arvioitiin toimivuutta ja sen vähentämää työmäärää. Työn väheneminen on selvää, kun sovellus tekee käsin tehtävää työtä automaattiseksi. Tutkimus mittaa hyvin sitä, jota on tarkoituskin mitata, joten työn validiteetin voidaan sanoa olevan hyvä.

#### **8.4. Suositukset sekä kohdeyritykselle ja tiedeyhteisölle**

Työn lopputuloksena syntyi automatisoitu ja muutoksiin reagoiva ETL-rutiini. Se on kohdeyritykselle ensimmäinen laatuaan. Kohdeyrityksessä tulisi pohtia, mitä muita kohteita kannattaisi automatisoida. Esimerkiksi tiedonlatausta välitietokannasta tietovarastoon voisi olla seuraava vaihe. Yksi potentiaalinen vaihtoehto olisi myös päivittää suoraan tietovarastoa automatisoiduilla ETL-prosesseilla, jolloin automatisoinnin ja ketteryyden hyödyt olisivat suuremmat. Tuotteistamista voisi kehittää myös OLAP-kuutioiden ja raportoinnin automatisoinnilla.

Kohdeyrityksessä tulisi tutkia tietovaraston tietomallin vakiinnuttamisen mahdollisuutta, ainakin asiakkuudenhallintajärjestelmistä ladatun datan osalta. Tietovarastoon ladattavat muut tiedot tulevat tyypillisesti asiakkaiden toiminnanohjaus-

tai laskutuksen järjestelmästä, jolloin ne ovat hyvin asiakaskohtaisia. Tietovaraston rakenteen vakiinnuttamista varten tulisi huolellisesti arvioida asiakkaiden tietotarpeita, jotta tarvittavat raportit ja mittaristot olisi mahdollista rakentaa.

Kohdeyrityksessä tulisi harkita ketterien menetelmien käyttöönottoa tietovarastointiprojekteissa. Niiden tuomat hyödyt on tiedostettu ja Scrum on alkanut tulla tutuksi työntekijöille. Yksi lähestymistapa olisi kokeilla ketteriä menetelmiä, kuten Scrumia yhdessä projektissa, minkä jälkeen se voitaisiin ottaa käyttöön kaikkiin liiketoimintatiedon hallintaprojekteihin, jos se koetaan hyväksi toimintatavaksi. Ketteriä menetelmiä tulisi soveltaa yrityksen tarpeisiin sopivaksi, eikä niitä tarvitse toteuttaa kirjaimellisesti.

Tällä tutkimuksella on suuri uutuusarvo, sillä ketterää liiketoimintatiedon hallintaa on tutkittu toistaiseksi hyvin vähän. Aiheesta on käyty keskustelua vasta 2010-luvulta alkaen, eikä Suomessa vielä ollenkaan. Ketterät menetelmät ovat vakiintuneet vasta ohjelmistokehitykseen, mutta liiketoimintatiedon hallintaan niitä on sovellettu vielä toistaiseksi vähän. Tärkeää olisi tutkia niiden soveltuvuus liiketoimintatiedon hallinnan projekteihin. Niiden hyödyt ja haitat tulisi empiirisellä tutkimuksella todeta sekä miten ketteriä menetelmiä kannattaisi soveltaa tietovarastointiprojekteissa.

Tietovarastojen tietomallit olisi hyvin mielenkiintoinen tutkimuksen kohde. Mikäli tietovarastoja halutaan yhdenmukaistaa, tulisi ymmärtää mahdollisimman hyvin eri tietovarastojen tietomalleja. Haasteena voi olla se, että tietovarastot ovat yrityksille tärkeitä ja sisältävät liikesalaisuuksia. Kattavaa tietoa tietovarastojen rakenteista voi olla vaikea saada luotettavasti. Tietovarastojen vakiinnuttamista varten tulisi ymmärtää yritysten tietotarpeita hyvin. Hyvä jatkotutkimusaihe olisi selvittää, miten yritykset yleensä hyödyntävät tietovarastojaan ja minkälaista raportointia ja mittaristoja heillä on käytössään. Myös muut ratkaisut, kuten virtuaalinen tietovarasto tulisi ottaa tutkimuksen kohteeksi.

Ketterän liiketoimintatiedon hallinnan tavoitteet ovat selvät. Ne nousevat liiketoiminnan tarpeista, markkinoiden kehityksen nopeutumisesta ja tietovarastoinnin heikkouksista. Muutoksiin on reagoitava nopeammin, kehityksen on oltava nopeampaa ja kustannustehokkaampaa. Menetelmät, joilla näihin tavoitteisiin päästään, eivät ole yhtä selkeät. Ketterää liiketoimintatiedon hallintaa tulisi tutkia myös terminä enemmän ja määrittellä se täsmällisemmin. Tulevaisuus näyttää, jääkö ketterä liiketoimintatiedon hallinta hetken ilmiöksi ja aikansa muotitermiksi, vai vakiintuuko se parhaaksi tavaksi toteuttaa liiketoimintatiedon hallinnan projekteja.

## LÄHTEET

- Abrahamsson, P., Salo, O., Ronkainen, J. & Warsta, J. 2002. Agile Software Development Methods: Review and Analysis. Oulu, VTT. 107 pages. [<http://www.vtt.fi/inf/pdf/publications/2002/P478.pdf>]. Luettu 11.12.2010.
- Agile Alliance. 2001. Manifesto for Agile Software Development. [<http://agilemanifesto.org/>]. Luettu 8.12.2010.
- Alavi, M. & Leidner, D. E. 2001. Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MISQ Review, Vol. 25, No. 1, Mar., pp. 107-136.
- Biere, M. 2003. Business Intelligence for the Enterprise. Upper Saddle River, New Jersey, IBM Press. 222 pages.
- Boehm, B. W. 1988. A Spiral Model of Software Development and Enhancement. IEEE Computer Society. Vol. 21, Issue 5, pages 61-72.
- Boehm, B. & Turner, R. 2004. Balancing Agility and Discipline: Evaluating and Integrating Agile and Plan-Driven Methods. Proceedings of the 26th International Conference on Software Engineering. IEEE Computer Society Washington, pp. 718-719.
- Brady, J. 2004. Data Warehousing with Proven Products and Methodology. Tree times, A Publication of Treehouse Software, Inc. Issue 5, December 2004, pages 1-8.
- Casters, M., Bouman, R. & van Dongen, J. 2010. Pentaho Kettle Solutions – Building Open Source ETL Solutions with Pentaho Data Integration. Indianapolis, Indiana, USA, Wiley Publishing, Inc, 674 pages.
- Chaudhuri, S. & Dayal, U. 1997. An overview of data warehousing and OLAP technology. Association for Computing Machinery. Vol. 26, Iss. 1, 10 pages.
- Chaudhuri, S. & Umeshwar, D. 1997. An Overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record. Vol 26, Iss. 1, pp 65-74.
- Chickowski, E. 25.8.2009. Six Steps to Agile BI. Smarter technology. Saatavilla: [<http://www.smartertechnology.com/c/a/Technology-For-Change/Six-Steps-to-Agile-BI/>]. Luettu 22.10.2010.
- Choo, W. C. 2001. Information management for the intelligent organization: the art of scanning the environment. 3rd printing, Medford, New Jersey, Information Today, Inc. 325 pages.
- Davenport, T. H. & Prusak, L. 1998. Working Knowledge: How Organizations Manage What They Know. Harvard Business School, Boston. 201 pages.
- Davenport, T., De Long, D. & Beers, M. 1999. Successful Knowledge Management Projects. Cortada, J., Woods, J. (edit.) The Knowledge Management Yearbook 1999-2000. United States of America, Butterworth-Heinemann. 521 pages.
- Dishman, P. & Calof, J. 2008. Competitive intelligence: a multiphasic precedent to marketing strategy. European Journal of Marketing. Vol 422, No. 7/8, pp. 766-785.
- Forsberg, K., Mooz, H. & Cotterman, H. 2003. Projektihallinta – Malli kaupalliseen ja tekniseen menestykseen. Jyväskylä, Edita Publishing Oy. 350 sivua.
- Halonen, P. & Hannula, M. 2007. Liiketoimintatiedon hallinta suomalaisissa suuryrityksissä. Tampere, Tampereen teknillinen yliopisto ja Tampereen yliopisto, eBRC Research Reports 37. 49 sivua.
- Hammergren, T. 30.8.2010a. Gartner on Agile BI. Balanced Insight Blog. Saatavilla: [<http://www.balancedinsight.com/blog/2010/08/gartner-on-agile-bi/>]. Luettu 16.1.2011.

- Hammergren, T. 6.9.2010b. What Agile BI Really Means. Balanced Insight Blog. Saatavilla: [<http://www.balancedinsight.com/blog/2010/09/what-agile-bi-really-means/>]. Luettu 17.1.2011.
- Helander, N., Kukko, M., Virtanen, P., Väisänen, J. & Yliniemi, T. 2007. Tietämyksenhallinnan käytännöt pirkanmaalaisissa pk-yrityksissä. Tampere, Tampereen teknillinen yliopisto, eBRC Research Report 40. 36 sivua.
- Henschen, D. 6.11.2010. Business Intelligence: How To Get Agile. InformationWeek. Saatavilla: [<http://www.informationweek.com/news/software/bi/showArticle.jhtml?articleID=228200231>]. Luettu 20.1.2011.
- Hirsijärvi, S., Remes, P. & Sajavaara, P. 1997. Tutki ja kirjoita. Tampere, Kirjayhtymä Oy, 432 sivua.
- Hovi, A., Hervonen, H. & Koistinen, H. 2009. Tietovarastot ja Business Intelligence. 1. painos, Porvoo, WS Bookwell. 196 sivua.
- Howson, C. 2007. Successful Business Intelligence: Secrets to Making BI a Killer App. United States of America, The McGraw-Hill Companies. 244 pages.
- Ihanus, M-L. 20.11.2009. Salesforce taipuu mies ja pöytä –yritykseenkin. Tietokonelehti. Saatavilla: [[http://www.tietokone.fi/uutiset/salesforce\\_taipuu\\_mies\\_ja\\_poyta\\_yritykseenkin](http://www.tietokone.fi/uutiset/salesforce_taipuu_mies_ja_poyta_yritykseenkin)]. Luettu 7.1.11.
- Kaario, K. & Peltola, T. 2008. Tiedonhallinta, Avain tietotyön tuottavuuteen. 1. painos. Porvoo, WS Bookwell. 164 sivua.
- Kasanen, E., Lukka, K. & Siitonen A. 1991. Konstruktiivinen tutkimusote liiketaloustieteessä, Liiketaloudellinen Aikakauskirja, No.3, ss. 301–329.
- Kauppi, E. 27.5.2010. Pentahon avoimen lähdekoodin bi-ratkaisut Suomeen. Tietoviikko. Saatavilla: [<http://www.tietoviikko.fi/bi/article423572.ece>]. Luettu 4.1.11.
- Kauppi, E. 21.1.2011. Tämä on CIO:n ykkösprioriteetti vuonna 2011. Tietoviikko. Saatavilla: [<http://www.tietoviikko.fi/cio/article565708.ece>]. Luettu 22.1.2011.
- Kimball, R., Reeves, L., Ross, M. & Thornthwaite, W. 1998. The Data Warehouse Lifecycle Toolkit. Expert Methods for Designing, Developing, and Deploying Data Warehouses. New York, John Wiley & Sons, Inc. 771 p.
- Korhonen, S. 10.1.2011. Oraclen uutuus on tarkoitettu SAP:n asikkaille. Tietoviikko. Saatavilla: [<http://www.tietoviikko.fi/cio/article558531.ece>]. Luettu 22.1.2011.
- Koskinen, A., Pirttimäki, V. & Hannula, M. 2005. Liiketoimintatiedon hallinta suomalaisissa suuryrityksissä vuosina 2002 – 2005. Tampere, Tampereen teknillinen yliopisto, eBRC Research Report 21. 38 sivua.
- Larman, C. & Basili, V. R. 2003. Iterative and Incremental Development: A Brief History. IEEE Computer Society. Vol. 36, Issue 6, pages 47-56.
- Lehtinen, A., Salminen, A. & Nurmeksela, R. 2005. Metatiedot suomalaisen lainsäädäntöprosessin tiedonhallinnassa. Raske2-projektin II väliraportti. Eduskunnan kanslian julkaisu 7/2005.
- Lönnqvist, A., Kujansivu, P. & Antola, J. 2005. Aineettoman pääoman johtaminen. Tampere, JTO-Palvelut Oy. 252 sivua.
- Markkula, M., Huotari, A., Huovinen, S., Karjula, K., Korhonen, R. & Krohn, I. 2001. Avauksia tietämyksen hallintaan, tulevaisuusvaliokunnan teknologian arviointeja 6. Eduskunnan kanslian julkaisu 1/2001, Oy Edita Ab, Helsinki. 190 sivua.

- Miettinen, M. 2002. Tietoturvan historiaa. Seminaarityö. Helsingin yliopisto, tietojenkäsittelytieteen laitos. 12 sivua.
- Moran, D. 12.5.2010. What is Agile BI? Your answers from business user to fluff. Pentaho Blog. Saatavilla: [[http://blog.pentaho.com/2010/05/12/pentaho\\_agilebi\\_contest\\_finalist/](http://blog.pentaho.com/2010/05/12/pentaho_agilebi_contest_finalist/)]. Luettu 8.1.11.
- Moss, L. T. & Atre, S. 2003. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. 3rd printing, Boston, Massachusetts, Addison-Wesley. 543 pages.
- Mundy, J., Thornthwaite, W. & Kimball, R. 2006. The Microsoft Data Warehouse toolkit: with SQL Server 2005 and the Microsoft Business Intelligence toolset. 3. ed. Indianapolis (Ind.): Wiley Publishing. 746 p.
- Nonaka, I. 1991. The Knowledge-Creating Company. Harvard Business Review. Vol. 69, No. 6, pp. 96-104.
- Okkonen, J. 2010. TITA-6600 Tietojohtamisen diplomityöseminaari, luentokalvot. Tampereen Teknillinen Yliopisto. Saatavilla rajoitetusti: [<http://moodle.tut.fi/file.php/1426/moddata/forum/73/41294/30082010.ppt>]. Luettu 8.9.2010.
- Parikka, P. 2009. Tietovarastoon perustuvan ennustejärjestelmän toteuttaminen. Diplomityö. Tampere, Tampereen teknillinen yliopisto, Tiedonhallinta. 115 sivua.
- Pirttimäki, V. & Hannula, M. 2002. Business Intelligence suomalaisissa suuryrityksissä 2002. Tampere, Tampereen teknillinen yliopisto, eBRC Research Report 4. 58 sivua.
- Pirttimäki, V. 2007. Business Intelligence as a Managerial Tool in Large Finnish Companies. Thesis for the degree of Doctor. Publication 646. Tampere, Tampere University of Technology. 129 pages.
- Powell, J. & Howson, C. 9.6.2010. Q and A: Agile BI Basics. TDWI. Saatavilla: [<http://tdwi.org/articles/2010/06/09/agile-bi-basics.aspx>]. Luettu 22.1.2011.
- Pulkkinen, M. 2008. Halpeneva bi buustaa bisnestä. Tietoviikko. Julkaistu 10.3.2008. Saatavilla: [<http://www.tietoviikko.fi/bi/article135977.ece?fail=f>]. Luettu 11.10.2010.
- Rantanen, T. 2007. Organisaatio-orientoitunut tietovaraston kehitysprosessi – tapaustutkimus tietovaraston suunnitteluprosessin hallinnasta tietovarastointiprojektin toimittajaorganisaation näkökulmasta. Pro gradu -tutkielma. Tampereen yliopisto, Tietojenkäsittelytieteiden laitos. 102 sivua.
- Royce, W. W. 1970. Managing the Development of Large Software Development and Enhancement. Proceedings, IEEE WESCON. Pages 1-9.
- Salesforce 2010. Annual Report. 93 sivua. Saatavilla: [<http://www.salesforce.com/assets/pdf/investors/2010AnnualReport.pdf>]. Luettu 7.1.11.
- Salo, O. 2006. Enabling Software Process Improvement in Agile Software Development Teams and Organisations. Doctoral Thesis. Espoo, VTT. 149 sivua.
- Saunders, M., Lewis, P. & Thornhill A. 2009. Research Methods for Business Students. Fifth edition, Harlow, England, Pearson Education Limited, 614 pages.
- Scrum Alliance 2005. A Playbook for Adopting the Scrum Method of Achieving Software Agility. [<http://www.scrumalliance.org/resources/66>]. Luettu 11.12.2010.

- Swoyer, S. 25.8.2010a. Agile Meets Integration at TDWI Conference. TDWI. Saatavilla: [<http://tdwi.org/articles/2010/08/25/agile-and-integration.aspx>]. Luettu 19.1.2010.
- Swoyer, S. 15.9.2010b. The Agile Data Warehouse: Keeping Users Happy. TDWI. Saatavilla: [<http://tdwi.org/articles/2010/09/15/agile-dw-keep-users-happy.aspx>]. Luettu 22.10.11.
- Takeuchi, H. & Nonaka, I. 1986. The New New Product Development Game. Harvard Business Review, Vol. Jan-Feb, pp. 137-146.
- Thierauf, R. J. 2001. Effective Business Intelligence Systems. Westport, Quorum Books. 370 pages.
- Todman, C. 2001. Designing a Data Warehouse: Supporting Customer relationship Management. Upper Saddle River, New Jersey, Prentice Hall. 323 p.
- Trujillo, J. & Luján-Mora, S. 2003. A UML Based Approach for Modeling ETL Procressed in Data Warehouses. Lecture Notes in Computer Science, Vol. 2813/2003, pp. 307-320.
- Virta, J. 21.7.2010. Q and A: Agile BI Architectures. TDWI. Saatavilla: [<http://tdwi.org/articles/2010/07/21/agile-bi-architectures.aspx>]. Luettu 22.10.10.
- Vuori, V. & Hannula, M. 2009. Liiketoimintatiedon hallinta suomalaisissa suuryrityksissä vuonna 2009. Työpäperi. Tampereen teknillinen yliopisto. 28 sivua.
- Wailgum, T. 24.2.2009. Businesses Are Still Crazy for BI After All These Years. CIO Magazine. Saatavilla: [[http://advice.cio.com/thomas\\_wailgum/businesses\\_are\\_still\\_crazy\\_for\\_bi\\_after\\_all\\_these\\_years](http://advice.cio.com/thomas_wailgum/businesses_are_still_crazy_for_bi_after_all_these_years)]. Luettu 22.1.11.
- Wailgum, T. 26.4.2010. Agile BI Development: Meet User Demand and Beat IT Complexity. CIO Magazine. Saatavilla: [[http://www.cio.com/article/591966/Agile\\_BI\\_Development\\_Meet\\_User\\_Demand\\_and\\_Beat\\_IT\\_Complexity?page=2&taxonomyId=3000](http://www.cio.com/article/591966/Agile_BI_Development_Meet_User_Demand_and_Beat_IT_Complexity?page=2&taxonomyId=3000)]. Luettu 9.1.11.
- Williams, A. 13.12.2010. Top 10 Cloud Computing Services for 2010. ReadWriteWeb. Saatavilla: [<http://www.readwriteweb.com/cloud/2010/12/top-10-cloud-computing-services-for-2010.php>]. Luettu 7.1.11.
- Winter, R. 2001. The Current and Future Role of Data Warehousing in Corporate Application Architecture. Proceedings of the 34th Hawaii International Conference on System Sciences – 2001, IEEE. 8 pages.
- Worthington, D. 31.3.2010. Pentaho consolidates tools to bridge developers, business. Software Development Times. Saatavilla: [<http://www.sdtimes.com/link/34237>]. Luettu 4.1.11.



## HAASTATTELURUNKO

### 0. Taustatiedot

- Nauhoittamiseen suostuminen
- Kerro itsestäsi
  - o koulutus, toimenkuva
- Työkokemus alalta, työkokemus yrityksessä, asema ja vastuualueet yrityksessä
- Minkälaisissa projekteissa olet ollut mukana

### 1. BI-projekti

- Käytetäänkö tiettyä projektimallia, minkälaista?
  - o Vesiputousmallia, syklistä mallia, ketteriä menetelmiä?
- Sovelletaanko ketteriä menetelmiä?
  - o miten/mitä?
  - o kuinka kauan on käytetty?
  - o oletko tyytyväinen ketteriin menetelmiin?
- Kuinka pitkiä BI-projektit tyypillisesti ovat?
  - o Mikä projekteissa kestää ajallisesti kauiten?
  - o Mikä projekteissa syö eniten resursseja?
- Kuinka monta henkilöä BI-projektiin tyypillisesti osallistuu?
  - o Minkälaisia rooleja henkilöillä on?
  - o Minkälaisissa rooleissa olet itse toiminut?
- Miten tietoa jaetaan? Kuuluuko projekteihin paljon tapaamisia?
- Miten projektin etenemistä seurataan?

### 2. BI-ympäristö

- Minkälaisia työkaluja käytetään?
  - o Mitä näistä olet itse käyttänyt? Oletko käyttänyt jotain muuta?
  - o Mitä käytät mieluiten ja miksi?
- Minkälaisia hyväksi havaittuja käytäntöjä tietovaraston rakentamisessa on?
  - o esimerkiksi nimeäminen
- Kuinka paljon aiempia ETL-rutiineja voidaan hyödyntää?
- Kuinka joustava BI-ympäristö on muutoksille?
  - o Tuleeko muutoksia paljon?
  - o Minkälaisia ovat muutokset?
  - o Kuinka nopeasti niihin tulee reagoida?
- Minkälaisia tiedon mallinnusmenetelmiä on käytetty?
- Minkälaista on ylläpito?
  - o Kuinka paljon ylläpitoon kuluu aikaa, esim. kuukaudessa?

### 3. Dokumentointi

- Minkälaisia dokumentteja BI-ympäristön rakentamisesta tyypillisesti syntyy?
  - o Missä vaiheessa projektia?
  - o Kuka dokumentit luo?
  - o Minkälaisia asiakkailta ja BI-ympäristön tekijöiltä?

- Onko dokumentointi mielestäsi riittävää?
- Hyödynnetäänkö dokumentteja tarpeeksi?

#### **4. Lopuksi**

- Mitä mielestäsi tarkoittaa agile BI?
  - Mitkä ovat vahvuuksia IPSS:n BI-projekteissa?
  - Mitä toiminnassa voisi kehittää? Miten?
  - Miten näet tulevaisuuden? Minkälaiseksi toiminta tulee mahdollisesti kehittymään?
  - Mitä muuta haluaisit kertoa? Tuleeko mieleesi vielä jotain?
- 
- Mielenpide haastattelusta
  - Kuka voisi olla seuraava haastateltava?