

## Tietovarastoinnin uudistamisprosessi pilvipalveluun

Jukka Malinen



<b>Tekijä</b> Jukka Malinen	
<b>Suuntautuminen</b> Liiketoiminnan teknologiat	
<b>Opinnäytetyön nimi</b> Tietovarastoinnin uudistamisprosessi pilvipalveluun	<b>Sivumäärä + liitesivumäärä</b> 44 + 4
<p>Kehittämistehtävän tavoitteena oli selvittää, mistä osa-alueista moderni tietovarastointialusta muodostuu ja kuinka se toteutetaan pilvipalveluun. Pilvipalvelu mahdollistaa nykyaikaisen ja kustannustehokkaan tietovarastointialustan valtuin toiminnallisuuksin. Jatkuvasti lisääntyvä tietomäärä mahdollistaa kilpailuetua yrityksille, jotka kykenevät hyödyntämään tarkoituksenmukaista eri lähteistä kerättyä dataa liiketoimintapäätöksissään oikea-aikaisesti. Jalostetun datan hyödyntämistä voidaan tehdä esimerkiksi analytiikan ja raportoinnin visualisoinnin keinoin Business Intelligence -työkaluja hyväksikäyttäen.</p> <p>Kehittämistehtävä toteutettiin toiminta- ja konstruktiivisia tutkimusmenetelmiä käyttäen. Työssä tunnistettiin ja toteutettiin uudistetun tietovarastoalustan keskeiset osa-alueet pilvipalveluun, joista saatava informaatio tukee yrityksen liiketoimintastrategiaa sekä liiketoimintaympäristössä tapahtuvien muutoksien ymmärtämistä. Työssä on hyödynnetty tekijän työhistorian kautta hankittua käytännön kokemusta tietovarastointiin ja raportointiin liittyvistä kehitystehtävistä finanssitoimialalta.</p> <p>Kehittämistehtävän teoriaosuus perustuu tietovarastoalustan suunnitteluvaiheessa tarvittavaan tietoperustaan sekä sen keskeisiin käsitteisiin, jotka muodostavat pilvipalvelussa toimivan tietovarastokokonaisuuden.</p> <p>Käytännön toteutustyöhön liittyvät empiiriset tiedot kerättiin varsinaisen kehittämissprojektiin liittyvien vaatimusmäärittely projektikokousten, teemahaastatteluiden ja kyselylomakkeista saataviin tietoihin perustuen.</p> <p>Tämän kehittämistehtävän tuloksena on toteutettu tietovarastointiratkaisu pilvipalveluun. Tietovarastoinnin uudistamisen hyötyinä yrityksen henkilöstö kykenee hyödyntämään ajantasaisista tiedoista etenkin liiketoimintapäätöksien tueksi. Tietovarastoinnin uudistaminen lisää ymmärrystä lisäksi liiketoiminnan tuote- ja palvelukehitykseen ympäristön muutostrendeistä.</p> <p>Kehittämistehtävän merkitys toimeksiantajalle on, että yrityksen kilpailukyky, kannattava kasvu ja asiakasymmärryksen muutoksien tunnistaminen paranee, koska yritys kykenee keräämään ulkoisista ja sisäisistä järjestelmistä tarvitsemansa tiedon sekä hyödyntämään keräämäänsä tietoa eli dataa lähes reaaliaikaisesti päätöksenteossa ja liiketoiminnan analytiikassa. Tietovarastoinnin uudistaminen pilvipalveluun tukee myös digitaalisten pankkipalveluiden kehittämistä, jossa eri lähteistä kerättyyn tietoon ja niiden yhdistelyyn perustuen, kasvatetaan eri liiketoiminta-alueiden automaatiota ja tekoälyn käyttömahdollisuuksia.</p>	
<b>Asiasanat</b> data, tietovarasto, pilvipalvelu, analytiikka, raportointi, business intelligence	

## Sisällys

1	Johdanto .....	1
2	Tietovaraston uudistaminen osana yrityksen kilpailukykyä .....	3
2.1	Rajaukset .....	3
2.2	Keskeiset käsitteet .....	4
3	Tietovarastointiuudistuksen teoriapohja .....	6
3.1	Informaatio ja data .....	6
3.2	Master data .....	7
3.3	Tietoarkkitehtuuri ja tietomallin suunnittelu .....	7
3.4	Relaatiotietokanta .....	8
3.5	Tietoallas ja tietovarasto .....	10
3.6	Pilvipalvelu .....	11
3.7	Tiedon visualisointi, business intelligence -työkalut ja analytiikka.....	13
4	Tietovarastointiprosessin kehitysprojekti ja tutkimusmenetelmät.....	14
4.1	Kehittämisen- ja tutkimusstrategia.....	15
4.2	Tietovarastointiuudistuksen kehitysmalli .....	16
4.3	Tiedon keruumenetelmät .....	18
4.4	Haastattelujen toteutus ja tietojen kerääminen kyselylomakkeella.....	19
4.5	Kehittämistehtävän tutkimuksen toteutus.....	20
4.5.1	Kyselylomakkeen avoimet sisältökysymykset .....	21
4.6	Tiedon analysointimenetelmät, teemahaastattelujen vastaukset ja analyysi .....	22
4.7	Nykytilan tunnistaminen .....	23
4.8	Pilvipalvelu ja SNOW-Flake tietovarastonhallintaratkaisu .....	25
4.9	Yhdenmukaisen käsitteistön ja tarvemäärityksen tärkeys tietovarastointiuudistuksessa .....	27
4.10	Tietovaraston mallintamisen arkkitehtuurinen prosessi .....	28
4.11	Raakadatan lataaminen .....	29
4.12	EDW-kerroksen taulumääritykset.....	30
4.13	Datamart-kerros raportoinnin ja analytiikan tukeen .....	32
5	Tulokset .....	36
5.1	Havainnot ja tulokset .....	36
5.2	Tulosten analysointi .....	37
5.3	Tutkimustulosten luotettavuus .....	37
5.4	Kehittämishankkeen eettisyys ja tietosuoja.....	38
6	Johtopäätökset ja jatkokehitysmahdollisuudet .....	40

6.1 Johtopäätökset.....	40
6.2 Jatkokehitysmahdollisuudet.....	40
Lähteet .....	43
Liitteet .....	45

# 1 Johdanto

Tieto mahdollistaa objektiivisen arvion yrityksen historiallisesta kehityksestä tai ennusteen odotetusta kehityssuunnasta. Datasta saatavan informaatioarvon kautta voi kerättyyn tietoon perustuen esittää yritykselle sekä valituille liiketoimintayksiköille kehittämiskohteita esimerkiksi kvartaali- tai vuositavoitteina. Hovi, Hervonen ja Koistisen mukaan (2009, 11) tiedot ovat yrityksen tärkeä resurssi ja pääoma, joka pitää olla hyvin kuvattuna ja saatavilla palvelemaan yrityksen lisääntyviä liiketoiminnan tarpeita.

Tarkoituksenmukainen tiedon sisältö on usein yhdistelmä sisäisten ja ulkoisten tietovarastojen yhdistelmästä, jossa on tärkeää kyetä esittämään johtopäätökset halutuilta ajanjaksoilta. Usein erityisen kiinnostavaksi koetaan tulevaisuuden ennusteet historialliseen tietoon perustuen, missä johtopäätöksiin on yhdistetty myös ulkoisia tietolähteitä. Yrityksien liiketoiminnan kehittämis- ja raportointitehtävistä saatu käytännön kokemus ovat osoittaneet, kuinka keskiössä on virheettömän ja tarkoituksenmukaisen tiedon hyödyntäminen päätöksentekoprosesseissa.

Tämän kehittämistehtävän tutkimuksen kohteena on yrityksen tietovarastointiratkaisun modernisointi pilvipalveluun. Kehittämistehtävässä selvitettiin keskeiset osa-alueet, joiden määrittämisen kautta moderni tietovarastointipalvelu voidaan toteuttaa julkiseen pilvipalveluun. Uudistettuun tietovarastointialustaan tuodaan sisäisistä ja ulkoisista lähdejärjestelmistä kaikki se oleellinen tieto, mikä tukee yrityksen liiketoimintaa ja strategisia tavoitteita. Kehittämistehtävän aikana yritykselle on luotu ja sovittu yhtenäiset käsitteet, mitkä helpottavat eri liiketoimintayksiköiden kehittämistoimintoja sekä ymmärrystä eri yksiköiden toiminnan tavoitteista. Kehittämistehtävän toteuttamisen kautta yrityksen kilpailukyky, kannattava kasvu ja asiakasymmärryksen muutoksien tunnistaminen paranee, koska yritys kykenee hyödyntämään keräämäänsä tietoa eli dataa kokonaisvaltaisesti sekä oikea-aikaisesti päätöksenteon tueksi.

Tämän kehittämistehtävän tuloksena syntyy tietovarastointi ratkaisu pilvipalveluun. Tietovarastoinnin uudistamisen hyötyinä yrityksen henkilöstö kykenee hyödyntämään ajantasaista tietoa liiketoimintansa tueksi, lisäksi analysointia tukevat business intelligence käyttöliittymät sekä tiedon visualisointia hyödyntävät raportit helpottavat liiketoiminnan kehittämisen seuranta suhteessa asetettuihin tavoitteisiin. Henkilöstö kykenee myös itse tuottamaan ja automatisoimaan tarvitsemiaan räätälöityjä raportteja. Analytiikka tukee ymmärryksen lisäämistä liiketoimintaan vaikuttavien muutosten ja trendien havaitsemiseksi sekä helpottaa erilaisten kampanjoiden suunnittelua valituille kohderyhmille.

Kehittämistehtävässä on hyödynnetty tekijän työhistorian kautta hankittua käytännön kokemusta tietovarastointiin ja raportointiin liittyvistä kehitystehtävistä finanssitoimialalta. Tekijä on määritellyt ulkoisilta kumppaneilta tilattavia tietovarastointiin liittyviä tietokokonaisuuksia sekä voinut kehittää yrityksen sisäisiä raportointikyvykkyyksiä ja analytiikkaa tukevia ratkaisuja liiketoiminnan hyödyksi. Kehittämistehtävässä hyödynnettiin lisäksi yrityksen sisäisiä asiantuntijoita ja kumppanuusyhteistyöhön liittyviä sidosryhmäverkostoja. Kehittämistehtävässä selvitettiin myös raportoinnin ja analytiikan merkitystä yrityksen kilpailukyvyyn edistämiseksi. Kehittämistehtävä on toteutettu monivaiheisena järjestelmäprojektina pääasiallisesti vuosien 2020 ja 2021 aikana, ja sen tavoitteena oli tukea yrityksen kannattavan kasvun strategiaa, missä yleisesti keskiössä on varmistaa onnistunut asiakaskokemus sekä kasvattaa ymmärrystä asiakkaan elinkaaren eri vaiheista.

Ulkoisista lähteistä saatavilla tiedon määrä on käytännössä rajatonta, tarkoituksenmukaisen tiedon käyttö ja oikea aikainen hyödyntäminen on keskeistä. Vären mukaan (2019, 5) kaikessa erilaisessa liiketoiminnassa käytetään dataa tavalla tai toisella, sekä erilaiset teknologiset innovaatiot ja analytiikka sekä tekoälyn tuleminen antavat datan hyödyntämiselle lisää uusia mahdollisuuksia. Mikäli tietolähteitä eli tietovarastoja on ulkoistettu liiaksi, voi nämä järjestelmäratkaisut aiheuttavat yrityksen kilpailukyvyille, liiketoiminnan kehittämiseen sekä esimerkiksi operatiiviselle tehokkuudelle merkittäviä puutteita.

## 2 Tietovaraston uudistaminen osana yrityksen kilpailukykyä

Kehittämistehtävän tavoitteena on lisätä uudistetun tietovaraston kautta yrityksen kilpailukykyä, helpottaa liiketoimintapäätöksiä tekemistä sekä ymmärtää asiakaskäyttäytymisen muutostrendejä datasta saatavaa informaatioarvoa hyödyntäen. Ennen kehittämistehtävää yrityksen tietovarastoinnin ratkaisut olivat keskeisiltä osin ulkoistettu. Kehittämistehtävässä tunnistetaan nykyisen tietovarastointiratkaisun puutteellisuudet ja ne keskeiset osa-alueet, mitä uusi moderni tietovarastointialusta pilvessä edellyttää.

Tavoitetilassa uusi tietovarastointialusta kykenee keräämään lukuisista eri ulkoisista ja sisäisistä lähdejärjestelmistä sen oleellisen tiedon, mitä yrityksen liiketoiminta todella tarvitsee. Tietovarastointialustan rakentaminen pilveen mahdollistaa kustannustehokkaan sekä vähän hallinnollista työtä vaativan ratkaisun, joka toimiessaan tukee yrityksen kannattavaa kasvua ja määriteltyä strategiaa. Työn yhtenä tavoitteena on, että tietovarastoinnin lopputuotokset tukevat yrityksen raportointia sekä analytiikkaa kerättyyn dataan perustuen. Lisäksi raporttien saatavuutta ja räätälöinnin mahdollisuutta on tarkoitus laajentaa läpi yrityksen eri organisaatiotasolle.

Kehittämistehtävän kohteena ja tutkimuskysymyksenä on selvittää sekä toteuttaa nykyainen tietovarastointialusta pilvipalveluun, jotta se tukee yrityksen liiketoiminnan määrittelmiä tarpeita sekä tukee sen kilpailukykyä. Kehittämistehtävässä tavoitetilan mahdollistamiseksi selvitettiin vastaukset seuraaviin kysymyksiin:

1. Mikä on tarkoituksenmukainen tietovarastoinnin kehittämisen prosessi?
2. Mitkä ovat ne osa-alueet, jotka ovat huomioitava, kun tietovarastointialustaa uudistetaan pilvipalveluun?
3. Mikä on se tehokas toimintamalli, jonka avulla uudistetusta tietovarastointiratkaisusta voidaan hyödyntää dataa liiketoiminnan tueksi eri organisaatioiden tasoille?

Uudistetun tietovarastointiratkaisun tekniset kuvaukset on toteutettu tässä materiaalissa siten, että tietosuojan ja salassapitovelvollisuuksiin liittyvät vaatimukset täyttyvät.

### 2.1 Rajaukset

Kehittämistehtävässä kuvataan ja tuetaan finanssitoimialalla toimivan yrityksen tietovarastointialustan kehitysprosessia. Kehittämistehtävä on kuvattu sellaisin yleisesti hyödynnettävin käsittein ja menetelmin, että yrityksen luottamukselliset tiedot ovat suojattu käytettä-

vien esimerkkien ja kerätyn empiirisen lähdeaineistojen osalta. Yrityksen järjestelmäarkkitehtuurin muutoksia kuvataan vain uudistuksen muutosvaikutuksien näkökulmasta yleisellä tasolla. Myös palvelujen ulkoistamisiin liittyvät yksityiskohtaiset selvityskartoitukset ovat rajattu kokonaan pois kehittämistehtävän sisällön kuvauksesta. Tietovarastoinnin kehitysprojektin eri vaiheet on kuvattu ylätasolla siten, että keskeisten vaiheiden kokonaiskuvan ymmärtäminen on mahdollista. Seuraavassa on esitelty ne ydinkäsitteet, jotka lukijan on hyvä ymmärtää kehittämistehtävän sisällön ymmärtämiseksi.

## **2.2 Keskeiset käsitteet**

Yrityksen keräävät ja tallentavat tietoa eli dataa liiketoimintansa edistämiseksi. Hernandezin (2000, 35) mukaan informaatio on dataa, joka on käsitelty tavalla, joka tekee siitä merkityksellistä ja siis hyödyllistä sitä käyttävälle tai katselevalle henkilölle sekä lisää että, data on arvoja, mitkä ovat talletettu tietokantaan.

Master dataksi Väre määrittelee (2019, 23) yrityksen sellaisen tiedon, mikä kertoo yrityksen myytävistä tai tarjottavista tuotteista, missä liiketoimintaa tehdään ja ketkä ovat asiakkaita. Master data on yrityksen toiminnan näkökulmasta tärkeintä tietoa, toiminnan jatkuvuuden näkökulmasta. Salon mukaan (2014, 9) Big Data viittaa käsitteenä kahteen asiaan, että datan määrä on lisääntynyt ja monipuolistunut sekä että, ratkaisuvaihtoehtoihin, joilla suurten datamassojen tallennusta, liikuttelua sekä sen hyödyntämistä voidaan lähestyä.

Tietoarkkitehtuuri on käsitteenä laaja ja vaihteleva, tietoarkkitehtuuriin voi Myllymäen ym. (2011, 34) mukaan kuuluvan esimerkiksi tietojen mallinnustavat, datoihin liittyvän määrittelyn sekä DW- ja BI-arkkitehtuurit sekä tietoihin liittyvän infrastruktuurin ja tiedon tallennuksen. Tietomallin Hovi ym. tiivistää (2005, 6) kuvausmenetelmäksi ja säännöiksi, jotka muodostavan tietomallin rakenteen sekä mallin väliset yhteydet.

Hovi, Hervonen ja Koistisen mukaan (2009, 192) tietovarasto on yhdenmukainen ja kokonaisuus yrityksen tietoja, jotka tukevat päätöksentekoa. Tietovarasto tukee raportointia ja analysointia eli business intelligence -käyttöä.

Tietoallas (Ari Hovi 2020) on arkkitehtuurisesti tallennusalue, johon on nopeaa tallettaa isoja tietomassoja sekä myös strukturoimatonta dataa. Edullisen talletuksen vuoksi erilaista dataa voidaan kerätä laajasti ja vasta myöhemmin päättää mitä osaa tiedosta käytetään varsinaisessa tietovarastossa.



Salon mukaan (2012, 16) käsitettä pilvi eli englanniksi ”cloud” käytetään kielikuvana, jolla viitataan internetiin, kun taas pilvipalveluilla tarkoitetaan tietotekniikkaresursseja kuten laskenta- ja tallennuskapasiteettia, sovelluksia ja palveluita, joita tarjotaan verkon välityksellä käyttöön, eikä käyttäjän tarvitse huolehtia pilvipalveluiden ylläpidosta.

### 3 Tietovarastointiuudistuksen teoriapohja

Kehittämistehtävän teoriaosuus sisältää ne keskeiset tietoperustan osa-alueet, jotka liittyvät käsitteinä ja menetelminä yrityksen tietovarastointiuudistuksen toteuttamiseksi pilvipalveluun. Tietoperusta tukee niitä valittuja osa-alueita, jotka yrityksen pitää ratkaista sekä tehdä päätöksiä rakentaessaan modernin tietovaraston. Käytännön tasolla tämä tarkoittaa järjestelmätekniisten kyvykkyyksien pystyttämistä pilviympäristöön sekä tietokantahallintaohjelmiston valintaa hyödyntäen kumppanuusyhteistyötä.

Lisäksi liiketoiminnalle merkityksellisen datan keräämistä ja tallentamista valitun tietomallin mukaisesti toteutetaan tietovarastointialustalla pilvipalveluun, siksi pitää tallennettava data ja käsitteistö määritellä yrityksessä ymmärrettävään sekä liiketoiminnan näkökulmasta merkitykselliseksi sisällöksi. Tallennetun tiedon eli datan konkreettinen hyödyntäminen tapahtuu esimerkiksi BI-työkaluja ja automaatio raportointia hyödyntäen, historiallinen tallennettu data ja siitä tehtävät ennusteet tukevat analytiikan hyödyntämistä liiketoiminnassa. Siten edellä mainittujen ja liiketoiminnalle merkityksellisen modernin pilvipalvelussa toimivan tietovaraston toteuttamisen perusteeksi on luvuissa 3.1 – 3.7 kuvattu tarpeellinen tietoperusta tämän kehittämistehtävän toteuttamiseksi.

Kehittämistehtävän teoriaosuudesta on rajattu pois uudistetun tietovaraston kytkeytyminen yrityksen kokonaistietoarkkitehtuuriin sekä kumppanuusyhteistyöhön liittyvän teorian tarkastelu järjestelmäprojektissa. Teoriaosuudessa ei käsitellä myöskään järjestelmäprojektinhallintaan liittyviä kysymyksiä. Edelliset rajaukset on tehty siksi, että yritys käyttää digitaalisissa palvelu ja alusta-arkkitehtuuriratkaisuissa lukuisia muita pilvipalveluiden toimittajia, joiden vastuualueiden prosesseja, eikä niiden vaikutuksia järjestelmäprojektin hallintaan olisi tarkoituksenmukaista avata luottamuksellisten ja yrityksen kilpailukykyyn vaikuttavien tekijöiden näkökulmasta.

#### 3.1 Informaatio ja data

Hovi, Huotari ja Lahdenmakin mukaan (2005, 4) tiedot ovat yritykselle tärkeä resurssi, yrityksen johto tekee tärkeitä päätöksiä toiminnasta tallennettujen tietojen perusteella, sekä lisäävät, että tiedon tallentaminen on pitänyt tapahtua järkevässä muodossa, jotta tietoja voidaan hyödyntää eri yhdistelmillä yrityksen tarpeisiin.

Väre lajittelee (2019, 15–16) erilaiset datatyypit analyttiseen dataan, mitä on rikastettu yrityksen ulkopuolisella datalla analytiikan edistämiseksi. Master dataan, mikä kuvaa yrityksen avain-, ydin- tai perustietoa. Metadataan, mikä kuvaa tiedon käsittemalliin liittyvää

määrittelydataa sekä tiedon sisältöä. Esimerkiksi tekstimuotoinen, numeerinen tai päivämäärämuotoinen tieto. Referenssidata ryhmittelee ja luokittelee muuta dataa esimerkiksi ennalta määrätyn standardin mukaisesti, esimerkiksi ISO 3166 maakoodit tai postinumerot. Transaktiodata kuvaa varsinaisesta liiketoiminnasta syntyvää tapahtumatietoa ja Big data tietoa, jota syntyy suuria määriä ja voi olla luonteeltaan myös kuva- tai videomuotoista.

### **3.2 Master data**

Master dataa ovat esimerkiksi yrityksen asiakas, tuote- ja organisaatorakenteeseen liittyviä tietoja. Väre tiivistää (2019, 16), että master data kertoo yrityksen keskeisen toiminnan tarkoituksen tuotteiden ja palveluiden merkityksen kautta, palveluita käyttävän yrityksen asiakkaiden määrityksen ja toimintaan liittyvät sidosryhmät (esim. yrityksen maantieteelliset toimipisteet, organisaation työntekijät, yrityksen sidosryhmäverkostot).

Master dataa ovat Myllymäen, Hinkan, Hirvensalon ja Hämäläisen (2011, 34) mukaan tiedot, jotka ovat yhteisiä kaikille organisaation eri tasoille. Master data on usein pitkäikäistä ja hitaasti muuttuvaa, vastakohtana master datalle on tapahtumatiedot, jota päivitetään usein. Master data kannattaa keskittää yrityksessä yhteen paikkaan, jotta eroavaisuuksia ei synny organisaation sisällä.

### **3.3 Tietoarkkitehtuuri ja tietomallin suunnittelu**

Tietomallin ja tietovaraston suunnittelu on tärkeä prosessi, jotta tietovaraston sisältö ja rakenne tukee liiketoiminnan todellisia tarpeita. Tietomallin suunnittelu liiketoiminta-asiantuntijoiden kanssa on tärkeää, jotta liiketoiminnan tarpeet ja tavoitteet ymmärretään. Samalla muodostetaan käsitteitä ja ne määritellään sovitun käsittemallin mukaisesti yhtenäiseksi, jotta yrityksessä ymmärretään yhtenäisten periaatteiden mukaisesti luotu tietomalli.

Tietomallin Hovi ym. mukaan (2005, 6) tietomalli kuvataan määrätyillä säännöillä, jotka kuvaavat syntyvät tietorakenteet sekä tietojen väliset yhteydet toisiinsa.

<b>Tietomalleissa käytettyjen termien vertailua</b>			
<b>Käsitelmä</b>	<b>Oliomalli</b>	<b>Relaatiomalli</b>	<b>Relaatiotietokanta</b>
Käsite	Luokka	Relaatiomalli	Taulu
Tieto (attribuutti)	Attribuutti	Attribuutti	Sarake
Esiintymä	Esiintymä	Monikko (tuple)	Rivi
Yhteys	Suhde	Viiteavain	Viiteavain
Perusavain	Object ID	Perusavain	Perusavain

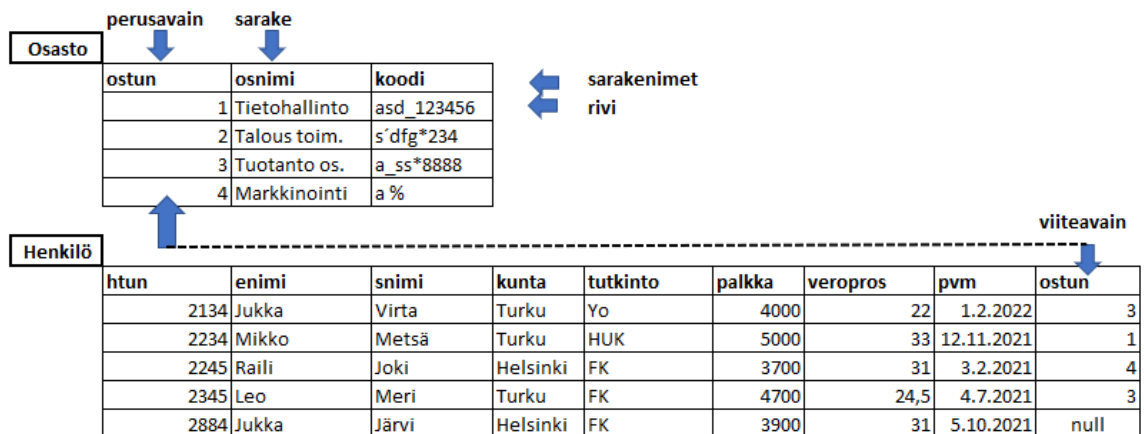
Taulukko 1. Tietomalleissa käytettyjen termien vertailua (mukaillen Hovi ym. 2005, 7).

Tietokannan suunnittelu on oleellisen tärkeää. Huonosti suunniteltu tietokanta on kuin talo huonoilla perustuksilla: aina joutuu paikkaamaan. Tietokantojen kohdalla paikkaaminen tarkoittaa tarpeettoman monimutkaista SQL-ohjelmointia. Tietokannan suunnittelussa on seuraavia vaiheita:

- käsiteanalyysi: tehdään karkeat piirustukset tietokannasta
- tarveanalyysi: täydennetään ja testataan tehtyä käsitelmä
- normalisointi: tarkistetaan, että malli on ns. kolmannessa normaalimuodossa (ei toisteisuutta)
- taulujen muodostaminen: muunnetaan käsitelmä relaatiokannan taulurakenteiksi
- suoristuskäyvyn tarkistus: lisätään tarpeelliset indeksit ja mahdollisesti muutenkin viritetään (Hovi 2010, 10).

### **3.4 Relatiotietokanta**

Relatiomallissa voidaan tunnistaa 3 keskeistä osaa, jotka ovat rakenne, käsittely- ja eheyssäännöt. Hovin (2010, 10) mukaan relaatiotietokannassa asiat kuten osastot ja henkilöt esitetään tauluina (table), tauluissa on sarakkeita (column) ja rivejä (rows). Sarakkeille määritetään tietoa koskeva tyyppi (esim. numeerinen tai merkkimuotoinen sekä tiedon pituus). Relatiotietokannoissa tiedot talletetaan aina tauluihin.



Taulukko 2. Relaatiotietokannan tauluja (mukaillen Hovi 2010, 6)

Vären mukaan (2019, 15) relaatiossa on vähintään kahden entiteetin välinen suhde (asia tai käsite, jota data kuvaa), usein entiteetillä on useita eri suhteita toisiinsa. Relatio siis kuvaa minkälainen käsiteriippuvuus näillä entiteeteillä on.

Relaatiotietokannat perustuvat IBM:n tutkija E. F. Coddin v.1970 julkaisemaan relaatiomalliin, joka määrittelee relaatiotietokannan teoreettisen pohjan. Relatiomalli perustuu joukko-oppiin, matematiikkaan ja predikaattilogiikkaan. Relatiomalli ei ota kantaa relaatiokannan fyysiseen toteutukseen, vaan se on jätetty tietokantatuotteiden toimittajille. (Hovi ym. 2005, 7).

Jokaisessa relaatiotietokannan tauluissa tarvitaan tunnisteita, jotta tietoja on mahdollista hyödyntää eri tauluista halutuun kyselyyn. Hovin mukaan (2010, 6) tauluissa tarvitaan perusavain tunnisteena (primary key, PK), mitkä löytyvät osasto taulun sarakkeesta "ostun" sekä henkilö taulun "htun" sarakkeista. Perusavaimen pitää olla yksilöivä tunniste, eikä samaa arvoa voi esiintyä perusavain sarakkeen useammalla rivillä, perusavain voi kuitenkin koostua useamman sarakkeen tiedoista.

Henkilöt liittyvät yrityksiin. Osastolla voi olla monta henkilöä, mutta yksi henkilö kuuluu aina yhteen osastoon. Tätä kutsutaan isä-lapsi-yhteydeksi tai yksi moneen yhteydeksi. Isällä voi olla monta lasta, mutta lapsella on yksi isä. Asian hoitamiseksi on lapsitaulussa eli henkilötaulussa linkkikenttä eli viiteavain (foreign key) ostun, joka viittaa osastotaulun perusavaimen ostun. Viittaavaa taulua sanotaan lapsitauluksi ja viittauksen kohteeseen olevaa taulua isätauluksi. Viiteavaimia tarvitaan, kun tauluja halutaan yhdistää eli kun tehdään liitoksia. (Hovi 2010, 6).

Relaatiotietomallin käsittelyyn liittyy joukko-opillisuus. Hovin ym. (2005, 10–11) mukaan taulu muodostuu riveistä, joihin voi kohdistaa voimakkaita joukko-operaatioita, kuten että, hae kaikki espoolaiset asiakkaat (valinta) ja heidän nimensä sekä osoite (projektio). Lisäksi relaatiomallin tietokanta on eheä, kun tiedot ovat oikein, ristiriidattomia sekä täsmäivät oikeaan maailmaan. Tietokannan eheys vaarantuu, mikäli esimerkiksi sama asiakas tai asiakkaan osoite tallennetaan kahteen kertaan, eikä ole varmuutta mikä osoite on oikea.

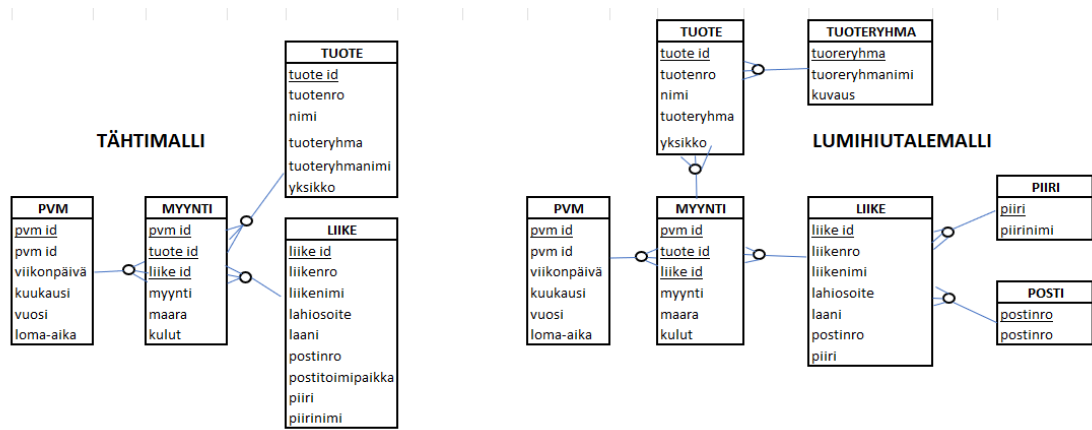
### 3.5 Tietoallas ja tietovarasto

Tietoallas tarkoittaa käytännössä tiedostojärjestelmää, johon voidaan tallentaa eri muotoista tietoa eri lähdejärjestelmistä, myös ns. jäsentämätöntä epästrukturoitua informaatiota, kuten videoita ja kuvia. Tammisen mukaan (2019) tietoallas on tekninen ympäristö, esimerkiksi pilvipalvelu, jossa on mahdollista varastoida ja käsitellä suuria määriä tietoja.

Varsinaiseen tietovarastoon varastoidaan ja tallennetaan määrätyn lataushetkin tietovarastomallin mukaisesti yrityksen sisäistä ja ulkoisista lähteistä kerättyä tietoa. Tietovarastossa on usein eri kerroksia, jotka jaetaan esimerkiksi raakadata ja esitysdata kerrokseen. Raakadata kerroksessa kerätty data on alkuvaiheessa yleensä alkuperäislähteen mukaisessa muodossa, mikä jatkojalostetaan yrityksen käyttötarpeiden mukaisesti tietomalliin rakennettujen sääntöjen mukaisesti. Lisäksi raakadatasta käytetään vain sitä tietoa, mikä on yritykselle merkityksellistä liiketoiminnan näkökulmasta. Esitysdata kerrosta on jalostettu valmiiksi, jotta erilaisilla raportointi- ja BI-työkaluilla tiedon hyödyntäminen sekä analysointi olisi helppoa.

Tietokanta (Database, DB) on moniselitteinen, jossain määrin tuotekohtainen käsite. Yleisesti ottaen tietokanta on loogisesti yhteenkuuluvien, tallennettujen tietojen joukko, jota voidaan helposti käsitellä tietokantakielellä (kuten SQL). Tietokannassa olevia tietoja hallinnoi erityinen ohjelmisto, tietokannan hallintajärjestelmä eli TKHJ (Database Management System, DBMS) (Hovi ym. 2005, 4).

Tietovarastomalleja on erilaisia, joiden käyttötarkoitus puoltaa tarvittavan mallin valitsemista. Tyypilliset mallit ovat ns. tähti- ja lumihiutalemalli. Tähtimalli nimensä mukaisesti muistuttaa ulkoisen rakenteensa takia tähteä, kun taas moniulotteisempi lumihiutalemalli muistuttaa yksittäistä lumihiutaletta.



Taulukko 3. Tietovaraston tähti- ja lumihiihtomallit (mukaillen Hovi ym. 2005, 136 ja 139)

Tietovarastomallit sisältävät fakta- ja dimensiotauluja. Faktataulut sisältävät tyypillisesti numeerisia arvoja, jotka kuvaavat esimerkiksi kappaleita tai euromääriä. Hovin ym. (2005, 135) mukaan, faktataulut ovat tapahtumatyyppisiä, kuten myyntiä tai kirjanpitoa kuvaavia arvoja, kun taas dimensiotauluihin kerätään tarvittavia luokittelutekijöitä, jotka ovat usein luonteeltaan staattisia. Taulukossa kolme (3) faktataulua kuvaa taulu "myynti", kaikki muut taulut koskien molempia malliesimerkkejä ovat ns. dimensiotauluja.

### 3.6 Pilvipalvelu

Pilvipalvelun määritelmänä voidaan käyttää Yhdysvaltojen julkishallinnon standardiin perustuvaa kuvausta, jonka Salo (2012, 17) tuo seuraavasti esiin, pilvipalvelut ovat toimintamalli, joka mahdollistaa pääsyn vapaasti konfiguroitaviin ja skaalautuviin tietotekniikkaresursseihin, jotka voidaan ottaa käyttöön tai poistaa käytöstä nopeasti. Lisäksi Salo lisää (2012, 20), että pilvipalvelut voidaan jakaa tyypillisesti kolmeen eri kategoriaan palvelusäilytyksen perustuen. IAAS-palvelu (infrastructure as a service) tarjoaa kokonaisvaltaista infrastruktuuri palvelua, PAAS-palvelu (platform as a service) tarjoaa yritykselle alustan sovellusten käyttämistä kehittämistarkoituksiin. SAAS-palvelussa (software as a service) yritys käyttää sovelluksia palveluina ilman niiden omistamista tai päivittämistarvetta, palvelun tarjoama yritys vastaa sovelluksen ylläpidosta sopimuksen mukaisesti.

Salon mukaan (2014, 93) pilvipalveluista voidaan tunnistaa erityispiirteitä, jotka ovat:

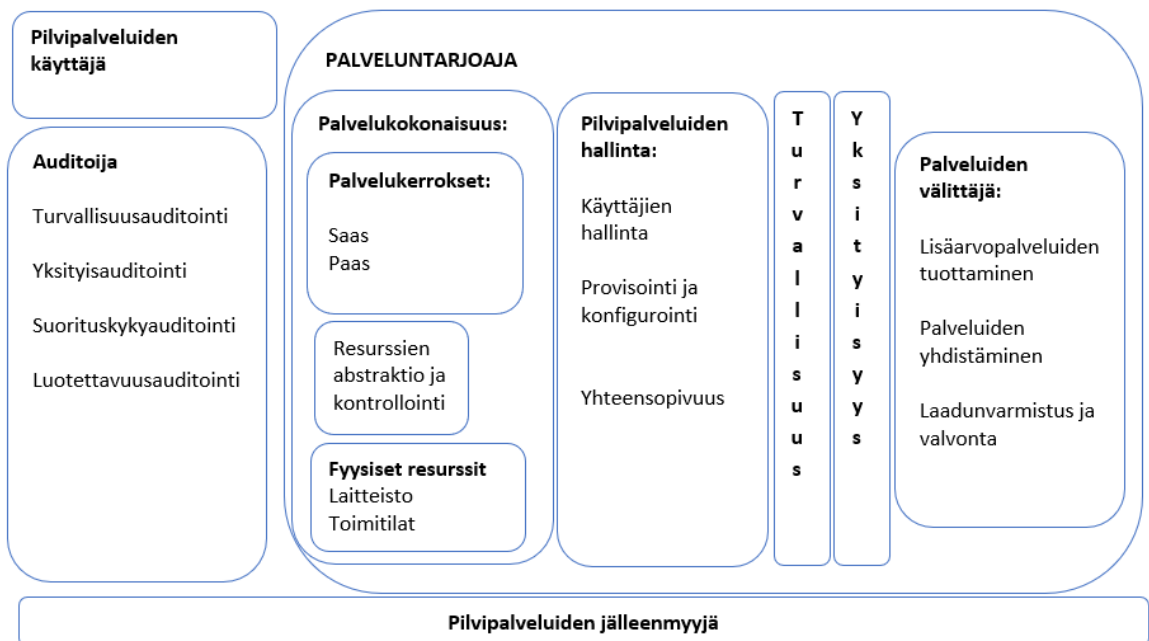
- palvelujen käytön aktivoiminen tai lopettaminen itsenäisesti ilman erillistä yhteydenottoa esimerkiksi asiakaspalveluun tai myyntiedustajaan
- palveluihin on suora pääsy erilaisilla päätelaitteilla, kun internet verkkoysteys on vain saatavilla
- resurssien yhteiskäyttö eri toimijoiden kanssa määrätyin rajoituksin ja suojaus

- palveluiden suuri joustavuus eli käytettävät palvelut skaalautuvat yrityksen tarpeen mukaisesti nopeasti korkeammalle tai matalammalle tasolle. Yritys voi lisätä esimerkiksi palveluiden laskenta tai tallennuskapasiteettia reaaliaikaisesti
- palveluiden käyttöasteen tarkka mittaaminen, jolloin yritys maksaa vain käyttämistään palveluistaan ja palveluntarjoaja tunnistaa yrityksen palveluiden resurssitarpeen

Salo mainitsee (2014, 123), että Amazon Web Service avasi datan tallentamiseen tarkoitettun S3-palvelun vuonna 2006, mikä on laajentunut laajaksi valikoimaksi erilaisia pilvipalveluita. AWS-palvelut voidaan katsoa kuuluvan IAAS-palvelukategoriaan, vaikka palveluiden laajentumisen myötä, sen voidaan katsoa sisältävän myös PAAS- ja SAAS-kategorioiden sopivia palveluita.

Salo lisää (2014, 124), että Amazon Webservice jakaa palvelunsa seuraaviin sisältöihin:

- laskentaan ja tietoliikennepalvelut
- tallennustilaan sekä sisällönjakamisen verkostoon
- tietokantapalveluihin
- analytiikan kehittämisspalvelut
- sovelluspalveluihin
- käyttöönotto ja hallintapalvelut
- AWS-markkinat



Kuva 1. Pilvipalveluiden arkkitehtuuri (mukaillen Salo 2012, 29)



### 3.7 Tiedon visualisointi, business intelligence -työkalut ja analytiikka

Tiedon visualisointi helpottaa käyttäjää ymmärryksen kasvattamisessa kuinka suuresta datamassasta voi tehdä nopeasti havaintoja ja johtopäätöksiä esimerkiksi tietyn tuotteen asiakaskäyttöasteen muutoksista ja trendeistä. Visualisoinnin kautta käyttäjän on helppompaa omaksua nopeasti tietoa ja havaita muutoksen nopeutta esimerkiksi pitkistä aikasarjoista, kun tiedot on esitetty graafisessa muodossa. Collin ja Saarelainen (2016, 206) kuvaavat toiminnallisuuden helppokäyttöisyyden merkitystä seuraavasti, ”Käyttöliittymässä on voitava porautua dataan tarkasti. Käyttäjä saa lisäarvoa, kun raportointi, erilaiset ennusteet, poikkeustiedot ja hälytykset visualisoidaan”.

Hovin ym. mukaan (2009, 74) business intelligence (BI) ratkaisut helpottavat yrityksen ja organisaatioiden liiketoimintaa kuvaavan informaation ymmärrystä ja merkitystä. Tämä kasvanut ymmärrys edesauttaa liiketoimintapäätöksissä sekä ohjaa yrityksen toimintaan oikeaan suuntaan. Lisäksi BI-ratkaisut mahdollistavat, että henkilöstön ei tarvitse tuntea syvällisesti tietojärjestelmiä tai niiden rakenteita, vaan voivat hyödyntää tietoja silloin, kun niitä todella tarvitsevat. Optimitilanteessa ratkaisut ovat yksinkertaisia ja helppokäyttöisiä, tiedot ovat visualisoitu ja niiden hakeminen sekä käyttö onnistuu kaikilta helposti kuin tavallinen nettisurffailu.

Analytiikan tarkoitus on hyödyntää kerättyä dataa ja muodostaa tiedosta analyysin perusteella lisäarvoa ja näin helpottaa tarkoituksenmukaista päätöksentekoa. Collin ja Saarelainen (2016, 206) kuvaa analytiikan hyödyn siten, että data mahdollistaa informaation, joka helpottaa päätöksentekoa, luo ennustamiseen malleja sekä tehostaa kannattavuutta ja laatua.

Aki Taanilan mukaan (2012) dataa analysoidaan, jotta saadaan käyttökelpoista tietoa suunnittelun ja päätöksenteon tueksi ja perusteluiksi. Kuvailevassa ja selittävässä analytiikassa tavoitellaan yleensä yhtä tai useampaa seuraavista. Jakaumien havainnollistaminen, erojen havaitseminen, riippuvuuksien tunnistaminen, kehityssuuntien esittäminen ja poikkeuksien löytäminen.

Koneoppimisessa koneelle ei ole määritelty toimintaohjetta jokaista erillistä tilannetta varten, vaan kone oppii itsenäisesti datasta. Suurin osa tekoälyn sovelluksista on koneoppimista. Ohjattu oppiminen: koneelle annetaan oikea vastaus opetusdatassa. Ohjaamaton oppiminen: kone päättää asioita datassa olevien säännönmukaisuuksien ja suhteiden pohjalta. Vahvistusoppiminen: koneelle annetaan palautetta siitä, kuinka onnistuneesti se toimii eri tilanteissa ilman, että annetaan oikeita vastauksia. (Merilehto 2018, 19).

## 4 Tietovarastointiprosessin kehitysprojekti ja tutkimusmenetelmät

Tässä luvussa kerrotaan ensin kehittämisprojektista ja sen sisältämästä tavoitteista tuottaa yrityksen omiin liiketoiminnallisiin tarpeisiin tarkoitettu tietovarastointiprosessi sellaisessa muodossa, että se kasvattaa myös yrityksen kilpailukyvykkyyttä markkinoilla. Edellisen lisäksi tässä selvitetään ne menetelmävalinnat, joita käytetään kehittämistehtävässä asetettujen tutkimuskysymysten ratkaisujen etsimiseksi.

Kehittämisestä kohdeyrityksenä on finanssitoimialan yritys, joka tarjoaa pankki- ja vakuutuspalveluita eri asiakassektoreille. Koska yrityksen keskeiset tietovarastointiratkaisut olivat ulkoistettu, tämä vaikutti käytettävissä olevan tiedon hyödyntämisen mahdollisuuksiin yrityksen eri organisaatio tasoilla, ja lisäksi vain rajallinen osa syntyvästä tiedosta oli hyödynnettävissä. Vanhat ICT-ratkaisut eivät myöskään tukeneet analytiikan kehittämistä liiketoiminnan tueksi, koska esimerkiksi riittäviä aikasarjoja tarvittavista tiedoista ei ole ollut saatavilla tai tiedot päivittyvät tietovarastoon liian harvoin analytiikan täysimääräisen hyödyntämisen näkökulmasta. Tämä kehittämisprojekti toteutettiin pääasiallisesti vuosien 2020 ja 2021 aikana ja sen kehittämisen kokonaishallinnasta vastasi tietohallintoyksikkö yrityksen projektien hyväksymismenettelyprosessien mukaisesti. Hovin ym. mukaan (2009, 144) liiketoimintatarpeiden määrittäminen on tärkeää tietovarastoprojektissa. Konkreettiset, realistiset ja eteenpäin vievät tavoitteet luovan pohjan onnistuneelle projektille. Tietovarastoinnin kehittämisprojektin keskeiset osa-alueet on kuvattu tarkemmin kuvassa kaksi (2).



Kuva 2. Tietovarastoinnin kehittämisprojektin keskeiset osa-alueet

Tyypillisesti projektisuunnitelma vaatii hyväksynnän, jonka jälkeen projektin varsinainen toteutusvaihe voi alkaa. Tarkemmassa vaihemallissa projektin toteutusvaihe jaetaan edelleen useisiin alivaiheisiin. Näin on järkevä menetellä varsinkin pidempien ja laajempien projektien kohdalla. Projektisalkunhallinnan kannalta on olennaista, että varsinkin pitkää projektia voidaan tarvittaessa arvioida myös toteutuksen aikana siirryttäessä alivaiheesta toiseen. Tällöin on mahdollista suunnata uudelleen projektia esimerkiksi strategisten tavoitteiden muuttumisen vuoksi. (Lehtonen, Lindblom, Simonen & Korpinen, 2006, 30.)

Tässä kehittämistehtävässä toteutetaan yritykselle omaan hallintaan liittyvä tietovarastoinnin kokonaisratkaisu pilvipalveluun, johon välitetään kaikki liiketoiminnan tarvitsema tieto. Tietovarastointi tukee liiketoiminnan analysointikyvykkyyksiä, tiedon keräämistä ja välittämistä valittuihin operatiivisiin järjestelmiin, automaattista raportointia sekä henkilöstön manuaaliraportoinnin tuottamisen tukea liiketoimintaymmärryksen lisäämiseksi. Tietoa voidaan hyödyntää eri organisaatiotasoilla BI-käyttöliittymä ja raportointitukijärjestelmin. Kuten luvussa kaksi (luku 2) on kerrottu kehittämisprojektille asetetuista tavoitteista, on siis oleellista selvittää tutkimuksen keinoin se tietovarastoinnin kehittämisprosessi, joka tukee yrityksen asettamia käytännön liiketoimintatavoitteita. Edellisen lisäksi on selvitettävä ne osa-alueet, jotka vaikuttavat merkitsevästi, kun tietovarastointialustaa uudistetaan pilvipalveluun. Tuloksena voidaan olettaa, että se uudistettu prosessi toteuttaa tehokkaan käytännön toimintamallin yritykselle, jonka tuotoksena liiketoimintajohto ja henkilöstö eri organisaatiotasoilla saa päätöksenteon tueksi ajantasaista ja tarkoituksenmukaista tietoa (dataa) pilvipalvelualustaan toteutetusta tietovarastosta.

#### **4.1 Kehittämisen- ja tutkimusstrategia**

Tietovarastoinnin kehittämisprosessi on monivaiheinen ja edellyttää laajaa sisäistä yhteistyötä yrityksen eri organisaatiotasoilla sekä valittujen ulkoisten sidosryhmien kanssa. Tutkimuksellinen kehittämistapa on yhdistelmä konstruktivisen ja toimintatutkimuksen menetelmistä. Bisterin mukaan (2019, 42.) toimintatutkimuksen lähestymistavalla tarkoitetaan, jossa tutkija osallistuu tutkittavana olevan kohdeyhteisön toimintaan ja arkeen. Tutkija pyrkii siis yhdessä yhteisön jäsenten kanssa ratkaisemaan yhdessä määritellyt ongelmat, saavuttamaan asetetut tavoitteet ja päämäärät.

Ojasalon, Moilasen ja Ritalahden mukaan (2015, 24) tutkimuksellisessa kehittämishankkeessa on keskeistä tunnistaa kehityksen kohde sekä kehittämiseen vaikuttavien tekijöiden merkitys.

Ojasalo ym. (2015, 68) esittävät konstruktiivisen tutkimusprosessin vaiheiksi:

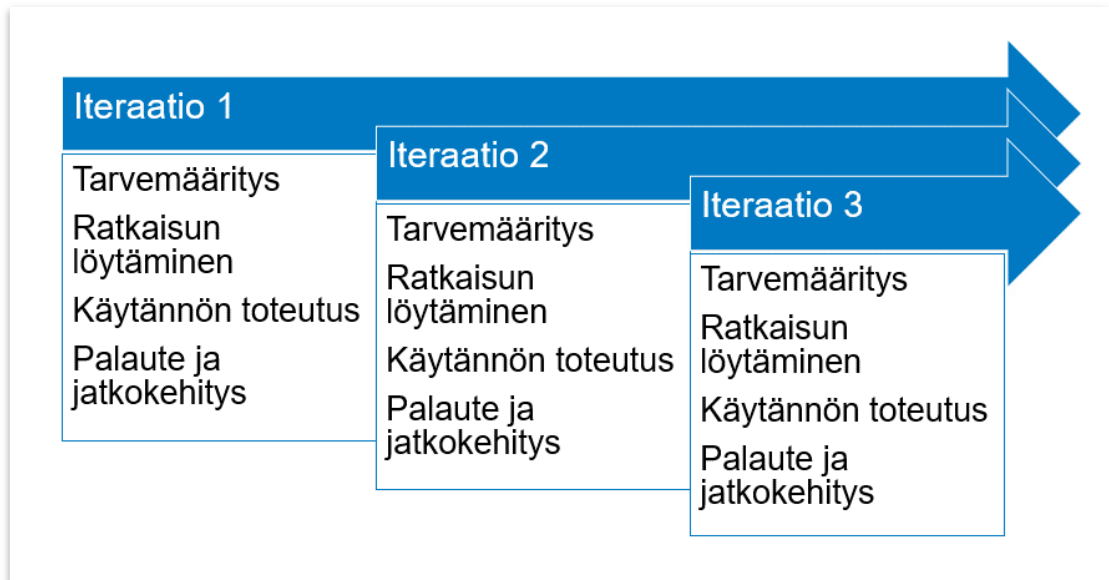
1. Mielekkään ongelman etsiminen
2. Syvällisen teoreettisen ja käytännöllisen tiedon hankinta tutkimuksen ja kehittämisen kohteesta
3. Ratkaisujen laatiminen
4. Ratkaisun toimivuuden testaus ja konstruktion oikeellisuuden osoittaminen
5. Ratkaisussa käytettyjen teoriakytkeiden näyttäminen ja ratkaisun uutuusarvon osoittaminen
6. Ratkaisun soveltamisalueen laajuuden tarkastelu

Tietovarastoa loppukäyttäjät voivat hyödyntää raportointityökalun avulla itsenäisesti, sekä myös valmisraporttien kautta. Tietovarasto tukee ensivaiheessa myös pilottiyksikön kampanjoiden luontia eri kohderyhmille kertyneet asiakasdatan perusteella. Ojasalo ym. (2015, 66) mukaan konstruktiivisessa tutkimuksessa on tavoitteena saada ongelmaan teoreettisesti perusteltu ja uusi ratkaisu, mikä luo liiketoimintaan uutta tietoa.

#### **4.2 Tietovarastointiuudistuksen kehitysmalli**

Tarvittavat muutostyöt kehitystehtävässä ovat hyvin laaja-alaiset liittyen tarvittaviin sidosryhmiin sekä järjestelmäalustojen tukiratkaisuun liittyen, tästä syystä kehitysmallissa kuvataan keskeiset kehityseriaatit eri lähdejärjestelmistä kerätyn datan välittämiseksi tietovarastoon. Eri liiketoimintayksiköihin liittyviä tiedon määrittelyprojekteja on käynnissä useita samanaikaisesti ja uudistuksen kompleksisuuden takia kehitysmenetelmänä käytetään muutaman viikonpituisia ajanjaksoja. Kehittämisessä käytetään ketteriä kehitysmenetelmiä, jossa käytännön iteraatiot voivat olla hyvin lyhyitä, vain muutaman päivän pituisia liiketoimintatarpeista tiedon mallinnustyötä tekeville.

Kuten kuvassa kolme (kuva 3, 17) on esitetty iteraatiomalli, on suuren tehtävän ratkaiseminen toisiaan seuraavissa sykleissä (kierroksissa) tuo projektiin hallittavuutta ja läpinäkyvyyttä, koska tuotos on tarkasteltavissa kunkin syklin jälkeen. Kokonaisuus rakentuu vähitellen pala palalta. Edellä kuvattu syklinen ja iteratiivinen toimintamalli on toimintatutkimukselle luonteenomaista. (Bister, T. 2019, 43.)



Kuva 3. Toistuvat iteraatiot yrityksen tietovarastoinnin kehitysprojektissa

#### **Kehittämisen toimintamallin keskeiset periaatteet:**

1. Eri liiketoimintayksiköiden tavoitetilan tarpeiden kartoitus (liiketoiminnan tukeminen)
2. Tietovarastointiratkaisun tukijärjestelmien ja alusta-arkkitehtuurien valinnat
3. Yksikön keskeisten lähdejärjestelmien selvitys ja tietosisällön tarvemääritys liiketoiminta-asiantuntijoiden kanssa
4. Lähdejärjestelmistä saatavan tiedon mallinnus loogiseksi ja fyysisiksi malleiksi huomioiden kokonaisuus
5. Tiedon mallinnus liiketoimintayksikön tavoitetilan mukaisesti määriteltyihin järjestelmiin sekä käyttöliittymiin (BI-käyttöliittymän raportointituki)
6. Käytännön toteutus ja testaus liiketoiminta-asiantuntijoiden kanssa
7. Testauksen perusteella havaittujen yksityiskohtien muutostarpeiden huomiointi ja muutokset
8. Tuotantoon vienti ja tuotannon verifiointi testaus liiketoiminnan asiantuntijoiden kanssa
9. Tuotannon muutos- ja jatkokehityshallinta sekä järjestelmien eheyden seuranta

Kehittämisen kohteena on tiedon varastointi ja tiedon helppo saatavuus tukemaan henkilöstön liiketoimintapäätöksiä. Tutkimusmenetelmän kohteena on tällöin koko organisaatiolle saavutettavat hyödyt, jotka mahdollistavat uudistetut järjestelmäratkaisut. Onnistuneen loppuratkaisun saavuttamiseksi tarvitaan henkilöstön eli koko kehittämisorganisaation osallistumista, palautetta sekä käytännön testaamista. Tähän perustuen kehittämismenetelmänä hyödynnetään myös osittain konstruktivisia tutkimusmenetelmiä.

### 4.3 Tiedon keruumenetelmät

Tiedon keruumenetelminä käytettiin liiketoiminnan ja tietohallinnon avainhenkilöiden haastatteluja. Tietoja kerättiin lisäksi sähköisen kyselylomakkeen kautta sekä erillisten projekti-palaverien informaatiota hyödyntäen. Avainhenkilöt valittiin asiantuntijuuden perusteella kehitysprojektin sisällön ja tavoitetilan ymmärrykseen perustuen.

Ojasalo ym. (2015, 37) mukaan, toimintatutkimuksessa on keskeistä, että yrityksen organisaation henkilöstö osallistuu aktiivisesti kehitystyöhön ja että, konstruktivinen tutkimus soveltuu hyvin lähestymistavaksi, kun esimerkiksi tehtävänä on luoda konkreettinen tuotos, esimerkiksi uusi tuote, tietojärjestelmä, malli tai suunnitelma.

Liiketoiminnan asiantuntijoille ja kehittämispäälliköille lähetettiin sähköinen kysely, jossa selvitettiin tietovarastointiprojektista saatavia näkökulmia, hyötyjä ja olemassa olevia puutteita. Haastatteluja ja tiedonkeruuta toteutettiin kehitystyön etenemisen eri sykleissä projektin osallisilta.

Keskeiset sisäiset avainkohderyhmät ovat haastattelujen ja tiedon keräyksen näkökulmasta henkilöresurssit, jotka kuuluvat organisatorisesti liiketoimintajohtoon, liiketoiminnan kehitys- ja tuotepäällikkö asiantuntijoihin sekä tietoarkkitehti ja DW/BI-asiantuntijatasoon. Hovin ym. (2005, 70) mukaan, haastatteluissa on hyödyllistä pyytää tulevaa tietovarastoa käyttäviltä nykyisiä raporttimalleja, jotta vuorovaikutteinen ymmärrys lisääntyy sekä ymmärrys organisaation tarpeista. Haastatteluissa on hyödyllistä myös selvittää käsitteiden ominaisuuksia sekä niihin liittyviä suhteita.

Järjestelmäratkaisujen toteuttaminen edellyttää lisäksi kumppanuusyhteistyöhön liittyvien avainhenkilöiden haastatteluja, johtuen organisaation ulkopuolella olevan datan hyödyntämiseksi keskitetysti yrityksen omassa tietokannassa. Haastatteluista saatavat tiedot dokumentoitiin kirjallisesti, joista saatavaa sisältöä hyödynnettiin varsinaisessa analyysissä sekä tehtävissä johtopäätöksissä. Lisäksi ulkopuolisen datan hyödyntämiseksi liittyi tiedonvälitykseen liittyviä järjestelmäselvityksiä.

Selvityksessä olivat myös olemassa oleva tietovarastointi ja raportointityökalujen järjestelmätekniset ratkaisut sekä vaadittavat muutosmäärittelyt käytännön toteuttamiseksi. Järjestelmäteknisten ratkaisujen kartoittamisessa hyödynnettiin organisaation sisäisiä ICT-yksikkö resursseja sekä ulkopuolisia yhteistyökumppanuuksia.

Kehittämistiimin jäsenet ottavat joko suoraan tai epäsuorasti osaa myös tutkimuksen toteuttamiseen. Juuri osallisena olevien itse suorittama aineiston keruu on eräs toimintatutkimukseen liitetty ominaispiirre. (Heikkinen, Huttunen & Moilanen, 1999, 75.)

#### 4.4 Haastattelujen toteutus ja tietojen kerääminen kyselylomakkeella

Kysymyksenasettelu haastatteluissa eroaa riippuen haastateltavasta henkilöstöstä tai esimerkiksi siitä, kuuluuko haastateltava organisaation ulkopuoliseen sidosryhmään. Haastattelukysymysten erilaisuus henkilösidonnoisesti perustuu myös haastateltavan rooliin kehitysprojektissa sekä yleisesti vastuualueensa kautta kerättyyn asiantuntemukseen. Lisäksi määrittelytietoa kerättiin projektipalaverista, joissa avointa ratkaisukeskeistä keskustelua käytiin esiin nousevista kysymyksistä. Haastateltaville kerrottiin etukäteen haastattelujen merkitys ja mitä hyötyjä niistä tavoitellaan tietovarastointiuudistukseen liittyen (taulukko 4). Lisäksi haastateltaville kerrottiin tutkimukseen liittyvän luottamuksellisen tiedon suojaamisesta sisältäen henkilötietosuojan anonymiteetista.

Tehtävänimike	Tehtäväalue	Haastattelu / Kyselylomake
Tietoarkkitehti (Johtava)	Tietovarastointi	H1
Tietoarkkitehti	Tietovarastointi	H2
DW/BI-asiantuntija	Tietovarastointi	H3
DW/BI-asiantuntija	Tietovarastointi	H4
DW/BI-asiantuntija (ulkoinen)	Tietovarastointi	H5
Tietoarkkitehti (ulkoinen)	Tietovarastointi	H6
Johtaja	Liiketoiminta	Kyselylomake
Kehityspäällikkö	Liiketoiminta	Kyselylomake
Kehityspäällikkö	Liiketoiminta	Kyselylomake
Kehityspäällikkö	Liiketoiminta	Kyselylomake
Asiantuntija	Liiketoiminta	Kyselylomake
Asiantuntija	Liiketoiminta	Kyselylomake
Asiantuntija	Liiketoiminta	Kyselylomake

Taulukko 4. Tietolähteet (haastattelu, kyselylomake, projektikokoukset).

Haastattelut keskittyivät siis tietovarastoinnin asiantuntijoihin ja datan hallinnan kehittäjiin, joiden vastuualueena oli hallita tietovarastoinnin kokonaisprojektia. Liiketoiminnan edustajien roolit projektissa keskittyivät enemmän tarvittavan datan ja raportoinnin määrittelyyn sekä testaukseen. Liiketoiminnan määrittely- ja selvitystyötä ohjasi tietovarastoinnin asiantuntijat, jotta tarvittavat selvitykset valmistuivat riittävän yksityiskohtaisella tasolla sekä varsinaisen liiketoiminnan tarpeet oli mahdollista selvittää.

## 4.5 Kehittämistehtävän tutkimuksen toteutus

Tutkimuksessa on käytetty haastattelumenetelmää, joka soveltuu tähän kehittämistehtävän tutkimusosuuden läpikäyntiin perusteilla, kun kehittämisprojektin tehtäväkokonaisuus on laaja sekä asiantuntijuus on jakautunut organisaatiossa. Haastattelut tehdään teknisesti tavoitetilan ja ymmärryksen lisäämiseksi puolistrukturoidun ja avoimen haastattelumenetelmin. Ojasalo ym. (2015, 108) mukaan puolistrukturoidussa haastattelussa kysymykset on jo etukäteen laadittu, mutta niiden esittämisjärjestys ja kysymyksien sisältö edelleen muuttua sekä avoin haastattelu luo vapauksia, koska aiheesta tai tunnistetusta ongelmasta voidaan avoimesti keskustella.

Hovin ym. mukaan (2009, 151) projektin onnistumisessa on olennaista, että projektiryhmään kuuluvat hallitsevat oman alueensa ja ovat halukkaita kehittämään sen toimintaa sekä ymmärtävät tietojen ja tietovaraston merkityksen yrityksen toiminnalle.

Kaikki haastattelut ja tutkimussisällön kerääminen on tapahtunut etäyhteyksien kautta, osa haastatteluista on tehty projektiin liittyvien etäkokouksien yhteydessä vain määrättyä osa-alueelta haastateltavan vastuualue huomioiden. Keskeisen aineiston kerääminen on toteutettu vuosien 2021 ja 2022 aikana. Käytännön tietovarastoinnin kyvykkyyksien kehitystyö on kuitenkin jatkuvaa muutosprosessia myös tulevaisuudessa liiketoimintatarpeet huomioiden.

Bisterin mukaan (2019, 35) kvalitatiivisen tutkimuksen luonteeseen kuuluu monipuolisen ja syvällisen kuvan muodostaminen tutkittavasta kohteesta. Tietojärjestelmiin liittyy sen elinkaaren eri vaiheissa paljon ihmistyötä, usein luovuutta ja korkeaa asiantuntemusta edellyttävää. Ja Bister lisää että, haastattelujen kautta välittyy faktatietoa, myös tunteita, asenteita, vaikeitakin pulmia, joiden esille saaminen muilla keinoin ei ole mahdollista.

### Tekninen toteutus

- Tietovarastoinnin järjestelmä arkkitehtuuri --> Minkälaisia ICT-ratkaisuja ja kumppanuuksia tarvitaan yrityksen olemassa oleva alusta-arkkitehtuuri huomioiden?
- Miten eri lähdejärjestelmien rajapintoja ja eri aineistomuotoja hallitaan?
- Kuinka tietovarasto kykenee tuottamaan tarvittavan datan eri käyttäjille sekä tuottamaan tarvittavia ennusteita käyttäjäkohtaisten tarpeiden mukaisesti?
- Mitkä tietovarastoinnin taulu ja tietosisältö rakenteet palvelevat liiketoiminnan tavoitetilaa? Mitkä arkkitehtuuriset rakenteet ja määrittelyt mahdollistavat tulevaisuudessa syntyvien muutostarpeiden tekemisen joustavasti?
- Kuinka järjestelmäarkkitehtuurin valinnoissa on huomioitu tietosuojaan liittyvät riskinäkökulmat?



Haastattelut ovat tärkeitä, koska ne antavat arvokasta tietoa, joka voi vaikuttaa tietokantarakenteen suunnitteluun. Esimerkiksi taulujen yhteyksiä käsitellessäsi voi tiettyyn yhteyteen liittyvän osallistumistyyppin ja osallistumisasteen selvittäminen olla joskus vaikeaa. Ainoa tapa, jolla voit selvittää näiden yhteyksien ominaisuuksien oikeanlaiset arvot, on haastatella asiasta tietäviä henkilöitä. (Hernandez 2000, 72).

### **Toimintojen kehittäminen**

- Lähtötilanne: Miten nykyiset puutteet datan laadussa ja sen hyödyntämisen mahdollisuuksissa tunnistetaan kokonaisvaltaisesti eri liiketoiminnoittain?
- Mitä tietoja eri yksiköt käyttävät, mitkä tiedoista ovat kriittistä ydintietoa?
- Mihin tietoa tarvitaan? Miten saatavilla olevaa tietoa käytetään eri organisaatioiden tasolla?
- Mikä on raporttien ja datan käyttöaste tällä hetkellä? Mikä on turhaa? Mikä on vajavaista?
- Mitä BI-työkaluja tarvitaan, jotta liiketoiminta kykenee hyödyntämään monipuolisemmin ja laajemmin saatavilla olevaa dataa hyödykseen
- Miten valmis- ja automaattiraportit tuotetaan kustannustehokkaasti?
- Mistä ulkoisista tietovarastoista ja lähteistä on tarkoitus välittää tietoa uudelle tietovarastointialustalle?
- Tekoälyn ja koneoppimisen kehittäminen? Mitä tarkoittaa käytännön tasolla?

### **Käyttöönotto ja jalkauttaminen**

- Kuinka testausvaihe toteutetaan?
- Tuotantoon menon vaiheet?
- Käytännön viestintä ja koulutus?
- Kuinka uudistetun tietovarastointiratkaisujen liiketoiminnallisia hyötyjä voidaan mitata?
- Jatkokehittämisen hallinta?

Haastatteluista saatu tieto kerättiin ja jäsennettiin aihealueittain kirjalliseen muotoon ilman nauhoituksia. Nauhoituksesta luopuminen oli tietoinen päätös vapaan ja rennon ilmapiirin varmistamiseksi. Bisterin mukaan (2019, 51) tutkimuksen teon yhteydessä kootusta aineistosta on pystyttävä tekemään luotettavia johtopäätöksiä. Tutkimuksen tavoitteena saattaa tulokset sellaiseen muotoon, että lukijan on helppo tulkita niitä. Tavoitteena on selkeys ja tiiviys.

#### **4.5.1 Kyselylomakkeen avoimet sisältökysymykset**

Microsoft Forms -kyselylomakkeelle selvitettiin seuraavia näkemyksiä liiketoiminnan asiantuntijoilta ja kehittämispäälliköiltä, joilla suora yhteys välillisesti kehitysprojektiin oman

liiketoiminnan kautta. Kyselyyn vastanneet osallistuivat liiketoimintansa kehitysalueelta raportointi määräytykseen tarkoituksenmukaisen datan hyödyntämiseksi sekä datasta määriteltävien raporttien muodostamiseksi niin numeeristen lukujen kuin visuaalisten kuvien (graafien) toteuttamiseksi.

Lomake sisälsi seuraavat avoimet kysymykset:

1. Tietovaraston uudistusprojektista aluksi, kuvaile lyhyesti, miten tämä helpottaa työtä tulevaisuudessa tai mitä hyötyjä siitä saat?
2. Mitä pidät itsellesi tärkeimpänä asiana työn edistämisen näkökulmasta, kun uudistusprojekti valmistuu?
3. Mitä hyötyjä näet uudistusprojektista liiketoimintayksiköllesi?
4. Mitä hyötyjä näet uudistusprojektista asiakkaille?
5. Mikä on arvokkain asia mielestäsi koko yrityksen näkökulmasta mitä tietovarastoalustan uudistusprojekti tarjoaa?
6. Mitä haasteita raporttien ja liiketoimintatiedon hankinnassa on ollut ennen uudistusprojektiä?
7. Mitä arvelet hyödyiksi, kun uusi tietovarastoalusta upotetaan ns. pilvialustaan? Mitä hyötyjä tästä meille tai asiakkaille voisi arviosi mukaan olla?
8. Miten analytiikka voisi auttaa työn tekemistäsi?
9. Mitä raporttien visualisointi hyödyttää? Esimerkiksi Power-Bi käyttöliittymän graafit?
10. Mitä raportointi tai liiketoimintatietotarpeita uudistusprojekti ei ratkaise työsi osalta?

#### 4.6 Tiedon analysointimenetelmät, teemahaastattelujen vastaukset ja analyysi

Tehtyjen haastattelujen, lomakekyselyjen sekä projektiin liittyvien kokousten perusteella analysoidaan kerättyä tietoa. Projektiin liittyi lisäksi eri vaiheisia järjestelmäarkkitehtuuriin liittyviä selvityksiä.

Tutkimuskysymykset (kehittämisen tehtäväkokonaisuus)	Tietoperusta (kappalenumero)	Haastattelu (H) Kysymyslomake (K)	Tutkimus- ja toteutus tulokset
Mikä on tarkoituksenmukainen tietovarastoinnin kehittämisenprosessi?	3.1 - 3.5	H1, H2, H3, H6 K1, K2, K5	4.7, 4.9 - 4.12
Mitkä ovat ne osa-alueet, jotka ovat huomioitava, kun tietovarastointialustaa uudistetaan pilvipalveluun?	3.6	H1, H2, H4 K3, K7	4.8
Mikä on se tehokas toimintamalli, jonka avulla uudistetusta tietovarastointiratkaisusta voidaan hyödyntää dataa liiketoiminnan tueksi eri organisaatioiden tasoille?	3.7	H3, H4, H5 K4, K6, K8, K9, K10	4.13

Taulukko 5. Tutkimuskysymykset ja analyysin rakentuminen

Ajallisesti ensimmäinen on aineiston esikäsittely- ja valmisteluvaihe. Toinen vaihe on varsinainen tutkimusanalyysi, jolloin tutkijan pöydälle ensimmäisessä vaiheessa asetetut palaset saavat jonkin merkityksensä, niitä suhteutetaan toisiinsa ja tutkimuksen tavoitteisiin, niitä eritellään ja yhdistellään, niistä jotkut hylätään tarkemman arvioinnin nojalla juuri siihen tutkimukseen, hyödyttöminä tai muuten tieteellisesti mielenkiinnottomina ja niin edelleen. (Pihlaja 2004, 153–154.)

Haettiin vastaukset ja määritykset seuraaviin kysymyksiin:

- Nykyinen järjestelmä arkkitehtuuri on kuvattu sekä tunnistettu keskeiset muutostarpeet nykyisen tietokannan ja raportointikyvykkyyksien kehittämiseksi
- Tavoitetila ja toimittajat ovat selvillä tietovarastoinnin järjestelmä ja alustaratkaisuihin huomioiden yrityksen kokonaisarkkitehtuuri
- Keskeisistä muutostiedoista on tehty määrittelyraportti sekä selvitetty vastaukset niiden ratkaisemiseksi
- Järjestelmiin liittyvistä muutostiedoista on laadittu vaatimus- ja määrittelydokumentaatio
- Kehittämistehtävälle on asetettu budjetti sekä kannattavuuslaskelma syntyvistä kuluista ja arvioiduista kustannussäästöistä
- Tunnistettu kehittämisprojektin tuottamat operatiiviset hyödyt sekä saavutettavat käytännön hyödyt liiketoiminnalle
- Kumppanuus- ja sidosryhmäverkostoihin liittyvät muutostyövaikutukset on selvitetty
- Ulkopuolisen asiantuntijuuden tarve määrätyille tukitehtäville on selvitetty
- Aikataulutavoitteet eri projektinvaiheille on vahvistettu
- Projektin aikana tapahtuvat raportoinnin määrittelytehtävät eri liiketoimintayksiköiden kanssa on aikataulutettu sekä projektin viestinnästä on toteutussuunnitelma

Analyysivaiheen lopputuotoksesta tuotettiin sisäinen kehittämisraportti, mikä hyväksyttiin kehitysprojektiin osallistuvien toimesta tarvittavissa päätöselimissä. ”Analysointiprosessissa on kolme osaa: tietojen keräystavan tarkastelu, tietojen esitystavan tarkastelu ja käyttäjien ja johdon haastattelujen tekeminen” (Hernandez 2000, 102).

Analyysimenetelmiksi kutsutaan tapoja ja menettelyitä, joiden avulla tutkija jäsentää suurta määrää aineistoa rakenteisempaan ja helpommin tulkittavaan muotoon (Bister, T. 2019, 43). Käytännön toteutuksen lopputuokset analyysiin perustuen ovat esitelty seuraavissa kappaleissa 4.7–4.13.

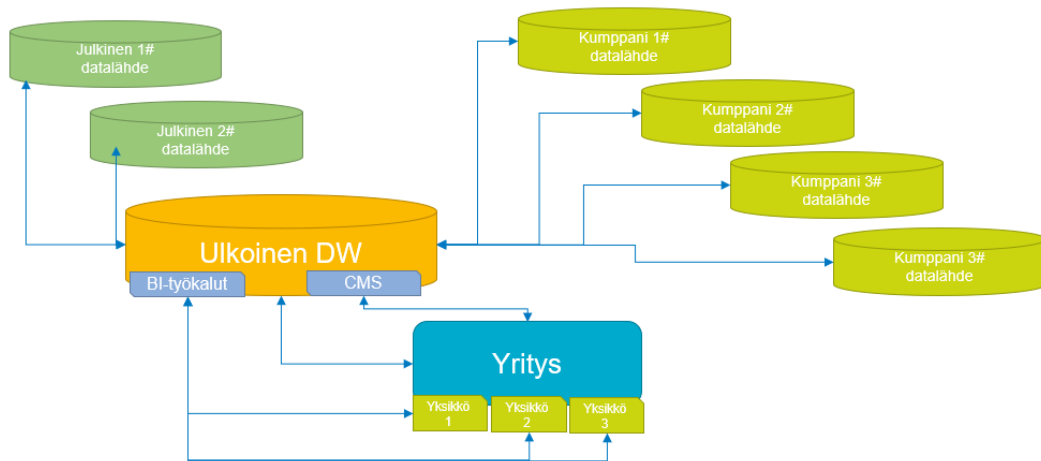
#### **4.7 Nykytilan tunnistaminen**

Nykytilaan liittyvässä selvityksessä tunnistettiin haastatteluihin ja projektipalaveriin liittyen ulkoistetun tietovaraston aiheuttavan lukuisia haasteita, jotka eivät tue yrityksen liiketoimintastrategiaa ja prosesseja optimaalisesti. ”*Tieto on jakautunut useampaan paikkaan*

ja se tieto, mitä meillä on haettavissa raporttien muodossa, ei ole ollut riittävää. Tiettyjä raporteja ei ole välttämättä ollut aina saatavilla ajallaan.”

Keskeisimmät puutteet liittyvät järjestelmäteknisiin rajoituksiin, rajoitettuun datan määrään sekä lähteisiin sekä yleisesti ulkoisen tietovarastoratkaisun muutoshallinta- ja ylläpitokustannuksiin. ”Toimittajien raporttien toimitushaasteet, tietojen koonti eri lähteistä tuo haasteita.”

Yrityksen ulkoistettu tietovaraston arkkitehtuuri ennen kehittämisprojektiä on kuvattu tarkemmin kuvassa neljä (kuva 4).



Kuva 4. Yrityksen ulkoistettu tietovaraston arkkitehtuuri ennen kehitystehtävää

Ulkoistettu tietovarastointiratkaisu aiheutti yritykselle seuraavia keskeisiä rajoitteita:

- Ulkoistettuun tietovarastoon on talletettu rajoitetusti yrityksen tarvitsemia datavaroja
- Ulkoistetun tietovaraston historiointitiedot kertyneestä datasta puutteelliset
- Ulkoistettu tietovarasto on vain rajoitetusti yhteydessä yrityksen käyttämien lähdejärjestelmien kanssa, eikä näin varastoi kertyvää dataa
- Ulkoistettuun tietovarastoon tehtävät muutokset kestävät kauan ja muutoshallinnasta aiheutuvia kustannuksia on vaikea ennakoida
- Ulkoistetun tietovaraston BI-käyttöliittymä palvelut ovat vain rajoitetusti yrityksen käytössä, eivätkä vastaa yrityksen eri organisaatiotasojen vaatimuksia käytettävyydestä, eikä tarvittavasta tietosisällöstä
- Ulkoistettuun tietovarastoon ei voi yhdistää yrityksen tavoitteiden mukaisesti modernin pilvipalveluympäristön lisäpalveluita, kuten analytiikan tukipalveluita

#### 4.8 Pilvipalvelu ja SNOW-Flake tietovarastonhallintaratkaisu

Selvityksessä esiin nousseet hyödyt, mitkä tukevat rakennettavan Snowflake-tietovarastointi ratkaisun rakentamista Amazon Webservicen tarjoamaan pilvialustaan. *”Pilvipalvelun kustannukset ovat alhaiset ja joustavuus lisääntyy”*.

Salon mukaan (2012, 175) pilvipalvelut korvaavat mm. tiedosto- ja sovelluspalvelimet, antavat mahdollisuuden käyttää omia räätälöityjä sovelluksia pienin kustannuksin ja lisää, että kiinteät kustannukset lähes häviävät, toiminta tehostuu, hallittavuus paranee, liikkuvuus kasvaa, prosessit nopeutuvat, henkilöstö on tyytyväisempää, kuten myös asiakkaat, ja ennen kaikkea liiketoiminnan kannattavuus kohenee. *”Helpottaa asioita kaikilla osa-alueilla. Sekä työnteon kannalta, että sen kannalta miten voimme tarjoamiamme palveluita kehittää kannattavammiksi meidän kannaltamme, sekä asiakkaiden kannalta paremmiksi.”*

Yleisesti pilvipalveluiden selkeitä etuja on, että palveluiden käytön aloittaminen ei edellytä merkittäviä alkuinvestointeja, kustannukset syntyvät ja perustuvat palveluiden sekä tallennuskapasiteetin käytön mukaan. Salon mukaan (2014, 94) tarjotut palvelut skaalautuvat joustavasti ja nopeasti ylös tai alaspäin. Kapasiteettirajoitetta ei asiakkaan näkökulmasta ole usein lainkaan. *”Toimittajariippuvuus vähenee, kustannushyödyt.”*

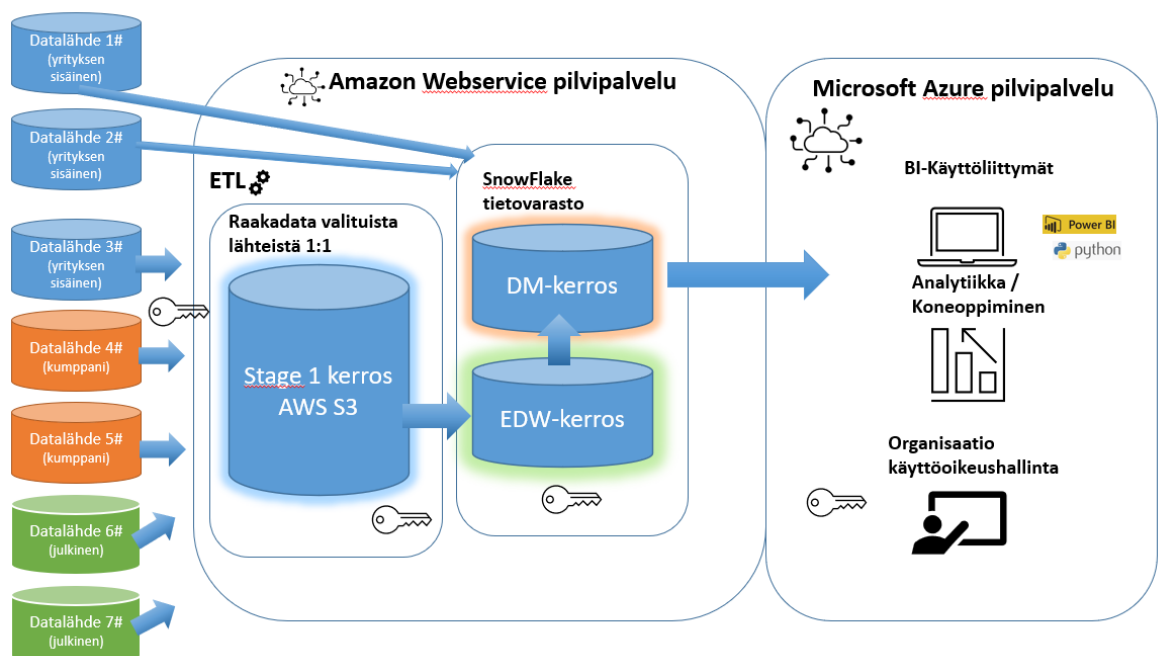
Pilvipalvelut tarjoavat mahdollisuuden kilpailuedun saavuttamiseen ja lähitulevaisuudessa yritykset, jotka eivät niitä käytä, menettävät kilpailukykynsä. ICT-monimutkaistuu ja monipuolistuu, ja yhä harvemmalla organisaatiolla on osaamisensa tai resurssiansa puolesta taloudellisesti järkevää tai edes mahdollista hallita tätä kokonaisuutta omin voimin. (Salo 2012, 174).

Pilvipalvelut tarjoavat eri toimialoille niiden tarvitseman tietoturvasuojan modernein datan suojausmenetelmin ja pilvipalvelualustalla toimiva Snowflake-tietovarastoinnin hallintaratkaisu sekä käyttöliittymä ovat aktivoitavissa käytännössä jopa suoraan verkkoselaimen kautta valituin lisäpalveluin mille tahansa yritykselle toimialasta riippumatta kuvan viisi ja kuusi mukaisesti (kuva 5 & 6). *”Pilvipalvelut tarjoavat erityisesti edullista tallennuskapasiteettia jatkuvasti kasvaville tietomäärille. Pilvipalveluun yritys voi automatisoida eri lähteistä määritetyn raakadatan käsittelyä sekä datamassojen historiointia, automaatioprosessit vapauttavat yrityksen omia resursseja tietovarastoinnin teknisistä ylläpitotoimista.”*

Table Name	Schema	Creation Time	Owner	Rows	Size
MY_TABLE	PUBLIC	8/23/18 8:52:37 PM	SYSADMIN	28	1.5KB

Kuva 5. "Snowflake" tietovarastoinnin hallintajärjestelmä (snowflake.com)

"Pilvipalvelut tukevat yrityksen vaatimia laskenta- ja analytiikkapalveluita sekä koneoppimista edistyneesti liiketoimintatarpeiden mukaisesti. Pilvipalveluiden tallennus ja laskentakapasiteettia on mahdollista nostaa automaattisesti yrityksen muuttuvien tarpeiden mukaisesti esimerkiksi kausivaihtelun takia. Pilvipalveluiden kyky kerätä, käsitellä ja jalostaa eri lähteistä kerättyä raakadataa liiketoiminnan hyödyksi on edistyneempää kuin perinteisten tietokantaratkaisujen kyvykkyys".



Kuva 6. Uudistetun tietovarastointialustan keskeiset elementit pilvipalvelussa

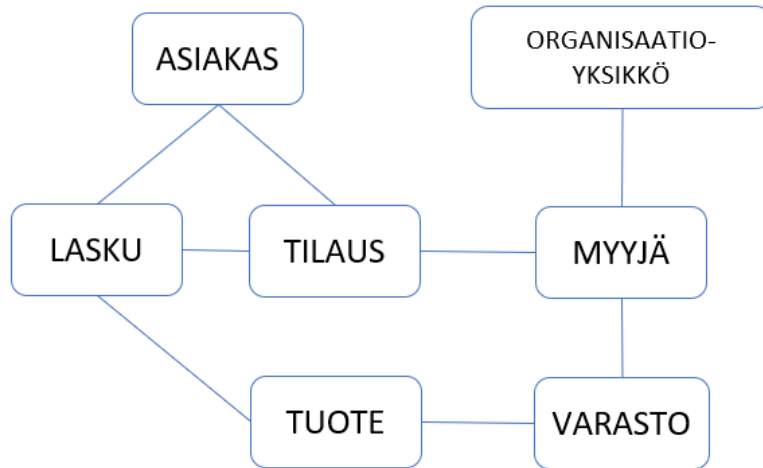
Salo huomauttaa (2014, 22–23), että perinteisesti yrityksen itsehallitsemassa palvelinkeskuksessa jopa 70 % IT-kustannuksista menee tutkimuksen mukaan olemassa olevan inf-

rastruktuurin ylläpitämiseksi. IAAS-ratkaisussa yritys ostaa palveluntarjoajan laitteiston resurssit käyttöönsä palveluna ja pitkäkestoisen sitoutumisen sijaan kapasiteettia voi lisätä joustavasti tarpeen mukaan. ”*Tieto on itsellä eli yksiköllä ja omalla alustalla, eikä missään muualla hajautettuna, sekä tiedon analysointi helpottuu projektin myötä. Tieto on valtaa ja siihen tuleekin panostaa.*”

#### **4.9 Yhdenmukaisen käsitteistön ja tarvemäärityksen tärkeys tietovarastointiuudistuksessa**

Kerätään vain sellaista dataa sisäisistä ja ulkoisista lähteistä, joilla on merkitystä yrityksen liiketoiminnalle. Ei kerätä lopulliseen tietovarastoon mitään sellaista tietoa, millä ei ole todellista käyttötarkoitusta. Tietovarastointi uudistuksessa yrityksen henkilöstön pitää ymmärtää ja määritellä yhteinen käsitteistö kerättävälle tiedolle, jonka merkityksen henkilöstö kykenee ymmärtämään sekä hyödyntämään. Käsitteistön määrittämisessä on tärkeää myös kuvata eri käsitteiden väliset loogiset suhteet. ”*Saamme tuotettua tuotteistamme dataa jatkossa myös eri yksiköille, eikä meidän tarvitse toimittaa lukuja heille erikseen, on tämä ainakin yksi helpottava tekijä myynnin kannalta. Eri yksiköt voi seurata omaa myyntiään tuotteidemme osalta ja tehdä tarvittavia liikkeitä myös itsenäisesti, esimerkiksi asettaa myyntitavoitteita*”.

Käsitteanalyysi on yleensä ryhmytystä, jossa liiketoiminnan edustajat ja IT-ammattilaiset yhdessä mallintavat tulevan järjestelmän kohdealuetta. Ylätason käsittemalli on kuvattu kuvassa seitsemän (kuva 7), mikä voidaan laatia seuraavilla säännöillä: mukaan otetaan vain ydinkäsitteet (kuten asiakas, tilaus, tuote jne.), joilla on yleensä selkeä perusavain. Pois jätetään vähäisen merkityksen omaavat riippuvat käsitteet, jotka kuvaavat vain yhtä toista käsitettä. (Hovi ym. 2005, 68–72).



Kuva 7. Esimerkki ylätasen käsittemallista (mukaillen Hovi ym. 2005, 72)

Vären mukaan (2019, 23) master dataan liittyy kaksi pääajatusta, se on liiketoiminnalle kriittistä tietoa, jota ilman organisaatio ei voi toimia. Ja lisäksi, että master dataa on jaettu läpi koko organisaation, jolloin kaikki eri toiminnot käyttävät tätä samaa dataa.

Hovin ym. mukaan (2009, 74) Käyttäjien on usein vaikeaa määrittellä, mitä tietoja he tarvitsevat päätöksentekonsa tai ylipäätään työnsä tueksi. Määrittelyä voi helpottaa rakentamalla tietovarastoalueen tietoja sisältävä prototyyppi, jonka avulla voidaan näyttää tietoja ja näin havainnollistaa ja visualisoida tulevaa ratkaisua. Käyttäjien on helpompi ottaa kantaa, kun on jonkinlainen ratkaisu nähtävillä. Hyvä proto tukee myös tietovaraston suunnittelua. *”Tapahtumien statistiikan tuottamiseen selkeä hyöty, kun tietoja voidaan poimia laajasti ja nopeasti esimerkiksi tapahtuma tyyppiin, tuotteen, toteutustavan tai sijainnin mukaan. Yksittäiset toimeksiannot hoituvat käden käänteessä. Parempi näkyvyys.”*

#### 4.10 Tietovaraston mallintamisen arkkitehtuurinen prosessi

Tietovarastoon ladataan määritetyistä lähdejärjestelmistä (sisäiset ja ulkoiset lähdejärjestelmät) kaikki tarvittava data. Tietovarastoon vietävä data käsitellään lopulliseen käyttömuotoon eri muokkauskerrosten avulla, jolloin jokaisessa latauskerroksessa on määrätty tehtävät, jotka edesauttavat tiedon lopullista käyttötarkoitusta liiketoimintahyötyjen saamiseksi. *”Keskitetty tietovarasto helpottaa, eikä tieto ole hajautettuna moneen paikkaan. Helpottaa kehitystä ja koordinoitua sillä nyt on vain yksi kontaktipiste.”*



Tietovarastokerrosten tunnistetut käyttötarkoitukset:

1. Stage 1-kerros, johon raakadata ladataan eri lähdejärjestelmästä kokonaisuudessaan suhteella 1:1 jatkojalostusta varten
2. EDW- kerros (enterprise dataware), jossa valittu raakadata historioidaan ja mallinnetaan yhteneväisten tietovarastointisääntöjen mukaisesti relaatiotauluihin.
3. DM-kerros (data mart), jossa määrättyjä relaatiotauluja voidaan käyttää halutuin suodatus- ja historiatiedoin. DM-kerroksessa tietovaraston taulut on jaettu ns. facta- ja dimensiotauluihin raportoinnin tarpeiden mukaisesti (facta- taulut=lukuarvoja, dimensiotaulut=luokittelutekijöitä ominaisuudesta).

#### 4.11 Raakadatan lataaminen

Raakadatan latauskerrosta (Stage-kerros) kutsutaan myös tietoallas kerrokseksi. ”Tietoallas-kerroksen etu yritykselle on, että tietoa voidaan ladata suuria määriä jatkojalostamista varten riippumatta tietomuodosta tai tietojen lataamista jo ennalta määritetyin muodoin tietoallas kerroksesta tietovarastoon.” Raakadatarakerros (Hendersson, Earley & Sebastian-Coleman 2017, 391) on välivarasto alkuperäisestä lähteestä kerätyille datalle, jossa se voidaan muuntaa ja valmistella ladattavaksi tietovarastoon. ”*Uskoisin tällä olevan ihan rahallista säästöä, kun emme ole jatkossa, toimittajiemme varassa raporttien kanssa niin ajallisesti kuin rahallisesti, kun saamme tiedot meille välitettävissä aineistoissa.*”

ETL tulee sanoista Extract – Transform – Load eli poiminta, muokkaus ja lataus. ETL-prosessissa tiedot luetaan operatiivisista järjestelmistä tai siirtotiedostoista, muokataan tietovarastotietokannan muotoon ja lopuksi kirjoitetaan tietovarastoon. Muokkauksessa on monia osatehtäviä, kuten virheiden tarkistus, koodien muunnokset ja tietojen historiointi. (Hovi ym. 2009, 48).

”*ETL-työkalua hyödyntäen siis raakadatatista hyödynnetään yrityksen tarvitsema data-osuus sekä muunnetaan ladattavaksi yrityksen varsinaiseen tietovarastokerrokseen määritetyn tietomuodon ja taulurakenteiden mukaisesti. ETL-työkalu tunnistaa määrittelyn mukaisesti uudet ladattavat aineistot tietoaltaasta ja muuntaa tiedon tarvittavaan muotoon, jotta aineisto voidaan ladata varsinaiseen tietovarastoon.*” Hovin ym. (2009, 25) mukaan tietovaraston latauksessa tarvitaan työtilaa, josta englannin kielessä käytetään termiä Staging Area ja työalueen tarkoituksena on avustaa ETL-prosesseja.

”*Eri lähdejärjestelmistä kerätyt tiedot on kuvattu selkokieლისinä sekä tietojen väliset suhteet on dokumentoitu tietovarastonhallintajärjestelmään kokonaishallinnan näkökulmasta. Työkalut ovat helpokäyttöisiä ja että meillä on kaikki tarvittava data yhdessä paikassa.*”

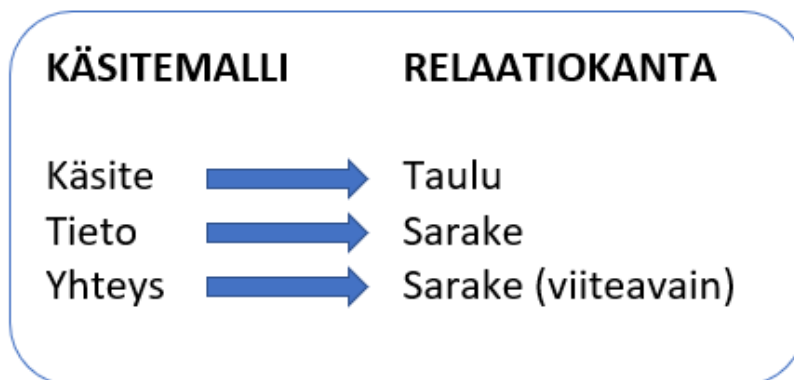
Tietovarasto-ympäristöön on tarjolla useita latausvälineitä (ETL-tools). Niihin voidaan yleensä tallentaa tietovaraston lataukseen liittyvää, teknistä metatietoa. Tietovarastoihin tarjolla olevissa kysely- ja raporttivälineissä on monesti oma metatietojärjestelmänsä (Hovi ym. 2005, 28).

#### 4.12 EDW-kerroksen taulumääriytykset

EDW-tietovarastointi kerros sisältää kaikki ne tiedot, joita yritys on määritellyt keräävänsä liiketoiminnan edistämiseksi. Tämä kerros sisältää yrityksen kaikkien yksiköiden tai organisaatiotasojen tarvitsemat tiedot valituista lähteistä, sisältäen myös historia ja muutostiedot. Hovin ym. (2009, 27) mukaan, EDW-kerroksessa tietoja hallitaan koko yritystasolla yli järjestelmä- ja organisaatorajojen ylitse.

*”Yrityksen EDW-tietovarasto sisältää lukuisista sisäisistä sekä ulkoisista lähteistä kerättyä tietoa, joten tietovaraston fyysisessä mallissa käytetyt dataan liittyvät termit, kuvaukset ja ylläpidon hallintaprosessit pitää olla dokumentoitu.”* Hovin ym. (2009, 27) mukaan työnjako on selvä, jolloin ICT-asiantuntijat rakentavat tietovaraston ja datamartti-aulut lataukseen ja loppukäyttäjät hyödyntävät tietoja selkeärakenteisten datamarttien tietoja BI-työkaluin.

Hovin ym. (2005, 104) mukaan, relaatiokantojen tietomalli on relaatiomalli, jossa objekteja on kahdenlaisia eli tauluja ja tietoja. Käsitellin yhteyksille ei ole suoraa omaa vastinetta relaatiomallissa, vaan niistäkin syntyy sarakkeita.



Kuva 8. Käsitellin relaatiokantaan (mukaillen Hovi ym. 2005, 104)

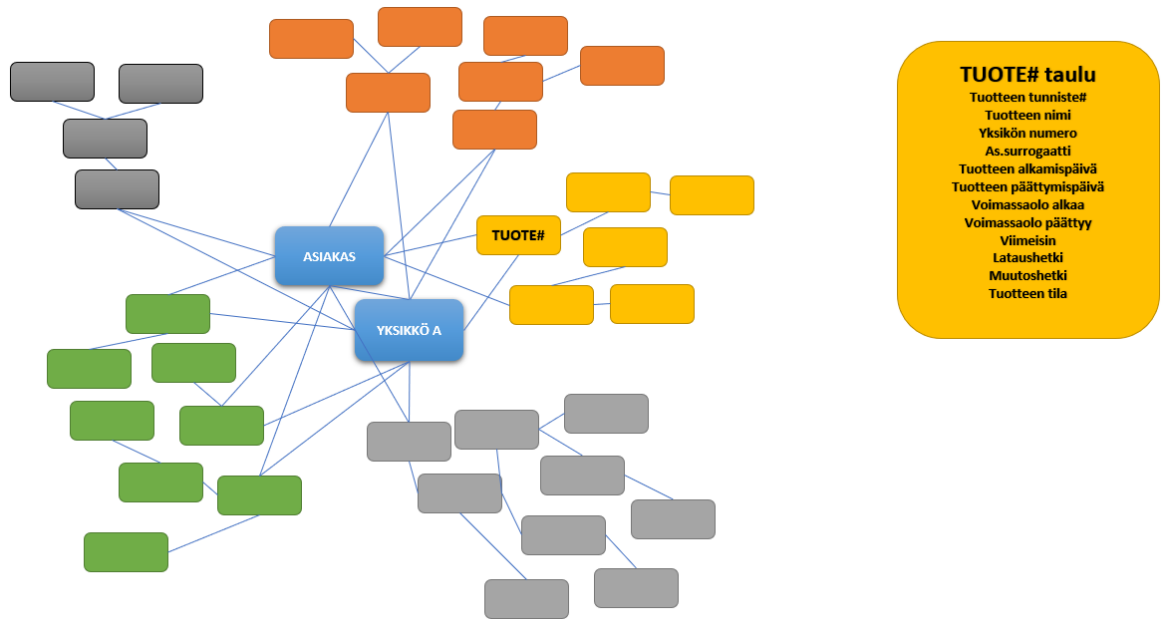
Kuvan kahdeksan (kuva 8) mukaisesti yhtenäisen käsitemallin jälkeen syntyville tauluille pitää määrittää hyväksyttävät tietotyypit, joita käytetään sisäisten ja ulkoisten lähdejärjestelmistä kerätylle datalle yhtenäisin määrittämissäännöin (taulukko 6).

Tietotyyppi	Selitys
char (n)	merkkimuotoinen, kiinteämittainen
varchar (n)	merkkimuotoinen, vaihtuvamittainen
decimal (n,m)	pakattu desimaaliluku, n=kokonaispituus, m=desimaalien määrä
number (n, m)	pakattu desimaaliluku
integer, int	suuri kokonaisluku (4 tavua)
smallint	pieni kokonaisluku (2 tavua), maksimiarvo n. 3 200
date, dateline	päivämäärä, aika

Taulukko 6. Taulun sarakkeen tietotyyppejä (mukaillen Hovi ym. 2005, 111)

*”Tietovarastoon luodaan tietotauluja merkityksellisistä asioista eli objekteista, jotka ovat liiketoiminnan edistämisen ja kehittämisen kannalta merkityksellisiä. Jokaiseen objektiin liittyy attribuutteja eli sarakkeita, joilla voidaan yksilöidä jokainen objektin rivi.”* Jokaisesta attribuutista kuvataan keskeiset tiedot ja lähdejärjestelmätiedot tietovaraston hallintajärjestelmään.

Käsittemallista taulurakenteeseen on tarkemmin kuvattu alla olevan kuvan mukaisesti (kuva 9).



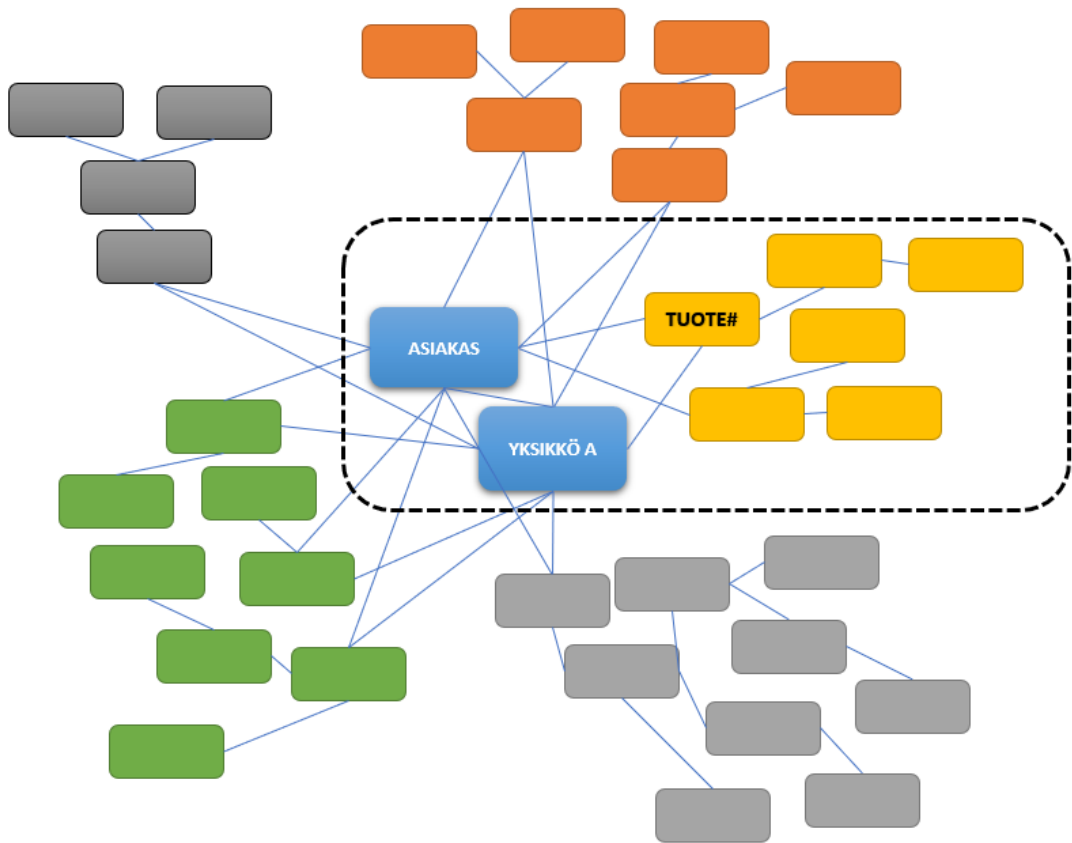
Kuva 9. Käsittemallista taulurakenteeseen

#### 4.13 Datamart-kerros raportoinnin ja analytiikan tukeen

Esityskerroksen (DM-kerros) tiedot ladataan ja muodostetaan EDW-kerroksen tietovarastosta. ”Esityskerrosta hyödynnetään valtuuksien mukaisesti eri liiketoimintayksiköissä BI-työkaluilla valmisraportin tai tekemällä raporteja itsenäisesti muuttuvien tarpeiden mukaisesti”. Datamart-kerros edesauttaa erilaisten liiketoiminnassa tapahtuvien trendien ja muutoksien vaikutusten tunnistamisessa. Microsoft Power BI:n visuaaliset näkymät yhdistettynä numeeriseen tietoon helpottavat datan antaman informaatioarvon ymmärtämistä. Lisäksi, sivun 33 kuvassa 10 on esitetty tarkemmin hyödynnettävissä olevat tietovaraston osajoukot esimerkiksi BI-työkalujen avulla, siten datamart-kerroksessa käytettävien taulujen avulla on helppo luoda tuotekohtaisia analyyseja tai valmistella kampanjoita valittuihin asiakaskohderyhmiin (kuva 10, 33).

”Analytiikan avulla voi tehdä faktaan perustuvia liiketoimintapäätöksiä, eikä tarvitse käyttää mutua.”

”Kannattavuuden seuranta ja ymmärryksen kasvaminen esim. asiakaskäyttäytymisen osalta, voidaan myös helpommin seurata tehtyjen toimenpiteiden vaikutusta liiketoimintaan.”

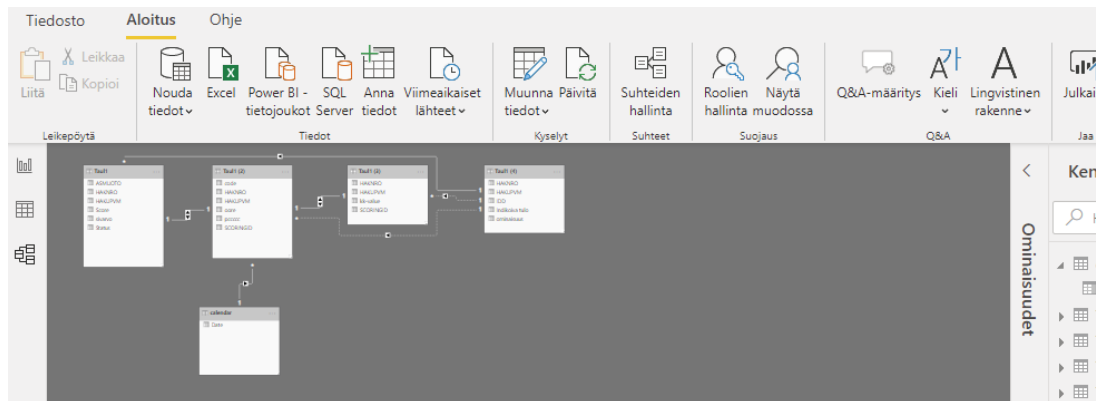


Kuva 10. Data marteissa hyödynnetään tietovaraston osajoukkoa esimerkiksi BI-työkaluin

Hovin ym. (2009, 24) mukaan datamarttien idea on tukea hyvin tietojen online-tyyppistä BI-käyttöä ja ne suunnitellaan ja toteutetaan yleensä jonkin käyttäjäryhmän tarpeiden mukaan: rakennetaan esimerkiksi henkilöstöhallinnan datamart, markkinoinnin datamart tai taloushallinnon datamart.

*”Eri yksiköille parempi näkyvyys omaan suorituskyykyyn ja kehityskohteisiin. Strategisten tavoitteiden mittaust.”*

Esimerkki datamarteista on esitetty kuvassa 11, josta saadaan tukea yksikkö- ja tuotekohtaisiin raportteihin sekä analytiikkaan



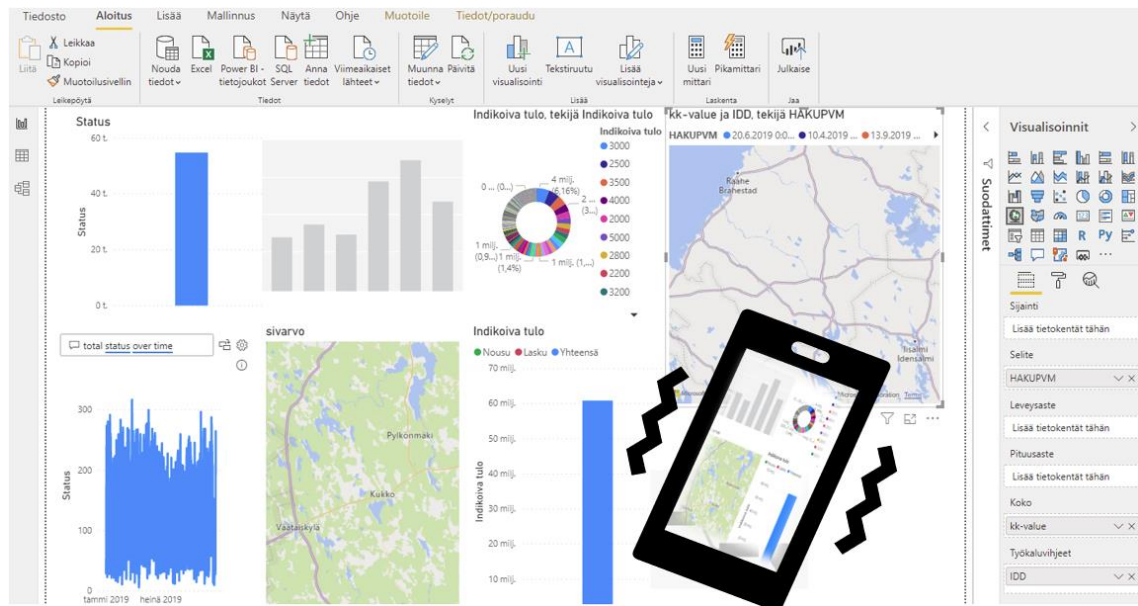
Kuva 11. Datamarteista tukea yksikkö- ja tuotekohtaisiin raportteihin sekä analytiikkaan

BI-työvälineillä voidaan helposti tuottaa haluttuja raportteja eri liiketoimintayksiköille valitun sisällön mukaisesti. Raportit voidaan tuottaa yrityksen intra-palveluun käytettäväksi tavanomaisen internetselaimen kautta hyödynnettäväksi. Raporttien visualisointi helpottaa tiedon ymmärrystä sekä parantaa raporttien yleistä käyttöastetta mielenkiintoa lisäten. ”Visualisointi helpottaa sitä, että ne säilyvät kiinnostavina ja parhaimmillaan erottuvat edukseen muiden lukuisien raporttien joukosta.” Raporttien käyttäjillä on mahdollista helposti ilman syvälistä raportointiosaamista suodattaa raporteista esimerkiksi haluamia aikanäkymiä tai valittuja tuoteominaisuuksien käyttöaste muutoksia. Suodattimet ovat helppokäyttöisiä sekä helposti ymmärrettäviä, sekä lisäävät informaatioarvoa.

Hovin ym. (2009, 104) mukaan päättäjät kaipaavat dataa, mikä on jo valmiiksi analysoitua sekä helppoa käsitellä. Interaktiivisuus toteutuu esimerkiksi siten, että käyttäjät voivat tarkastella haluamiaan tietoja tekemällä rajauksia sen perusteella, mitä tietoja haluavat tarkastella ja analysoida. ”...ehdottomasti hyödyntää lukujen tulkitsemisessä todella paljon, ja mitä olen tutustunut tähän mennessä, graafit ovat todella hienosti ja selkeästi toteutettu. Työtehtävät ovat jatkuvaa palveluiden kehittämistä. Analytiikan avulla pysymme paremmin mukana kehityksessä.”

Koponen & Hildén (2019, 29) esittävät, että visualisoinnin onnistuminen määräytyy viime kädessä sen mukaan, auttaako visuaalinen ilmentyvä ja sen muoto lukijaa ymmärtämään datan informaatioarvoa paremmin, vai onko se vain tarpeeton yksityiskohta. Visualisointi on silloin tarkoituksenmukaista, kun lukija ymmärtää heti mitä se esittää ja löytää sen

kautta vastauksia kysymyksiin tai hahmottamaan datan välittämää informaatiota helpommin.



Kuva 12. BI-työkaluin jalostetaan Datamarteista raportteja automaatioajoin tai manuaalisesti liiketoimintatarpeen mukaisesti eri käyttöliittymiin

Säännöllisten raporttien päivitykset voidaan automatisoida tietovaraston päivitettyjen aineistoajojen mukaisesti, mikä vähentää manuaalista työtä. Lisäksi BI-asiantuntijat kykenevät helposti toimittamaan räätälöityä raportteja tai tietosisältöjä erillisten satunnaisten tarpeiden mukaisesti. ”Pystyy nopeasti ja helposti tarkistamaan reaaliaikaista dataa ilman että minun tarvitsee penkoa sitä monesta paikkaa.”

Kirkkojen mielestä data auttaa niitä löytämään tehokkaammin uusia jäseniä. Kirkot haluavat tietää, kuka olet, ne haluavat ennustaa kapasiteettiasi antaa ja todennäköisyyttä siihen, että jätät ohjelman kesken – ne ovat ihan samoja konsepteja, jotka soveltuvat pankille. (Raivio 2022, 17).

## 5 Tulokset

### 5.1 Havainnot ja tulokset

Yrityksen tietovarastointi alustan uudistaminen sisäisten liiketoimintatarpeiden perusteella julkiseen pilvialustaan tuo merkittäviä kokonaishyötyjä. Yrityksen liiketoimintaprosesseissa voidaan käyttää aina tarkoituksenmukaista tietoa sisäisten tai sidosryhmien lähdejärjestelmistä. Liiketoiminta kykenee hyödyntämään automaation synnyttämiä analyysejä sekä heuristia päätöksenteossa, automaatio mahdollistaa myös loppuasiakkaille strategian mukaisia asiakaskokemuksia kustannustehokkaasti. Yritys maksaa käyttämästään pilvialustasta sekä liitetyistä teknologiaratkaisuista käytön mukaan, tämä tuo merkittäviä kustannussäästöjä sekä mahdollistaa joustavia muutostarkeita liiketoimintavaatimusten mukaisesti. Pilvialusta ratkaisut ja tiedonsiirto on aina salattu tietoturvan näkökulmasta toimialavaateet huomioiden. Pilvialustan ratkaisun ytimenä toimii Snowflake-tietoalusta, mikä on skaalautuva tietovarastointiratkaisu kerätyn datan hyödyntämiseksi.

Kehittämistehtävän tulokset, jotka mahdollistavat tietovarastointialustan uudistamisen:

- Tietovarastointikokonaisuuteen liittyvät kumppanuus ja sopimusvastuut ovat laadittu ja hyväksytty.
- Pilvipalvelualustan toiminnallisuudet ovat aktivoitu käyttöön (AWS).
- Tietovarastoinnin hallintajärjestelmä on implementoitu pilvipalveluun (Snowflake).
- Määritetyistä lähdejärjestelmistä sekä eri reskontraliittymistä voidaan ladata ja välittää tietoa uuteen tietovarastoalustaan (sisäiset ja ulkoiset lähdejärjestelmät).
- Tietovarastoalustaa voidaan hyödyntää yrityksen eri liiketoimintayksiköissä ja organisaatiotasolla raportointityökaluilla (Business Intelligence -työkalut).
- Raportointityökalut päivittävät automaattisesti määritellyjä liiketoiminnan tukiraportteja valituin lataushetkin, käyttäjäkohtaisia raportteja ja informaatiota voidaan tuottaa myös räätälöidysti eri tarpeisiin (esim. raportoinnin suodattimet)
- Raportointityökalut tuottavat visuaalisia raportteja käyttäjille helpottamaan johdon päätöksien tekemistä.
- Valittujen pilotti liiketoimintayksiköiden kanssa on määritelty liiketoimintaa tukevat valmisraporttien informaatio sisällöt ja graafiset näkymät. Käyttöliittymänä hyödynnetään Microsoftin Power BI -raportointityökalua.
- Tietovarastointialustaa hyödynnetään valituissa asiakasrajapinnan työkaluissa, jotka parantavat asiakaskokemusta ja henkilöstön ajankäyttöä sekä lisäävät analytiikan käyttömahdollisuuksia sekä tavoitteiden ja myynnin kehittymisen seuranta.



## 5.2 Tulosten analysointi

Kaksi ensimmäistä tutkimuskysymystä liittyivät liiketoimintaa tukevan tietovarastoinnin kehittämisenprosessiin sekä osa-alueiden johdonmukaiseen hallintaan uudistamisprosessin toteuttamiseksi myös järjestelmätekniikan vaatimusten näkökulmasta. Tutkimuksen aikana tunnistettiin, että tietovarastointiuudistus on monivaiheinen sekä laaja kehitysprojekti, minkä suuntaviivat määrittävät alkuvaiheessa erityisesti yrityksen liiketoimintatarpeet. Liiketoimintatarpeiden näkökulmasta toteutettu tietovarastoinnin uudistamisen projekti edellyttää liiketoiminnan edustajien lisäksi erityisesti tietoarkkitehtuurin ja tietovarastointiammattilaisten syvällistä asiantuntemusta sekä käytännön kokemusta vastaavista projekteista. Koska käytännön toteuttaminen edellyttää lukuisia teknologisia ja kokonaisarkkitehtuuriin liittyviä päätöksiä siten, että yrityksen kokonaisarkkitehtuurin tarkoituksenmukaiset toiminnallisuudet eivät vaarannu. Toisaalta keskitetyn tietovarastointiratkaisun tuottaminen pilvipalveluun myös vapauttaa sekä vähentää yrityksen sisäisiä resurssitarpeita tietovaraston hallintatehtäviin liittyen, koska valitut ylläpito ja hallintapalvelut samalla ulkoistetaan sopimuksien mukaisesti.

Kolmas tutkimuskysymys liittyi tietovarastosta saatavien datan välittämiseksi liiketoiminnan hyödyksi eri organisaatiotasolle? Keskiössä oli tunnistaa ne keskeiset ja merkitykselliset tiedot, mitä liiketoiminta tarvitsee päätöksentekoon sekä analytiikkaan. Tässä roolit olivat sikäli selkeitä, että liiketoiminnan edustaja tunnistaa tarvitsemat tiedot sekä osaa arvioida oikeita tietolähteitä tiedonkeräämiselle. Ja tietovarastoinnin sekä BI-työkaluja hallitsevien rooli on välittää tarvittavat tiedot oikeista lähdejärjestelmistä tietovarastoon sekä jalostaa edelleen tietoa haluttuun visuaaliseen muotoon BI-työkaluja hyödyntäen. Lisäksi keskeistä oli järjestää riittävä koulutus läpi organisaation, jotta liiketoiminnan edustajat osaavat käyttää ja hyödyntää valituin käyttöliittymin kerättyä tietoa.

## 5.3 Tutkimustulosten luotettavuus

Mäkisen mukaan (2006, 92) tutkimuksissa käytetään paljon haastatteluja, että kyselyjä. Ne pitävät sisällään monia eettisiä ongelmia. Miten varmistetaan se, että vastaukset ovat luotettavia? Miten koehenkilöt tai haastateltavat pitää huomioida?

Haastatteluja ja tietoa kerätessäni kehittämistehtävää varten, kerroin avoimesti miten saatuja tietoja hyödynnetään tutkimuksessa. Ja erityisenä kriteerinä oli varmistaa riittävä anonymiteetti kerättyjen aineiston käytössä. Tutkimustehtävää helpotti kehittämistehtävän

keskeisten osa-alueiden riittävä tunteminen, jolloin haastatteluissa oli mahdollista keskittyä olennaisten aihealueiden käsittelyyn. Haastatteluja en tietoisesti tallentanut vaan vastustiedot kirjoitettiin käsin muistioksi.

Mäkinen lisää (2006, 92), että monet tieteen sisäiset haastatteluihin liittyvät eettiset kysymykset koskevat reliabiliteettia ja validiteettia. Jos kysymyslomakkeet on laadittu niin, että ne eivät anna kysymyksenasetteluun luotettavia vastauksia, tutkija tuhlaa sekä omia että laitoksensa resursseja ja käyttää samalla vastaajien aikaa hyödyttömään toimintaan.

Liiketoiminnan edustajille välitetyissä kyselomakkeissa vastausvaihtoehdot olivat tietoisesti jätetty avoimeksi, että minkäänlaista johdattelua ei tapahdu esimerkiksi monivalintavastauksista aiheutuen. Tästä oli myös merkittävää etua, koska avointen vastausmahdollisuuksien kautta sai laajemman käsityksen liiketoiminnan edustajien näkökulmista tavoitte-tilaan liittyen.

Kehittämishankkeessa huomioidaan tietosuojalaki, laki luottolaitostoiminnasta, luottolaitoslaki sekä kuluttajansuojalakiin liittyvät tietosuojaan liittyvät vaatimukset. Kehitysprojektiin liittyvien yrityksen ulkopuolisten osapuolien kanssa solmitaan sopimukset, joissa vastuut, vaatimukset sekä tuotantovaiheen ylläpito tehtävät ovat määriteltä.

#### **5.4 Kehittämishankkeen eettisyys ja tietosuoja**

Kehittämishankkeessa huomioidaan tietosuojalaki, laki luottolaitostoiminnasta, luottolaitoslaki sekä kuluttajansuojalakiin liittyvät tietosuojaan liittyvät vaatimukset. Kehitysprojektiin liittyvien yrityksen ulkopuolisten osapuolien kanssa solmitaan sopimukset, joissa vastuut, vaatimukset sekä tuotantovaiheen ylläpito tehtävät ovat määriteltä.

Rekisterinpitäjän osoitusvelvollisuus. Tietosuoja-asetuksen mukaan rekisterinpitäjä on se, joka päättää henkilötietojen käsittelyn käyttötarkoituksista ja keinoista. Rekisterinpitäjä vastaa siitä, että organisaatio noudattaa tietosuoja-asetuksen vaatimuksia koko yrityksen tasolla. Datan hallinnassa tämän haasteen ratkaisee jalkautettu datan hallintamalli ja sen kautta selkeät roolit ja vastuut datan käsittelylle. Näiden roolien ja vastuiden kautta rekisterinpitäjä tietää ja hallitsee henkilötietojen käsittelyn prosessien dokumentointia ja sitä, että prosesseja noudatetaan. Osoitusvelvollisuuteen liittyy myös datan käsittelyn prosessien määrittely. Osana datanhallintaa on määritellyt ja dokumentoidut prosessit datan luomiseksi, ylläpitämiseksi ja poistamiseksi. Näillä prosesseilla voidaan vastata tietosuoja-asetuksen vaatimukseen siitä, että data on ajanmukaista, oikein ja että tarpeeton tai virheellinen data poistetaan viipymättä. Hyvin kuvattu datan elinkaari auttaa tunnistamaan

mikä data on milloinkin tarpeetonta, jotta se voidaan poistaa vaatimuksenmukaisesti. (Väre 2019, 51).

EU:n tietosuoja-asetuksen eli ns. GDPR (General Data Protection Regulation) asettaa henkilötietojen käsittelylle ja tietosuojalle vaatimuksia myös tietovarastointikäsitteilyn periaatteissa. Tietovarastoinnin kehittämishankkeessa on jo mallinnusvaiheessa huomioitu, että esimerkiksi henkilötunnusta ei käytetä tietokannan teknisenä avaimena tietojen yhdistämiseksi muuhun kokonaistietoon, vaan henkilötiedosta on muodostettu erilliset keinotekoiset avaimet henkilötiedon minimointiperiaatteiden varmistamiseksi. Lisäksi henkilöstön käyttöliittymissä tai automatiikan säännöllisesti tuottamissa raporteissa henkilöä identifiivaa tietoa ei tuoda esiin (esim. asiakkaan nimi, kotiosoite, puhelinnumero). Henkilötietoja voi käyttää vain erillisen käyttöoikeuden omaavat henkilöt käyttötarkoitussidonnaisuuden perusteella huomioiden tietosuoja-asetuksen määräykset. Kun tietovarastoa hyödynnetään esimerkiksi asiakaskampanjointiin liittyvän tiedon keräämiseksi, ei kampanjointia kohdisteta asiakkaisiin, jotka eivät ole antaneet suostumusta suoramarkkinointiin tai määrättyjen palvelukanavien käyttämiseksi viestinnässä.

Oikeus saada pääsy tietoihinsa tai tulla unohdetuksi. Osana master datan hallintaa tehtävät tietovirrat, tietomallit sekä datan poistamisen prosessi ovat myös tärkeitä työkaluja tietosuoja-asetuksen täyttämiseksi. Ne edistävät organisaatiota tunnistamaan, mitä dataa on missäkin järjestelmässä ja mihin kaikkeen sitä käytetään. Niiden avulla voidaan määritellä, kuuluuko tieto ylipäänsä tarkastusoikeuden tai pyynnöstä poistettavien piiriin. Osoitusvelvollisuus edellyttää organisaation ulkopuolisten lähteiden tunnistamista ja keille tietoa luovutetaan käytettäväksi. Nämä edellä mainitut täytyy huomioida, kun henkilö pyytää tietojansa nähtäville tai haluaa ne poistettavan. Tietovirtojen kuvaaminen auttaa myös tässä. (Väre 2019, 51–52).

## **6 Johtopäätökset ja jatkokehitysmahdollisuudet**

### **6.1 Johtopäätökset**

Nykyisessä työssäni tekijä on voinut osallistua tämän projektin tavoitetilan asetantaan oman liiketoiminta-alueeni näkökulmasta, kun varsinainen kehitysprojekti vielä odotti käynnistymistään. Työtehtävissä tekijä on voinut käyttää laaja-alaisesti ulkoistetun palveluntarjoajan tarjoamaa tietovarastoa sekä raporttien tuottamiseksi tarkoitettuja BI-käyttöliittymiä. Käytännössä tämä on tarkoittanut eri raporttien tai analyysien tuottamista eri liiketoimintayksiköiden tarpeisiin. Tyypilliset raportit ovat sisältäneet määrättyjen asioiden historiallista kehittymiseen perustuvaa havainnointia ja siitä tehtävää analyysiä johtopäätöksiin. Raporttien käyttäjät ovat voineet arvioida oman yksikön kehittymistä eri liiketoiminta-alueilta verraten muiden yksiköiden toimintaan.

Muut ulkoistetun tietovarastoinnin käyttökohteet ovat kohdistuneet erilaisten kampanjakohteryhmien luomiseen, jossa tunnistettu ja ennakoitu määriteltyjen ryhmien tarpeita sekä niiden täyttämistä tarjotuilla palveluilla. Ulkoistetun palveluntarjoajan raportointi- ja tietovarastointiratkaisut ovat täyttäneet pääasiallisesti liiketoiminnan tarpeet. Kuitenkin keskeisiä puutteita on voinut tunnistaa, jotka liittyvät käytettävissä olevan datan määrän puutteellisuuteen niin sisällön kuin käytettävissä olevan historiointitietojen osalta. Lisäksi käytettävä data ei ole ollut reaaliaikaista ja tietovaraston päivitysajot ovat olleet hitaita ajallisesti. Uusi tietovarastointialusta yhdistettynä nykyaikaisiin BI-työkaluihin helpottaa erityisesti raportoinnin automatisointia, tiedon helpompaa omaksumista tiedon visualisoinnin kautta läpi organisaation. Ylätasolla yrityksen kyky hyödyntää eri tietolähteitä sekä tehdä tarvittavia muutoksia tietovarastoinnin ratkaisuihin muutostilanteissa, helpottuu ja nopeutuu merkittävästi pilviympäristössä tapahtuvan tietovarastoalusta ratkaisun johdosta. Yritys kykenee käyttämään tarvittaessa tuoretta, lähes reaaliaikaista dataa. Syntyvien kustannustenhallinta on täysin yrityksensä itsensä hallinnassa. Tietovarastointialustan tietoja on mahdollista käyttää modernein analytiikkatyökaluin itsenäisesti (esim. Python & Power BI) sekä pilvialustan tarjoamin lisäpalveluratkaisuin.

### **6.2 Jatkokehitysmahdollisuudet**

Yrityksen strategia ohjaa liiketoimintaa asetettujen painopisteiden ja tavoitteiden näkökulmasta. Yrityksen ympäristössä tapahtuvia muutoksia strategia ei käytännön tasolla tunnista, strategiassa mahdollisia muutostarpeita käytännön tasolla ohjaa liiketoiminnasta ja

ympäristöstä saatava informaatio eli data. Datan perusteella analytiikan sekä koneoppimisen menetelmiä hyödyntäen voidaan ennustaa parhaiten yrityksen toimintaympäristöön sekä asiakaskäyttäytymiseen tapahtuvia muutoksia. Kun data ohjaa asiakasliiketoimintaa, voidaan ennakoida esimerkiksi, milloin, mitä, missä kanavassa ja millä viestinnällisellä kärjellä eri asiakaskohderyhmille voidaan tarjota oikein ajoitettuja ja laadukkaampia asiakaskokemuksia asiakastarve huomioiden. Mielikuva siitä, että autoa ajaessa nähdään etuikkunasta tulevat korkean todennäköisyyden tapahtumat ennakkoon ja kuljettajaa voidaan näistä informoida etukäteen, kuten parantuvista sääolosuhteista tai vaaraa aiheuttavista tiellä olevista kelirikoista.

Tekoälylaitteetkin oppivat, eivät kuitenkaan inhimillisesti vaan sen mukaan, mitä ihminen antaa sille eväitä koneellista toimintaa varten ohjelmoijien kirjoittamien algoritmien perusteella. Kun lapsen kokonaisvaltainen kasvu aikuiseksi mitataan noin kahdessa vuosikymmenessä, tekoälyn kykyjen kasvu yksittäisten tehtävien hallitsijaksi mitataan vaikeustasosta riippuen minuuteissa tai viikoissa, joskus vuosissa. (Siukonen & Neittaanmäki 2019, 31).

Koneoppimisella voidaan automatisoida ja nopeuttaa asiakaspalvelun prosesseja, jotka perustuvat määriteltyihin herätteisiin. Koneoppiminen lisää yrityksen kustannustehokkuutta sekä tehostaa operatiivisia prosesseja ja vähentää manuaaliyötä. Asiakaspalvelua voidaan tarjota ympärivuorokauden hyödyntäen robotiikkaa, esimerkiksi ensivaiheen selvitysratkaisujen toteuttaminen ja palvelun ohjaaminen robottibotteja hyödyntäen.

Organisaation ei tule pitää datatieteen projektia kertaluonteisena asiana. Kerätäkseen pitkän aikavälin edut organisaation on rakennettava kapasiteettia toteuttaa datatieteen projektit usein ja käyttää projektin tuloksia. Datatieteen pitäminen strategiana edellyttää korkeimman johdon pitkän aikavälin sitoutumista. (Kelleher & Tierney, 2021, 222).

Tietovarastoinnin uudistaminen pilvipalveluun tukee yrityksen digitaalisten pankkipalveluiden kehittämistä, jossa eri lähteistä kerättyyn tietoon ja niiden yhdistelyyn perustuen kasvatetaan eri liiketoiminta-alueiden automaatiota ja tekoälyn käyttömahdollisuuksia. Yrityksen henkilöstö kykenee seuraamaan liiketoiminnan kehittymistä ja trendejä lähes reaaliaikaisesti kerättyyn dataan perustuen. Kustannustehokkuus paranee, koska laskutus perustuu ensisijaisesti käytön perusteella tapahtuvaan laskutukseen. Palveluiden kapasiteettia ja valikoimaa on mahdollista kasvattaa ilman raskaita muutosprojekteja. Tietovarastoinnin ratkaisu pilvipalvelussa tukee luontevasti yrityksen alusta-arkkitehtuurin toiminnallisuuksia,

koska merkittävä osa järjestelmien ekosysteemeistä käyttää jo verkko- tai pilvialustapalveluissa tapahtuvia prosesseja. Tietovarastoinnin uudistaminen mahdollistaa aidosti siis sen, että kerätty ja analysoitu data ohjaa liiketoimintaa yrityksen valitun strategian mukaisesti.

## Lähteet

Collin, J., Saarelainen, A. 2016. Teollinen internet. Alma Talent Oy. Helsinki. E-kirja. Luettu: 1.1.2022.

Hendersson, D., Earley, S. & Sebastian-Coleman, L. 2017. Data Management body of knowledge. 2 painos. Technics Publications. Basking Ridge. Yhdysvallat.

Hernandez, J. 2000. Tietokannat suunnittelu käytännössä. Gummerrus Kirjapaino Oy. Jyväskylä.

Ari Hovi. 2020. Katsottavissa: <https://www.arihovi.com/materiaalit/datapedia-data-alan-termit-avattuna/>. Luettu 20.2.2022.

Hovi, A. 2010. SQL-opas. 9. uudistettu painos. Docendo Finland Oy. Jyväskylä.

Hovi, A., Hervonen, H., Koistinen, H. 2009. Tietovarastot ja business intelligence. WS Bookwell. Porvoo.

Hovi, A., Huotari, J., Lahdenmäki, T. 2005. Tietokantojen suunnittelu & indeksointi. WS Bookwell. Porvoo.

Kelleher, J., Tierney, B., 2021. Datatiede. Painoliber Oy. Helsinki.

Koponen, J., Hildén, J., 2019. The data visualization handbook. Aalto ARTS Books. Espoo.

Lehtonen, P., Lindblom, L., Korpinen, S., Simonen, J., 2006. Projektisalkunhallinta – Kehitystoiminnan strateginen johtaminen. Edita Prima Oy. Helsinki.

Merilehto, A. 2018. Tekoäly matkaopas johtajalle. 3. uudistettu painos. Alma Talent Oy. Helsinki.

Mäkinen, T. 2006. Tutkimusetiikan ABC. 2006. Kustannusosakeyhtiö Tammi. Helsinki.

Ojasalo, K., Moilanen, T., Ritalahti, J. 2015. Kehittämistyön menetelmät. 3–4. uudistettu painos. Sanoma Pro Oy. Helsinki. E-kirja. Luettu 5.1.2022.

Raivio, J. 2022. Massadataa hyödyntävät myös kirkot. Helsingin Sanomat 30.1.2022, s. 17.

Snowflake. Tietokannanhallintajärjestelmä. Katsottavissa: <https://docs.snowflake.com/en/other-resources.html#videos>. Luettu 13.2.2022.

Siukonen, T., Neittaanmäki, P. 2019. Mitä tulisi tietää tekoälystä. Docendo Oy. Jyväskylä.

Aki Taanila. 2012. Katsottavissa: <https://tilastoapu.wordpress.com/etusivu/>. Luettu 13.3.2022.

Tuomo Tamminen. 2019. Katsottavissa: <https://www.suomidigi.fi/sv/node/499>. Luettu 20.2.2022

Salo, I. 2012. Hyötyä pilvipalveluista. Saarijärven Offset Oy. Saarijärvi.

Salo, I. 2014. Big Data & pilvipalvelut. Docendo Oy. Jyväskylä.

Väre, T. 2019. Master data. Alma Talent Oy. Helsinki.



## Liitteet

Liite 1. Kyselylomake (lomakkeessa olevat kuvat ovat julkaisuvapaita Microsoft Form palvelusta).

**1**



The diagram illustrates the Data Warehouse and ETL process. It shows a flow from 'register' to 'collect' to 'analyze'. On the left, various data sources are listed: ERP, FIN, CRM, EXT, and others, categorized as 'relational'. These feed into a 'Staging Area (SA)' which is also 'relational'. From the SA, data is moved to a 'Central Warehouse (CWH)' which can be 'relational or dimensional'. Below the CWH is an 'Operational Data Store (ODS)'. On the right, 'BI tools Reporting tools' are shown, which access data from 'dimensional' sources (1, 2, ...) and 'cubes/datamarts'. The process is labeled 'Data warehouse and ETL'.

Tietovaraston uudistusprojektista aluksi, kuvaile lyhyesti miten tämä helpottaa työtäsi tulevaisuudessa tai mitä hyötyjä siitä saat?

Kirjoita vastaus

**2**

Mitä pidät itsellesi tärkeimpänä asiana työn edistämisen näkökulmasta kun uudistusprojekti valmistuu?

Kirjoita vastaus

**3**

Mitä hyötyjä näet uudistusprojektista liiketoimintayksiköllesi?

Kirjoita vastaus

4

Mitä hyötyjä näet uudistusprojektista asiakkaille?



Kirjoita vastaus

5

Mikä on arvokkain asia mielestäsi koko yrityksen näkökulmasta mitä tietovarastoalustan uudistusprojekti tarjoaa?



Kirjoita vastaus

6

Mitä haasteita raporttien ja liiketoimintatiedon hankinnassa on ollut ennen uudistusprojektiä?

Kirjoita vastaus

7



Mitä arvelet hyödyiksi kun uusi tietovarastoalusta upotetaan ns. pilvialustaan? Mitä hyötyjä tästä meille tai asiakkaille voisi arviosi mukaan olla?

Kirjoita vastaus

8

Miten analytiikka voisi auttaa työn tekemistäsi?

Kirjoita vastaus

9



Mitä raporttien visualisointi hyödyttää? Esimerkiksi Power-Bi käyttöliittymän graafit?

Kirjoita vastaus

10



Mitä raportointi tai liiketoimintatietotarpeita uudistusprojekti ei ratkaise työsi osalta? Listaa muutama esimerkki mitä mieleen tulee?

Kirjoita vastaus