

Sakari Laulajainen

DATA-ARKKITEHTUURI STRATEGISEN PÄÄTÖKSENTEON TUKENA

Kandidaatintyö
Tekniikan ja luonnontieteiden tiedekunta
12/2020

TIIVISTELMÄ

Sakari Laulajainen: Data-arkkitehtuuri strategisen päätöksenteon tukena
(Data architecture to support strategic decision making)
Kandidaatintyö
Tampereen yliopisto
Teknis-taloudellinen kandidaattiohjelma, tuotantotalous
12/2020

Datan määrän jatkuva kasvaminen luo yrityksille tarpeen hallita ja hyödyntää saatavilla olevaa tietoa liiketoiminnan kehittämiseksi. Toimiva data-arkkitehtuuri on yksi tärkeimmistä työkaluista, joiden avulla yritys pystyy tehokkaasti hyödyntämään tietoa päätöksenteon tukena. Data-arkkitehtuurin merkitys näkyy erityisesti strategisessa päätöksenteossa, jossa tietoa yleensä tarvitaan paljon ja useista eri lähteistä. Tässä työssä pyritään selvittämään yleisiä periaatteita ja prosesseja, joita yritykset voivat hyödyntää oman data-arkkitehtuurinsa kehityksessä.

Työssä käsitellään ensin tietoa ja sen merkitystä päätöksenteon tukena. Tiedon merkitystä painotetaan erityisesti strategisen päätöksenteon näkökulmasta, jossa tietoa yleensä hyödynnetään paljon laajemmin ja perusteellisemmin kuin operatiivisessa päätöksenteossa. Tiedon hyödyntämisen jälkeen työssä siirrytään eri tietotyyppien käsittelyyn. Tässä luvussa kuvataan minkälaista tietoa yrityksillä voi olla käytettävissään. Tietotyyppien kaksi eri pääluokkaa ovat hiljainen ja eksplisiittinen tieto. Eksplisiittinen tieto jaetaan edelleen strukturoiduksi ja strukturoimattomaksi tiedoksi. Tietotyyppien jälkeen kuvataan myös yrityksille tyypillisiä tietolähteitä, joista tietoa voidaan kerätä. Tietolähteet voidaan jakaa korkealla tasolla yrityksen sisäisiin ja ulkoisiin tietolähteisiin.

Tiedon merkityksen, tietotyyppien ja tietolähteiden käsittelyn jälkeen työssä siirrytään tarkastelemaan yksityiskohtaisemmin data-arkkitehtuuria. Aluksi kuvataan yleisiä periaatteita, joita data-arkkitehtuurin kehittämisessä tulisi aina ottaa huomioon. Näitä ovat esimerkiksi datan laadun varmistaminen, tiedon turvaaminen ja master datan hallinta. Tämän jälkeen kuvataan erityisesti strategisen päätöksenteon tukena toimiva keskitetty arkkitehtuurimalli. Keskitetyssä arkkitehtuurissa tiedonhallinta pyritään toteuttamaan mahdollisimman keskitetysti, jotta tietoa voidaan esimerkiksi paremmin yhdistellä ja hallita. Keskitetty arkkitehtuuri on myös toteutukseltaan selkeä ja helpommin hallittavissa ja jatkokehitettävissä.

Yleisten periaatteiden käsittelyn jälkeen työssä tarkastellaan data-arkkitehtuurin kehitysprosessia ja yhtä tyypillistä teknistä toteutusta keskitetyille data-arkkitehtuurille. Kehitysprosessin eri vaiheita ovat lähtötilanteen arviointi, tietotarpeiden ja tavoitteiden määrittely, käsitteellinen mallinnus, tekniset ratkaisut ja datan hyödyntäminen. Tässä prosessikuvauksessa prosessin eri vaiheet ovat tyypillisesti osittain päällekkäisiä. Prosessi on myös luonteeltaan jatkuva, koska yritykset kehittävät data-arkkitehtuuriaan jatkuvasti uusien tarpeiden mukaan.

Työn johtopäätöksissä todetaan, että toimiva data-arkkitehtuuri riippuu merkittävästi tarkastelun kohteena olevasta yrityksestä. Yritysten koko, saatavilla oleva tiedon määrä ja esimerkiksi yrityksen toimiala ovat merkittäviä tekijöitä sopivan data-arkkitehtuurin valinnassa. On kuitenkin olemassa myös yleisiä periaatteita ja toimintatapoja, joita kaikkien olisi hyvä noudattaa data-arkkitehtuuria kehitettäessä.

Avainsanat: Data-arkkitehtuuri, tiedonhallinta, strateginen päätöksenteko, tietovarasto

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

SISÄLLYSLUETTELO

1. JOHDANTO	1
2. TIETO PÄÄTÖKSENTEON TUKENA	3
2.1 Tiedonhallinta.....	3
2.2 Operatiivinen ja strateginen päätöksenteko.....	4
3. TIETOTYYPIT.....	7
3.1 Hiljainen tieto	7
3.2 Eksplisiittinen tieto	8
3.2.1 Strukturoitu data	8
3.2.2 Strukturoimaton data.....	10
4. TIETOLÄHTEET	12
4.1 Sisäiset tietolähteet.....	12
4.2 Ulkoiset tietolähteet.....	12
5. DATA-ARKKITEHTUURI	14
5.1 Periaatteet	14
5.2 Keskitetty arkkitehtuuri.....	15
5.3 Data-arkkitehtuurin kehitysprosessi.....	16
5.3.1 Lähtötilanteen arviointi	16
5.3.2 Tietotarpeiden ja tavoitteiden määrittely	18
5.3.3 Käsitteellinen mallintaminen.....	19
5.3.4 Tekniset ratkaisut.....	19
5.3.5 Tiedon hyödyntäminen.....	22
6. JOHTOPÄÄTÖKSET	25
LÄHTEET	27

1. JOHDANTO

Digitalisaation myötä erilaisten tietojärjestelmien lisääntyminen yritysten toiminnassa johtaa yritysten käytettävissä olevan datan määrän kasvamiseen. IDC:n ennustuksen mukaan uuden, tallennetun ja kopioidun datan määrä tulee olemaan vuonna 2025 yhteensä 175 tsettatavua (ZB), joka tarkoittaa että vuodesta 2018 datan määrä olisi maailmassa lähes kuusinkertaistunut (Reinsel et al. 2018 s.3). Iso osa tästä datasta on yritysten hallussa. Vuoden 2019 alussa yritysten tietojärjestelmiin on tallennettuna enemmän dataa kuin kaikkiin muihin maailman päätelaitteisiin yhteensä (Reinsel et al. 2018 s.3). Tietojärjestelmät sisältävät ja käsittelevät yrityksen toimintaan liittyvää tietoa ja tätä tietoa pystytään hyödyntämään kehitettäessä yrityksen toimintaa. Usein tietojärjestelmien ensisijainen tehtävä ei ole tallentaa tietoa esimerkiksi analytiikkaa varten, vaan tehostaa yrityksen toimintaa esimerkiksi automatisoimalla tai helpottamalla eri työtehtäviä. Tietojärjestelmiin tallentuvaa dataa pystytään kuitenkin käyttämään usein myös muissa käyttötapauksissa, kuin sen alkuperäisessä käyttötarkoituksessa kyseisen tietojärjestelmän sisällä.

Yrityksen käytettävissä olevan datan hyödyntäminen vaatii tuekseen prosessit siitä, kuinka dataa kerätään, analysoidaan ja käytetään lopulta päätöksenteon tukena. Prosessien lisäksi datan automaattinen käsittely vaatii myös teknisen toteutuksen edellä mainittujen tehtävien suorittamiseen. Tätä prosessia, missä data liikkuu yrityksen sisällä ja datan hallintaa kutsutaan yleisesti data-arkkitehtuuriksi. Toinen hieman tarkempi määritelmä data-arkkitehtuurille on: "Data-arkkitehtuuri on joukko sääntöjä, periaatteita, standardeja ja malleja, jotka määrittelevät mitä ja miten dataa kerätään, varastoidaan, hallitaan, hyödynnetään ja integroidaan organisaation sisällä ja sen tietokantajärjestelmissä" (Data Architecture, 2017, käännetty lähteestä).

Datan määrän kasvaessa yritysten haasteena on kehittää itselleen sopiva data-arkkitehtuuri, jotta olemassa olevaa dataa pystytään täysimääräisesti hyödyntämään päätöksenteon tukena. Tässä työssä käsitellään yritysten data-arkkitehtuuria erityisesti strategisen päätöksenteon tukena. Työ pyrkii vastaamaan tutkimuskysymykseen, millaisia yleisiä periaatteita ja prosesseja on olemassa, joita hyödyntämällä yritys voi kehittää itselleen strategista päätöksentekoa tukevan data-arkkitehtuurin.

Työ alkaa käsittelemällä tiedon merkitystä yrityksen sisäisessä päätöksenteossa. Tässä luvussa kuvataan, miksi yritysten tulisi systemaattisesti kerätä, käsitellä ja analysoida niiden käytettävissä olevaa dataa. Tämän lisäksi käsitellään operatiivisen ja strategisen päätöksenteon eroja ja syvennytään hieman tarkemmin strategiseen päätöksentekoon. Seuraavaksi työssä siirrytään eri tietotyyppien kuvaamiseen. On tärkeää ymmärtää minkälaista dataa yrityksellä on käytettävissään ja miten eri tietotyypit poikkeavat toisistaan esimerkiksi tiedon tallentamisessa ja sen käsittelyssä. Eri tietolajien kuvaamisen jälkeen käydään läpi myös erilaisia tietolähteitä, joita yrityksillä saattaa olla käytettävissään. Tietolähteitä voi olla hyvin erilaisia ja tietolähteiden moninaisuus on merkittävä tekijä yritysten data-arkkitehtuurissa. Tämän jälkeen työssä esitellään tyyppillisiä teknisiä ratkaisuja, jotka lopulta luovat tarvittavan infrastruktuurin datan käsittelyyn ja hyödyntämiseen.

Työssä on käytetty tiedonhakumenetelminä pääasiassa internetissä vapaasti saatavilla olevia lähteitä. Näitä ovat muun muassa alan erilaiset artikkelit ja kirjallisuus. Hakusanoina on käytetty esimerkiksi seuraavia: *strategic decision making*, *data-architecture*, *data management*, *tiedonhallinta* ja *data warehouse*. Tiedonhaussa on pääasiassa pyritty käyttämään kansainvälisiä lähteitä niiden paremman saatavuuden takia. Näiden lähteiden lisäksi työssä on myös hyödynnetty alan tunnettujen toimijoiden kotisivuja ja siellä saatavilla olevaa tietoa esimerkiksi käsitteiden määrittelyissä.

Tässä työssä ei tehdä eroa eri tiedon tasoille, vaan esimerkiksi käsitteillä tieto, informaatio ja data tarkoitetaan samaa asiaa. Tähän ratkaisuun on päädytty, koska tarkoituksena on kuvata yleisiä periaatteita kaikenlaisen tiedon hallinnassa. Liian tarkat määritelmät tekisivät tekstistä vaikeammin ymmärrettävää, eivätkä ne tukisi juurikaan työn tavoitteita.

2. TIETO PÄÄTÖKSENTEON TUKENA

2.1 Tiedonhallinta

Yritysten päätöksentekijöille on tärkeää, että heillä on saatavissa oikeanlaista tietoa auttamaan päätöksenteossa. Oikeanlainen tieto auttaa päätöksentekijöitä tekemään tietoon pohjautuvia ja hyvin perusteltuja päätöksiä yrityksen toiminnan kehittämiseksi. Oikeanlaisella tiedolla tarkoitetaan sellaista tietoa, joka on käyttäjälleen relevanttia ja joka on käyttäjän saatavilla oikeaan aikaan ja oikeanlaisessa muodossa. (Laihonen et al. 2013 s.44) Tällä tavalla määriteltynä onkin selvää, että yritykset pystyvät oikeanlaisen tiedon avulla tekemään parempia päätöksiä ja luomaan itselleen mahdollisesti myös kilpailuetua.

Oikeanlaisen tiedon saaminen sitä tarvitseville päätöksentekijöille vaatii tuekseen toimivat prosessit. Yleisesti tiedonhallinnalla tarkoitetaan prosesseja, joilla pyritään varmistamaan tietojen saatavuus, löydettävyyys ja hyödynnettävyyys eri tarkoituksiin (Finto 2018). Tiedonhallintaa voidaan kuvata prosessimallilla, johon kuuluvat tietotarpeiden tunnistaminen, tiedon hankinta, tiedon organisointi ja varastointi, tiedon jakelu, tiedon käyttö ja muutokset yrityksen toiminnassa (Laihonen et al. 2013 s.25). Tämän prosessimallin mukaiset vaiheet eivät kuitenkaan välttämättä etene edellä mainitussa järjestyksessä, vaan eri vaiheita voidaan suorittaa samanaikaisesti (Laihonen et al. 2013 s.26). Tiedonhallinnalla pyritään siis loppujen lopuksi siihen, että päätöksenteon tueksi saadaan jo aiemmin mainittua relevanttia ja oikeanlaista tietoa.

Tiedonhallinnan merkitys korostuu, kun maailmassa liikkuu yhä enemmän digitaalisessa muodossa olevaa dataa. Datan määrän lisääntyessä myös yritykselle mahdollisesti hyödyllisen tiedon määrä kasvaa. Mahdollisuuksien lisäksi tiedon lisääntyminen voi kuitenkin luoda yrityksille myös haasteita. Liiallinen tiedon määrä voi johtaa tietotulvaan, jonka johdosta tietoa tarvitseva asiantuntija ei ehdi keskittymään hänelle olennaisiin asioihin (Laihonen et al. 2013 s.15). Myös päätöksenteon kannalta olennaisen tiedon tunnistaminen voi vaikeutua, mikäli saatavilla olevaa tietoa on valtavasti. Tämän lisäksi myös totuttautuminen suureen määrään tietoa saattaa johtaa tilanteeseen, jossa päätöksiä ei pystytä enää tekemään ilman suurta tietomäärää päätöksenteon tukena. Tämä voi johtaa kyvyttömyyteen tehdä päätöksiä tai turhaan tehottomuuteen päätöksentekoprosesseissa.

Toinen yleinen haaste tiedonhallinnassa liittyy toimimattomiin tai yhteensopimattomiin tietojärjestelmiin, joita yrityksellä on käytössään (Laihonen et al. 2013 s.15). Yrityksen

tietojärjestelmissä oleva tieto on usein erittäin relevanttia yrityksen toiminnan kehittämisen kannalta. Näiden tietojärjestelmien keräämä data olisikin siis oleellista saada hyötykäyttöön, jotta olemassa olevaa tietoa pystyttäisiin täysimääräisesti hyödyntämään.

Tarkasteltaessa tiedonhallinnan prosessimallia ja tiedonhallintaan liittyviä tyypillisiä haasteita huomataan, että toimiva data-arkkitehtuuri tukee vahvasti lähes kaikkia tiedonhallinnan prosessin vaiheita ja auttaa vastaamaan sen mukanaan tuomiin haasteisiin. Tiedon hankinta, tiedon organisointi ja varastointi, tiedon jakelu ja tiedon käyttö ovat varmasti selkeimmät vaiheet, joissa hyvä data-arkkitehtuuri on lähes välttämätöntä. Digitalisaation myötä tiedonhallinta tulee myös todennäköisesti määritelmällisesti lähentymään data-arkkitehtuurin määritelmää.

2.2 Operatiivinen ja strateginen päätöksenteko

Yrityksissä tehdään sekä operatiivisen, että strategisen tason päätöksiä. Strategiset päätökset ovat tyypillisesti vaikutuksiltaan merkittäviä ja vaikuttavat pitkällä aikavälillä yrityksen toimintaan ja sen tavoitteiden saavuttamiseen. Operatiiviset päätökset vaikuttavat lyhyellä aikavälillä yrityksen toimintaan ja ovat tyypillisesti yrityksen toistuvissa liiketoimintaprosesseissa tehtäviä päätöksiä. Operatiivisessa päätöksenteossa päätösten tavoitteet ovat lyhyen aikavälin tavoitteita. (Abahmane & Binkkour 2008, s.2) Operatiivisessa päätöksenteossa päätöksentekotilanteet ovat siis usein nopeita ja niissä hyödynnettävä tieto tulee olla nopeasti saatavilla.

Operatiivisessa päätöksenteossa hyödynnettävän tiedon erityispiirteitä ovat esimerkiksi suppeammat tietotarpeet, tiedon säännöllinen päivittyminen, tarvittavan tiedon tarkkuus ja usein sisäinen tietolähde. (Abahmane & Binkkour 2008, s.4-7) Operatiivisessa päätöksenteossa yrityksen henkilöstöllä on usein käytettävissään eri tietojärjestelmiä, joiden avulla päätöksenteon tueksi saadaan nopeasti melko yksinkertaista tietoa. Operatiivisia tietojärjestelmiä voivat olla esimerkiksi toiminnanohjaus-, asiakastieto- ja toimitusketjujen ja logistiikan hallintajärjestelmät (Abahmane & Binkkour 2008, s.3) Data-arkkitehtuurin tehtävänä operatiivisessa toiminnassa on näiden järjestelmien ja niiden välisen tietoliikenteen kehittäminen palvelemaan paremmin yrityksen henkilöstön päätöksentekoa.

Strateginen päätöksenteko on luonteeltaan hyvin erilaista kuin operatiivinen päätöksenteko. Strategisten päätösten merkitys yrityksen menestykselle on suuri ja strategiset päätökset ovat luonteeltaan hyvin kompleksisia ja moniulotteisia.

Strategisissa päätöksissä yritys joutuu usein sitoutumaan vahvasti tehtävään päätökseen ja siitä seuraaviin toimenpiteisiin. (Gänswein, 2011, s.19-20) Laajemman vaikuttavuuden takia strategiseen päätöksentekoon voi perustellusti kuluttaa enemmän resursseja, mikäli se parantaa tehtyjä päätöksiä. Strategiseen päätöksentekoon voidaan siis käyttää enemmän aikaa ja sen tueksi pystytään perusteellisemmin keräämään ja analysoimaan päätöksentekoa tukevaa tietoa. Nämä asiat tekevät kuitenkin strategisesta päätöksentekoprosessista usein operatiivista päätöksentekoprosessia monimutkaisemman. Tarkastellaan seuraavaksi hieman tarkemmin strategista päätöksentekoa ja sen erityispiirteitä.

Yrityksen strategiseen päätöksentekoon liittyvän prosessin eri vaiheita voidaan kuvailla informaation hyödyntämisen tai poliittisen käyttäytymisen näkökulmista. Poliittisen käyttäytymisen näkökulma viittaa siihen, että päätöksiin saattaa vaikuttaa hyödynnettävän tiedon lisäksi myös päätöksentekijöiden omat intressit ja näiden saavuttamiseen käytettävät taktiikat. Yritykset voidaan nähdä poliittisina järjestelminä, joissa henkilöstön intressit voivat olla keskenään osittain ristiriitaisia ja näiden intressien tavoittelu voi vaikuttaa tehtäviin päätöksiin. (Gänswein, 2011, s.25) Poliittisen taktikoinnin ja henkilöstön omien intressien ajaminen ei todennäköisesti johda päätöksenteossa optimaaliseen lopputulokseen. Päätöksenteon taustalla oleva tiedon määrä ja laatu vaikuttavat todennäköisesti merkittävästi siihen, kuinka paljon poliittista taktikointia päätöksentekotilanteeseen liittyy. Päätöksenteon tukena käytettävän tiedon riittävä määrä ja laatu antavat vähemmän tilaa erilaisten poliittisten intressien ajamiseen.

Informaation hyödyntämisen näkökulmalla viitataan siihen, että strategiseen päätöksentekoon tarvitaan tietoa päätökseen vaikuttavista tekijöistä, oletuksia tulevaisuuden tapahtumista ja vaihtoehtoisista ratkaisuista ja niiden vaikutuksista. Päätöksentekoprosessiin liittyy epävarmuutta ja informaation hyödyntämisen lisääminen vähentää tätä epävarmuutta. (Gänswein, 2011, s.25) Päätöksenteossa halutaan ottaa kattavasti huomioon eri päätöksistä aiheutuvat mahdolliset seuraukset, eri skenaarioiden toteutumiseen vaikuttavat tekijät ja tietoon perustuvat arviot eri tapahtumien todennäköisyyksistä. Näin kattavien ja kompleksisten tietotarpeiden tyydyttämiseen tietoa tulee hyödyntää usein monista eri lähteistä. Tieto vähentää päätöksentekoon liittyvää epävarmuutta ja täten kehittää tehtävää päätöksentekoa.

Strategisessa päätöksenteossa hyödynnettävän tiedon erityispiirteitä ovat esimerkiksi tietotarpeiden suuri laajuus, aggregoitu tieto, tilastollisten menetelmien hyödyntäminen, datan pitkäaikainen kerääminen ja sisäiset ja ulkoiset tietolähteet. (Abahmane & Binkour 2008, s.4-7) Strategisessa päätöksenteossa operatiivisten järjestelmien tiedon lisäksi tarvitaan siis usein laajasti kerättyä ja analysoitua dataa. Data-arkkitehtuurin yksi

merkittävä tavoite on luoda oma infrastruktuuri tämänkaltaisen päätöksenteon tueksi. Tämä tarkoittaa käytännössä monien eri tietolähteiden yhdistämistä, datan pitkäaikaista keräämistä ja erilaista analytiikkaa, jolla tietoa saadaan jalostettua paremmin päätöksenteon tueksi. Tässä työssä käsitellään data-arkkitehtuuria pääasiallisesti strategisen päätöksenteon tukena ja kuvataan edellä mainitun infrastruktuurin kehityksessä käytettäviä yleisiä prosesseja ja periaatteita.

3. TIETOTYYPI

3.1 Hiljainen tieto

Yrityksillä on usein käytössään hyvin monenlaista tietoa ja lukuisia mahdollisia tietolähteitä. Osaa tästä tiedosta voidaan kuvata hiljaisena tietona. Hiljainen tieto tarkoittaa yrityksen henkilöstön kerryttämää osaamista ja intuitiota. Tällaista tietoa voi olla vaikea sanallistaa tai tallentaa eksplisiittisesti. (Laihonen et al. 2013 s.18) Hiljaisen tiedon luonteen vuoksi sen hallinnan voidaan arvioida olevan paljon haastavampaa kuin eksplisiittisen tiedon, jota voidaan helposti dokumentoida ja tallentaa esimerkiksi yrityksen tietojärjestelmiin. Tietojärjestelmien avulla kerättyä tietoa on helpompi käsitellä ja lopulta jakaa sitä tarvitseville.

Hiljaisessa tiedossa tietolähteenä toimii siis yrityksen kokenut henkilöstö, joka on pitkäaikaisella työskentelyllään kerryttänyt osaamista omalta aihealueeltaan. Hiljaisen tiedon eri tietotyyppejä on haastavaa määrittää, koska määritelmällisesti tieto ei ole eksplisiittisessä muodossa.

Hiljainen tieto ja eksplisiittinen tieto eivät kuitenkaan ole täysin erotettuja toisistaan, vaan hiljaista tietoa pystytään osin muokkamaan myös eksplisiittiseen muotoon. Hiljaisen tiedon konvertointia eksplisiittiseen muotoon voidaan kuvata esimerkiksi SECI-mallilla. Tämän prosessimallin ensimmäisessä vaiheessa kokemattomimmat työntekijät työskentelevät kokeneiden työntekijöiden kanssa, joilla on paljon kokemustensa avulla kerrytettyä hiljaista tietoa. Seuraavaksi kokemattomimmat työntekijät pyrkivät kirjaamaan uuden tietonsa eksplisiittiseen muotoon. Uusi eksplisiittinen tieto yhdistetään yrityksen muuhun eksplisiittiseen tietoon, josta yrityksen muut työntekijät sisäistävät tämän uuden yhdistetyn tiedon jälleen hiljaiseksi tiedoksi. (Kimble 2013) Tätä prosessimallia tarkasteltaessa herää kysymys, miksi kokenut työntekijä ei vain itse pyri eksplisiittisesti kirjaamaan hänelle kertynyttä hiljaista tietoa. Tämä olisi varmasti joissain tapauksissa osittain mahdollista, mutta kokeneet työntekijät eivät välttämättä tiedosta, mitä hiljaista tietoa heille on kertynyt. Kokemattomampi työntekijä on mahdollisesti perehtynyt vain eksplisiittiseen tietoon työtehtävistään ja aihealueestaan ja tunnistaa tämän avulla ehkä helpommin, mikä kaikki on hiljaista tietoa.

Yrityksellä voi olla omat prosessinsa, kuinka hiljaista tietoa saadaan jaettua esimerkiksi uusille työntekijöille ja yleisesti koko organisaatioon. Tiedon muuttaminen eksplisiittiseen muotoon helpottaa kuitenkin tiedonhallintaa, koska tietoa pystytään tällöin käsittelemään digitaalisesti. Tiedon digitaalinen hallinta vastaa helpommin esimerkiksi seuraaviin

kysymyksiin: mistä tieto on peräisin, kenen saatavilla tieto on ja mitä kaikkea tietoa yrityksellä on käytettävissään. Hiljainen tieto tuo mukanaan myös haasteita tiedon säilymiselle osaavan henkilöstön jäädessä esimerkiksi eläkkeellä tai vaihtaessaan työpaikkaa. Näiden asioiden vuoksi eksplisiittinen tieto saatetaan nähdä yrityksissä halutumpana ja hyödyllisempänä kuin hiljainen tieto.

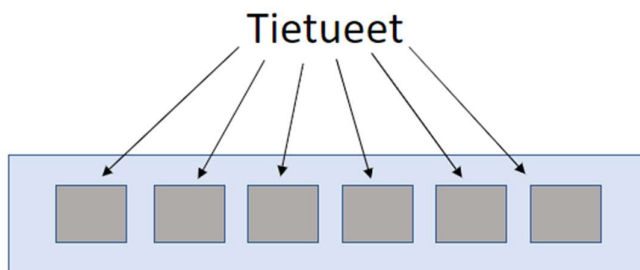
Myös data-arkkitehtuurin näkökulmasta voidaan ajatella, että hiljainen tieto on hyödynnettävissä vain jos siitä saadaan konvertoitua eksplisiittistä tietoa. Näin tiedon keräämistä, tallentamista, integroimista ja käyttöä pystytään hallinnoimaan ja luomaan erilaisia sääntöjä ja periaatteita, miten tämä kaikki toteutetaan. Konvertoinnin jälkeen data voi olla joko strukturoitua tai strukturoimatonta. Näiden käsitteiden eroja käsitellään seuraavissa alaluvuissa.

3.2 Eksplisiittinen tieto

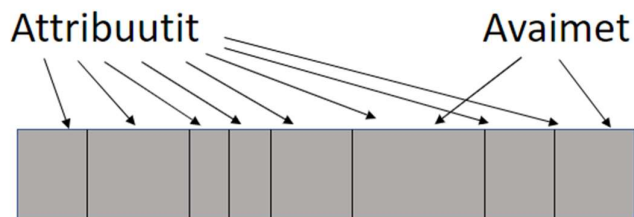
3.2.1 Strukturoitu data

Eksplisiittistä tietoa voidaan jakaa erilaisiin alakategorioihin. Tässä työssä on päädytty jakamaan se strukturoituun ja strukturoimattomaan dataan. Aluksi käsitellään strukturoidun datan määritelmää ja sen erityispiirteitä, jonka jälkeen siirrytään strukturoimattomaan dataan.

Strukturoidulla datalla tarkoitetaan sellaista dataa, joka on hyvin määriteltyä ja rakenteeltaan samanlaista. Tämänkaltaista dataa on yrityksessä kuitenkin selvästi vähemmän kuin strukturoimatonta dataa. Datan suhteellisen pienestä määrästä huolimatta strukturoitu data on yleensä hyvin arvokasta liiketoiminnalle. (Inmon et al. Luku 1.1) Tämä on ymmärrettävää, koska monet yritysten perustoiminnot tuottavat tämäntyyppistä dataa ja tietoa perustoiminnoista voidaan pitää lähtökohtaisesti arvokkaana. Alla olevissa kuvissa esitetään tyypillinen rakenne strukturoidulle datalle, joka voi kuvastaa esimerkiksi myyntitransaktioita.



Kuva 1. Strukturoidun datan kuvaus



Kuva 2. Strukturoidun datan kuvaus yhdestä tietueesta

Ensimmäisessä kuvassa esitetään joukkoa strukturoidun datan tietueita. Yksittäinen tietue kuvastaa tässä esimerkissä yhtä myyntitransaktiota. Tähän myyntitransaktioon liittyy useita eri attribuutteja, joita voivat olla esimerkiksi myyntipäivä, kappalehinta, kappalemäärä ja asiakastunniste. Yksi tai useampi näistä attribuuteista toimii tietuen yksilöivänä avaimena. Tässä esimerkissä avaimena voisi toimia esimerkiksi surrogaattiavain nimeltä "myyntitapahtuma_ID". Surrogaattiavaimella tarkoitetaan keinotekoisia avainta, jonka tarkoitus on yksilöidä tietue tietojärjestelmässä ja joissain tapauksissa nopeuttaa esimerkiksi tietokantahakuja (Linstedt & Olschimke 2016, s.97-98). Avaimet liittyvät erityisesti strukturoitua dataa käsitteleviin tietojärjestelmiin ja tiedon tallentamis- ja käsittelytapaan, eivätkä kuvasta strukturoidun datan luontaisia ominaisuuksia.

Tyypillisiä esimerkkejä strukturoidusta datasta ovat esimerkiksi tiedot myydyistä tuotteista, asiakastiedot tai käyttäjien klikkaukset yrityksen verkkosivuilla. Erilaista strukturoitua tietoa ja sen mahdollisia hyödyntämistapoja on melko helppo keksiä. Esimerkiksi myytyjen tuotteiden tiedoilla voidaan arvioida uuden tuoteryhmän kiinnostavuutta tai ennakoida kysynnän muutosta eri vuodenaikoina.

Strukturoitua dataa on helppo käsitellä ja täten datasta saadaan myös melko helposti tehtyä erilaisia analyyseja ja johtopäätöksiä. Helpon käsittelyn lisäksi strukturoitu data on usein toisteista ja jatkuvaa (Inmon et al. Luku 1.1).

Strukturoidun datan helpon käsittelyn, korkean arvon ja toisteisuuden/jatkuvuuden takia strukturoidun datan hallinta tulisikin olla ensimmäisiä asioita joita ottaa huomioon yrityksen data-arkkitehtuurin toteutuksessa. Myöhemmin työssä esitellään tyypillinen ratkaisu strukturoidun datan käsittelyssä koko yrityksen tasolla.

3.2.2 Strukturoimaton data

Strukturoimattomalla datalla tarkoitetaan sellaista dataa, jolla ei ole ennalta määriteltyä muotoa ja se ei sovi tallettavaksi esimerkiksi perinteiseen relaatiotietokantaan. Strukturoimaton data voi sisältää esimerkiksi kuvia, tekstiä, ääntä ja videokuvaa. (Isson & Hwang 2018, Luku 2) Strukturoimatonta dataa on yrityksiä saatavilla todella paljon. Yrityksen strukturoimattoman datan määrän on arvioitu olevan jopa yli 80% kaikesta yrityksen datasta (Inmon et al. Luku 4). Tämä on ymmärrettävää, koska esimerkiksi järjestelmien lokitiedot, sähköpostikeskustelut ja yrityksen nauhoittamat puhelut voidaan luokitella yrityksen strukturoimattomaksi dataksi.

Tyypillisesti yrityksissä on sekä toistuvaa strukturoimatonta dataa, että kertaluonteista strukturoimatonta dataa. Toistuva strukturoimaton data on rakenteeltaan ja sisällöltään lähes samanlaista, kun taas kertaluonteinen strukturoimaton data ei ole rakenteeltaan eikä sisällöltään samankaltaista. (Inmon et al. Luku 4). Yritysten tietojärjestelmien ja koneiden tuottamat lokitiedot ja erilaiset mittarit ovat hyviä esimerkkejä toistuvasta strukturoimattomasta datasta. Kertaluonteista strukturoimatonta dataa taas ovat esimerkiksi asiakaspalautteet, sähköpostit ja yritykseen liittyvät sosiaalisen median kommentit.

Erottelu toistuvaan ja kertaluonteiseen strukturoimattomaan dataan on tärkeää, koska tyypillisesti niiden sisältämän tiedon arvo poikkeaa merkittävästi toisistaan. Toistuvassa strukturoimattomassa datassa on harvoin lisäarvoa tuottavaa tietoa, kun taas kertaluonteisessa strukturoimattomassa datassa sitä on paljon (Inmon et al. Luku 4). Aiemmassa kappaleessa annetuista esimerkeistä työntekijöiden väliset sähköpostit voivat sisältää todella arvokasta tietoa. Yrityksen laitteiden tuottamat lokitiedot taas eivät normaalissa toiminnassa tuota juurikaan hyödyllistä tietoa yritykselle.

Huolimatta strukturoimattoman datan suuresta määrästä, sitä ei hyödynnetä juurikaan yrityksen päätöksenteossa. Selvästi suurin osa yrityksen päätöksenteosta perustuu strukturoituun tietoon, vaikka strukturoidun tiedon määrä on yrityksissä huomattavasti pienempää. Tämä johtuu pääasiassa strukturoidun datan käsittelyn helposta automatisoinnista ja analysoinnista. (Inmon et al. Luku 4). Edellisessä kappaleessa kuitenkin todettiin, että kertaluonteinen strukturoimaton data voi olla yrityksille todella hyödyllistä. Tämänkaltaisen tiedon hyödyntämisessä yrityksen tulisi kuitenkin arvioida tarkasti omia tietotarpeitaan ja käytettävissä olevia resurssejaan.

Strukturoimattoman datan vaikeammasta hallinnoimisesta ja käsittelystä huolimatta yritysten tulisi ainakin perehtyä sen tuomiin mahdollisuuksiin. Erityisesti kertaluonteisen

strukturoimattoman datan hallinnointi tulisi ottaa huomioon yrityksen data-arkkitehtuuria suunniteltaessa.

Strategisen päätöksenteon ja sitä tukevan data-arkkitehtuurin kannalta eri tietotyyppien tunnistaminen ja niiden eri ominaispiirteiden ymmärtäminen on erittäin tärkeää. Tietotyyppien erilainen hyödynnettävyys ja käsittelytavat vaikuttavat yrityksen data-arkkitehtuuria koskeviin päätöksiin. Tarkastelemalla esimerkiksi tiedon sisältämää arvoa liiketoiminnalle ja arvioimalla tiedon hankkimiseen ja käsittelyyn liittyviä haasteita, voidaan näiden pohjalta arvioida esimerkiksi erilaisia teknisiä ratkaisuja tai priorisoida seuraavia kehitysaskelia.

Tietotyyppien erilaisen hyödynnettävyyden ja käsittelytapojen pohjalta voidaan johtaa yleisiä periaatteita, kuinka eri tietotyyppistä voidaan ottaa huomioon yrityksen data-arkkitehtuurin kehityksessä. Yrityksen henkilöstölle kertynyttä hiljaista tietoa tulisi systemaattisesti pyrkiä muuttamaan eksplisiittiseen muotoon. Tällä tavalla tietoa pystytään paremmin hallinnoimaan ja jakamaan yrityksen sisällä sitä tarvitseville päätöksentekijöille. Eksplisiittisestä tiedosta strukturoidun datan helpon käsittelyn, korkean arvon ja toisteisuuden/jatkuvuuden takia se on usein hyvä lähtökohta data-arkkitehtuurin kehittämiseksi. Tämänkaltaiseen tietoon ja sen mahdollisuuksiin tulisi siis usein ensimmäiseksi kiinnittää huomiota yrityksen data-arkkitehtuuria suunniteltaessa. Strukturoimattoman datan merkittävästi vaikeampi käsittely ja suuri määrä tekee sen hyödynnettävyydestä selvästi hankalampaa. Strukturoimatonta tietoa tulisikin useissa tapauksissa alkaa hyödyntämään vasta hieman myöhemmässä vaiheessa yrityksen data-arkkitehtuurin kehitystä.

4. TIETOLÄHTEET

4.1 Sisäiset tietolähteet

Edellisessä osiossa keskityttiin erilaisiin tietotyyppeihin, joihin dataa voidaan luokitella. Tässä osiossa siirrytään tarkastelemaan, mistä tämä tieto tulee yrityksille. Tietolähteet voivat olla yrityksen sisäisiä tai ulkoisia tietolähteitä. Internetin tultua yleiseen käyttöön yrityksillä on käytössään valtavasti ulkoisia tietolähteitä. Sisäisiä tietolähteitä taas yrityksillä on käytettävissään huomattavasti vähemmän. BI-Surveys tekemän kyselyn mukaan yrityksillä on käytössään keskimäärin viisi sisäistä tietolähdettä päätöksenteon tueksi (BI-Survey). Vaikka ulkoisia tietolähteitä on olemassa paljon enemmän, tiedon luotettavuus on yleisesti paljon huonompaa kuin sisäisten tietolähteiden. Myös suuri osa yrityksen saatavilla olevista ulkopuolisista tietolähteistä sisältää yrityksen kannalta epärelevanttia tietoa. Molempien tietolähteiden käyttöä olisi kuitenkin syytä harkita, koska myös ulkopuoliset tietolähteet voivat sisältää todella arvokasta ja helposti saatavilla olevaa tietoa.

Tyypillisiä yrityksen sisäisiä tietolähteitä ovat yrityksen eri tietojärjestelmät. Erilaisia sisäisiä tietojärjestelmiä ovat esimerkiksi asiakastieto-, toiminnanohjaus- ja varastonhallintajärjestelmät. Nämä kaikki sisältävät yrityksen toiminnan kannalta todella relevanttia tietoa. Tietojärjestelmien lisäksi myös kaikki tietojärjestelmien ulkopuoliset tiedostot lasketaan yrityksen sisäisiksi tietolähteiksi. Näitä voivat olla esimerkiksi henkilökohtaiselle tietokoneelle tai pilvipalveluun tallennetut analyysit, esitysmateriaalit ja muistiot. Myös aiemmin mainittu yrityksen työntekijöiden hiljainen tieto voidaan laskea sisäiseksi tietolähteeksi.

4.2 Ulkoiset tietolähteet

Yrityksen ulkoisia tietolähteitä voivat olla esimerkiksi avoimen datan lähteet tai sosiaalinen media (Pitt 2019). Monet tahot tarjoavat kaikille avointa dataa, jota voidaan hyödyntää yrityksen päätöksenteossa. Esimerkiksi julkisen sektorin Tilastokeskus tarjoaa paljon sen hallussa olevaa avointa dataa kenelle tahansa. Sosiaalista mediaa voi hyödyntää tietolähteenä analysoimalla esimerkiksi yleisiä keskustelunaiheita tai tekemällä sentimenttianalyysia siitä, mistä ihmiset yleisesti pitävät (Pitt 2019). Voidaan kuitenkin ajatella, että yrityksen henkilöstö käyttää ainakin tiedostamattaan monia yrityksen ulkopuolisia tietolähteitä omien päätöksensä tueksi. Tämä jätetään kuitenkin yrityksen data-arkkitehtuurissa tarkastelun ulkopuolelle, jotta rajauksesta tulisi

selkeämpi. Ulkopuolisia tietolähteitä tarkastellaan siis vain systemaattisen tiedonhallinnan näkökulmasta.

Yritykset voivat siis hyödyntää tietolähteinään sekä sisäisiä, että ulkoisia tietolähteitä. Yrityksen sisäiset tietolähteet pitävät sisällään yrityksen kannalta lähes täysin relevanttia tietoa, kun taas suurin osa yrityksen ulkoisista tietolähteistä sisältää yrityksen kannalta epärelevanttia tietoa. Myös tiedon saatavuus, laatu ja alkuperä ovat usein paremmin tiedossa ja hallinnassa yrityksen sisäisissä tietolähteissä. Näiden tekijöiden ansiosta yritysten tulisi usein keskittyä aluksi sisäisiin tietolähteisiin kehittäessään omaa data-arkkitehtuuriaan. Ulkoisia tietolähteitä ei kuitenkaan kannata kokonaan jättää huomioimatta ja joissain tilanteissa ulkoiset tietolähteet voivat olla jopa paljon merkittävämpiä kuin sisäiset tietolähteet. Erityisesti strategisen päätöksenteon tukena hyödynnetään usein sekä sisäisiä, että ulkoisia tietolähteitä (Abahmane & Binkkour 2008, s.4).

5. DATA-ARKKITEHTUURI

5.1 Periaatteet

Data-arkkitehtuuriin ja datan hallintaan on kehitetty paljon yleispäteviä periaatteita, joita dataa käsittelevien yritysten olisi hyvä noudattaa. Näitä ovat esimerkiksi datan laadun varmistaminen, datan turvaaminen ja lakien/säännösten noudattaminen, master datan hallinta, datan omistajuus ja data-arkkitehtuuri (NodeGraph 2020). Tässä työssä datan hallinta on sisällytetty data-arkkitehtuurin käsitteeseen, joten näitä periaatteita voidaan käsitellä tässä tapauksessa osana data-arkkitehtuurin periaatteita.

Datan laadun varmistamisella tarkoitetaan sitä, että data on esimerkiksi luotettavaa, tarpeeksi tarkkaa, ja ei sisällä esimerkiksi duplikaatteja tai muuten virheellistä tietoa (NodeGraph 2019). Datan laadun varmistaminen on erittäin tärkeä osa yrityksen data-arkkitehtuuria. Mikäli datan laatuun ei kiinnitetä huomiota, niin pahimmassa tapauksessa päätöksiä voidaan tehdä virheellisen tiedon perusteella. Myös päätöksentekijöiden luottamus dataan ja sen käytettävyyteen heikkenee, mikäli datasta paljastuu merkittäviä puutteita. Tämä voi johtaa siihen, että dataa ei enää pyritä hyödyntämään päätöksenteon tukena. Datan laadun auditointi ja huonolaatuisen datan poistaminen/muokkaaminen tulisi siis olla osa yritysten data-arkkitehtuuria.

Toisena periaatteena on datan turvaaminen ja erilaisten säädösten noudattaminen. Datan turvaamiseen liittyy monia eri näkökulmia. Dataa voidaan turvata varmuuskopioinneilla mahdollisten virhetilanteiden varalta, mutta datan turvaamisella voidaan tarkoittaa myös tietoturvaa. Kerätty data voi olla todella arkaluontoista ja liiketoimintakriittistä. Tällaisen tiedon ei haluta päätyvän kilpailijoille tai muille yritysten ulkopuolisille henkilöille. Datan saatavuuden tulisi myös olla tarpeen mukaan rajattua, jotta esimerkiksi luottamuksellista tietoa ei ole kaikkien saatavilla. Myöskään osaamattomalle henkilöstölle ei tule antaa suoraa pääsyä datan hallintaan liittyviin järjestelmiin, jotta turhilta vahingoilta voidaan välttyä.

Myös erilaisten säädösten noudattaminen täytyy huomioida kun dataa aiotaan tallentaa ja käsitellä. Esimerkiksi ihmisten henkilötietojen tallentamisessa tulee ottaa huomioon tätä koskeva lainsäädäntö. Datan määrän lisääntyminen tuo tätä näkökulmaa jatkuvasti enemmän esille, ja se tulee todennäköisesti myös vaikuttamaan datan hallintaan liittyvään regulaatioon. Yritysten tulee siis kiinnittää tulevaisuudessa entistä enemmän huomiota erilaisiin lakeihin ja säädöksiin, jotka liittyvät datan hallintaan.

Master datalla tarkoitetaan sellaista dataa, jota pidetään organisaatiossa kyseisen datan ainoana oikeana lähteenä. Master datan hallinnalla pyritään luomaan, hyödyntämään ja hallitsemaan tämänkaltaista dataa. (Linsteadt & Olschimke 2016, s.229-231) ”MDM on kokoelma sääntöjä, prosesseja ja ohjelmistoja, joiden tehtävänä on mahdollistaa varsinaisen master datan tuottaminen.” (Heinänen 2020) Master datan hallinnalla (Master Data Management – MDM) pyritään siis ratkaisemaan ristiriitoja, joita eri tietolähteisiin kertyneellä tiedolla voi olla keskenään. Tämä voidaan nähdä yhtenä isona kokonaisuutena tiedon luotettavuuden parantamisessa.

Viimeisenä periaatteena tässä listauksessa on datan omistajuus. Datan omistajuudella tarkoitetaan sitä, että yrityksellä tulisi olla henkilöt, jotka ovat datan hallinnasta kokonaisvastuussa. Tätä vastuualuetta on tietysti järkevää muuttaa tilanteen mukaan, mutta periaatteessa kaikelle datalle tulisi olla liiketoiminnan näkökulmasta omistaja. Omistaja voidaan määritellä esimerkiksi liiketoimintayksiköittäin tai jollain muulla datan sisällön kannalta loogisella tavalla. Datan omistaja on henkilö, jolle voi esittää erilaisia kysymyksiä, raportoida virheitä tai antaa esimerkiksi kehitysehdotuksia dataan liittyen. Datan omistaja osaa delegoida kysymykset ja ehdotukset oikeille henkilöille ja toimia rajapintana yrityksen muulle henkilöstölle.

Edellä mainitut periaatteet liittyvät kaikki vahvasti onnistuneeseen datan hallintaan ja data-arkkitehtuuriin. Seuraavassa kappaleessa käsitellään yhtä oleellista data-arkkitehtuurin piirrettä strategisen päätöksenteon tukena, eli keskitettyä arkkitehtuurirakennetta.

5.2 Keskitetty arkkitehtuuri

Datan tehokas kerääminen, muokkaaminen ja hallinta on järkevää tehdä mahdollisimman keskitetysti. Tämä tarkoittaa, että tietolähteistä saatava tieto pyritään ensin tuomaan mahdollisimman keskitetysti yhteen paikkaan tai järjestelmään. Tämän jälkeen myös datan mallinnus ja mahdollinen muokkaus toteutetaan yhdessä paikassa. Vasta tarjottaessa yhdistettyä ja käsiteltyä dataa sen loppukäyttäjille, täytyy tietoa levittää taas organisaation sisällä. Tyypillisesti ilman systemaattista data-arkkitehtuuria erilaista dataa löytyy yrityksessä todella monesta paikasta. Tähän liittyy monia haasteita, johon keskitetty data-arkkitehtuuri auttaa vastaamaan.

Ensimmäisenä haasteena on monimutkainen data-arkkitehtuuri. Monimutkaiset järjestelmäintegraatiot ja datan siiloutuminen luovat haasteita data-arkkitehtuurin ylläpitämiselle ja erityisesti sen jatkokehittämiselle. Keskitetty tiedon tallentaminen suoraviivaistaa data-arkkitehtuuria ja tekee siitä helpommin hallittavaa. Tiedon

tallentamisessa data liikkuu vain yhteen suuntaan, eli kohti keskitettyä tallennuspaikkaa. Täältä data päätty sen mahdollisen muokkaamisen ja analysoinnin jälkeen lopulta loppukäyttäjälle. Ratkaisu on arkkitehtuuriltaan selkeä ja helposti ylläpidettävä.

Tieto kaikesta saatavilla olevasta datasta on myös helposti saatavilla. Mikäli kaikki tieto tallennetaan keskitetysti, niin tarvittavaa tietoa on helppo etsiä. Eri tietolähteiden ja niiden sisältämän datan yhdistäminen on myös helpompaa loppukäyttäjälle, koska kaikki tieto on jo valmiiksi mallinnettuna keskitetysti. Keskitetty tiedon tallentaminen myös ehkäisee niitä ristiriitoja, jotka syntyvät kun samanlaista dataa haetaan eri tietolähteistä.

Myös datan alkuperä on helposti selvitettävissä datan keskitetyn hallinnan avulla. Mikäli data on haettu suoraan esimerkiksi yrityksen lähdejärjestelmästä sen tietovarastoon, niin tiedon alkuperä on helppo tallentaa ja myöhemmässä vaiheessa tarvittaessa selvittää. Myös mahdolliset muutokset datassa ovat voineet tapahtua vain tietovarastossa, joten näiden muutoksien lähteen paikannus ja tehdyt muutokset on helpommin selvitettävissä.

Data-arkkitehtuurin täydellinen keskittäminen on kuitenkin usein lähes mahdotonta. Valmiiksi olemassa olevat integraatiot ja datan liikkumisen eri operatiivisten järjestelmien välillä johtaa käytännössä siihen, että täysin suoraviivaista data-arkkitehtuuria ei pystytä toteuttamaan. Tämän lisäksi erilaiset tietotyypit voivat vaatia erilaisia tallentamis- ja käsittelytapoja. Tästä johtuen näiden tehokkaaseen hallintaan voidaan vaatia useampia järjestelmiä. Data-arkkitehtuurin tavoitteena on kuitenkin hyvä pitää suoraviivaista ja keskitettyä arkkitehtuuria. Tästä joudutaan kuitenkin esimerkiksi edellä mainituista syistä osittain poikkeamaan.

5.3 Data-arkkitehtuurin kehitysprosessi

Tässä luvussa tarkastellaan yleisiä vaiheita, joita tulisi data-arkkitehtuuria toteutettaessa ottaa huomioon. Luku noudattaa prosessimaista kuvausta, jossa alaluvut ovat kuitenkin todellisuudessa usein osittain päällekkäisiä. Tässä työssä esitetty järjestys ei myöskään ole ainoa oikea vaihtoehto, vaan ainoastaan yksi mahdollinen prosessikuvaus data-arkkitehtuurin toteutuksesta. Prosessi on myös jatkuva ja sen eri vaiheita toistetaan jatkuvasti kehitettäessä yrityksen data-arkkitehtuuria. Prosessikuvausta voi siis hyödyntää, vaikka yrityksellä olisi jo olemassa olevaa arkkitehtuuria datan keräämiselle, käsittelylle ja hyödyntämiselle.

5.3.1 Lähtötilanteen arviointi

Data-arkkitehtuurin systemaattisessa toteutuksessa on hyvä lähteä liikkeelle lähtötilan arvioinnilla. Lähtötilan arvioinnilla pyritään saamaan käsitys yrityksen nykytilasta data-arkkitehtuurin suhteen. Yrityksillä on tyypillisesti jo käytössään erilaisia tietojärjestelmiä ja yrityksen hallussa olevaa tietoa käytetään jo päätöksenteon tukena. Näiden prosessien perusteellinen selvittäminen tulisi olla ensimmäinen askel systemaattisen data-arkkitehtuurin kehittämisessä. Tavoitteena on siis hankkia kattava ymmärrys saatavilla olevasta datasta, sen laadusta, alkuperästä ja tämän hetkistä hyödyntämistavoista.

Lähtötilan arviointi tarkoittaa eri tietolähteiden listausta, arviointia ja tietoa siitä, mitkä tietojärjestelmät ovat esimerkiksi integroituna toisiinsa. Tietolähteiden arvioinnin yhteydessä tulisi myös tarkasti perehtyä niiden sisältämään dataan. Sama tieto saattaa sijaita useassa eri järjestelmässä ja näiden järjestelmien sisältämä datan laatu ja granulariteetti voivat olla hyvin erilaista. Samankaltaista dataa saattaa löytyä eri järjestelmistä integraatioiden myötä, mutta myös integraatioiden puute voi johtaa samankaltaisen datan tallentamiseen useisiin eri järjestelmiin. Yrityksessä voi esimerkiksi olla oma asiakastietojärjestelmä, jossa asiakastiedot ovat tarkimmalla tasolla ja myyntijärjestelmä, jossa asiakkaista on tallennettuna vain perustiedot.

Tietolähteiden ja niiden sisältämän datan listauksen ja arvioinnin jälkeen tulisi tarkastella kuinka dataa tällä hetkellä hyödynnetään yrityksessä päätöksenteon tukena. Yrityksessä saatetaan jo tehdä henkilöstön toimesta esimerkiksi omaa vastuualuetta koskevaa raportointia ja analyttikkaa ilman systemaattisia prosesseja. Näiden tunnistaminen auttaa tunnistamaan paremmin, mitä tietotarpeita yrityksen päätöksentekijöillä on ja minkälaisia haasteita esimerkiksi datan laadussa, saatavuudessa tai sen hyödyntämisessä on.

Osana lähtötilan arviointia pystytään myös tunnistamaan muita kehityskohteita yrityksessä. Tietolähteiden ja -järjestelmien sisältävän datan sisällöstä ja liikkumisesta saadaan tarkempi kuva, joka voi auttaa tunnistamaan mahdollisia heikkouksia esimerkiksi järjestelmien välisissä integraatioissa tai järjestelmien sisältämän tiedon laadussa. Näiden havaintojen perusteella voidaan lähteä tekemään operatiivisten järjestelmien kehitystä tai luomaan uusia integraatioita niiden välille. Tämänkaltaisen sovelluskehitys voidaan myös luokitella data-arkkitehtuurin kehittämiseksi.

5.3.2 Tietotarpeiden ja tavoitteiden määrittely

Edellisessä aluvuossa käsitelty data-arkkitehtuurin lähtötilanteen arviointi menee usein vahvasti päällekkäin tietotarpeiden ja tavoitteiden määrittelyn kanssa. Lähtötilannetta määriteltäessä kokonaiskuva yrityksen tietotarpeista vahvistuu ja hyvin toteutetun data-arkkitehtuurin mahdollistamat tavoitteet datan hyödyntämiselle selkeytyvät. Tavoitteena tietotarpeiden ja tavoitteiden määrittämisessä on osaltaan tunnistaa tärkeimmät asiat, joihin data-arkkitehtuurissa tulisi ensimmäiseksi kiinnittää huomiota. Tämän lisäksi tulisi ottaa huomioon tulevaisuuden tavoitteet ja niihin mahdollisesti liittyvät uudet tietolähteet ja hyödyntämistavat.

Tärkeimpien tietotarpeiden tunnistamisen avulla voidaan arvioida, mistä tiedosta ja tietolähteistä kannattaa lähteä liikkeelle. Data-arkkitehtuurin rakentaminen on jatkuva prosessi, jossa kannattaa aloittaa kaikista kriittisimpien tietotarpeiden tunnistamisesta. Tässä on kuitenkin hyvä muistaa esimerkiksi kustannustekijät ja tietotarpeiden vaativuus. Tietotarpeita tunnistettaessa voidaan löytää todella helposti hyödynnettävää tietoa, jolla on kuitenkin merkittävää arvoa päätöksentekijöille. Tietotarpeita ja tavoitteita tunnistettaessa on siis järkevää ottaa huomioon tietotarpeiden tyydyttämiseen vaadittavat resurssit ja suhteuttaa tätä tiedosta saatavaan lisäarvoon.

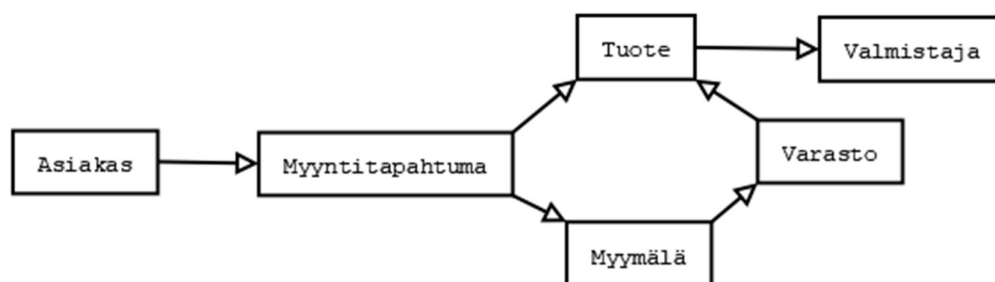
Hyvällä tavoitteiden määrittelyllä voidaan jo ennakolta varautua mahdollisiin muutoksiin hyödynnettävissä tietolähteissä, tietotyypeissä tai tiedon uudellisissa hyödyntämistavoissa. Data-arkkitehtuuria voidaan alkaa lähtökohtaisesti rakentamaan sillä ajatuksella, että tietynlaisiin muutoksiin ollaan varauduttu. Tämä vähentää tulevaisuudessa tehtäviin muutoksiin käytettävää työmäärää ja muutoksista mahdollisesti aiheutuvia virhetilanteita.

Tietotarpeiden ja tavoitteiden määrittelyssä pyritään arvioimaan minkälaisiin kysymyksiin tiedon avulla halutaan pyrkiä vastaamaan nyt ja tulevaisuudessa. Yrityksen eri toiminnoilla voi olla hyvin erilaisia tietotarpeita ja tavoitteita. Tietotarpeiden laajuuden ja tiedon hyödyntämisen mahdollisten vaikutusten arvioinnilla pystytään paremmin suunnittelemaan tarkoituksenmukainen data-arkkitehtuuri. Esimerkiksi hyvin vähäisten tietotarpeiden täyttämiseen ei kannata suunnitella valtavaa ja monimutkaista arkkitehtuuria. Tiedon merkityksellisyyden tulisi siis olla järkevässä suhteessa sen hyödyntämiseen vaadittavaan työmäärään ja muihin resursseihin.

5.3.3 Käsitteellinen mallintaminen

Lähtötilanteen arvioinnin ja tietotarpeiden/tavoitteiden määrittelyn jälkeen yrityksen kannattaa luoda kattava käsitteellinen malli saatavilla olevasta datasta ja eri tietojen välisistä suhteista.

Käsittemalli on kaavio, jossa on kuvattuna liiketoimintaan liittyvät entiteetit ja niiden väliset suhteet. Käsitemallin tarkoituksena on ymmärtää ja dokumentoida yrityksen liiketoimintaa datan näkökulmasta. (P. Stiglich) Alla esimerkki osasta yksinkertaista käsittemallia:



Kuva 3. Esimerkki osasta yksinkertaista käsittemallia

Tämä voisi esimerkiksi olla osa huonekaluliikkeen tekemää käsittemallia, jossa kuvataan tyypillistä liiketoiminnan myyntitapahtumaa, siihen liittyviä entiteettejä ja entiteettien välisiä suhteita. Käsittemalliin voi kuvassa olevien entiteettien lisäksi lisätä muun muassa tekstiä kuvaamaan entiteettien välisiä suhteita tai kirjata entiteetteihin liittyviä attribuutteja. Näiden lisäksi käsittemalliin voidaan sisällyttää myös eri tietolähteet, joista käsittemallin eri osat ovat saatavilla. Käsittemalli voidaan rakentaa joko kaiken olemassa olevan tiedon perusteella, tai sitä voidaan päivittää sen mukaan, kun eri tietolähteitä ja niiden sisältämää tietoa sisällytetään osaksi arkkitehtuuria.

Käsittemallin avulla saadaan teknisestä toteutuksesta riippumaton kuvaus siitä, miten yrityksen eri tietoja tulisi yhdistää toisiinsa. Käsitteellisen mallintamisen pohjalta pystytään aloittamaan datan mallintamisen tekninen toteutus hyödyntämällä järjestelmien välisiä integraatioita, tiedon keskitettyä varastointia ja erilaisia datan mallinnusmenetelmiä.

5.3.4 Tekniset ratkaisut

Tyypillinen tekninen ratkaisu datan keskitettyyn hallintaan on koko yrityksen laajuinen tietovarasto (EDW, Enterprise Data Warehouse). Yrityksen tietovarasto on tietokanta tai kokoelma tietokantoja, joka keskittää yrityksen eri tietolähteistä saatavan datan ja jakaa sen lopulta takaisin yrityksen käyttöön (Precisely). Tietovarasto toimii siis keskitettynä järjestelmänä, johon tietoa kerätään, muokataan, analysoidaan ja jaetaan lopulta eteenpäin yrityksen päätöksentekijöille. Tietovaraston sisäiselle tiedon mallinnukselle ja arkkitehtuurille on olemassa useita erilaisia malleja. Näitä ovat esimerkiksi Data Vault -mallinnus, Kimballin kahden kerroksen arkkitehtuuri ja Inmonin kolmen kerroksen arkkitehtuuri (Linstedt & Olschimke 2016, s.11-13). Käsitellään seuraavaksi tietovaraston sisäistä arkkitehtuuria ja teknistä toteutusta yleisellä Kimballin kahden kerroksen arkkitehtuurin menetelmällä.

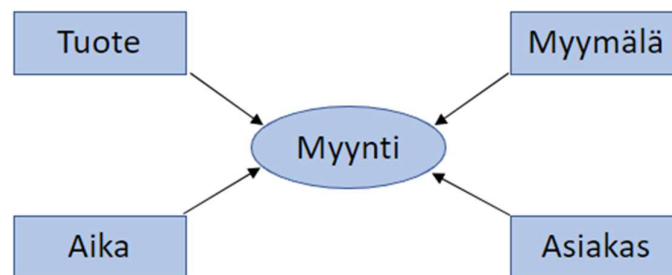
Kimballin esittelemä kahden kerroksen arkkitehtuuri koostuu nimensä mukaisesti kahdesta eri kerroksesta. Ensimmäinen kerros sisältää eri tietolähteistä haetun raakadatan sellaisenaan. Tämän kerroksen tarkoituksena on vähentää tietolähteisiin kohdistuneiden hakuoperaatioiden määrää ja aikaa, joka datan hakemiseen kuluu. Mikäli tietolähteisiin tehtävät hakuoperaatiot ovat monimutkaisia, niin ne kuormittavat tietolähteinä olevia järjestelmiä. Tämä ei ole toivottua, joten hakuoperaatiot pyritään pitämään yksinkertaisina ja tarvittavat muokkaukset dataan pyritään tekemään vasta tietovaraston sisällä. Ensimmäinen kerros tarvitaan myös, mikäli tietolähteistä haettava data on saatavilla eri aikoina. Lataukset tietovaraston ensimmäiseen kerrokseen voidaan tällöin toteuttaa eri aikoina ja kun kaikki tarvittava tieto on saatavilla, niin tietovaraston jälkimmäinen kerros voidaan päivittää. (Linstedt & Olschimke 2016, s.12)

Data haetaan eri tietolähteistä ohjelmallisesti erilaisten ETL (extract, transform and load) työkalujen avulla. ETL-työkalut on kehitetty lataamaan ja muokkaamaan dataa eri tietolähteistä ja tallentamaan sen lopulta haluttuun kohteeseen (Linstedt & Olschimke 2016, s.3). ETL-työkalut voivat siis esimerkiksi tehdä ohjelmallisesti kyselyjä lähdejärjestelmän tarjoamaan rajapintaan ja tallentaa rajapinnasta saatua dataa tietovarastossa olevaan tietokantatauluun. Dataa voitaisiin osana ETL-prosessia myös muokata ennen sen tallentamista kohdejärjestelmään. Haetusta datasta voitaisiin esimerkiksi tässä vaiheessa poistaa henkilötietoja tai dataa voitaisiin aggregoida halutulla tavalla. Kimballin kahden kerroksen arkkitehtuurissa datan haluttu tallennuskohde on arkkitehtuurin ensimmäinen kerros, johon data haetaan tietolähteistä sellaisenaan. Kun kaikki halutut tiedot ovat saatavilla arkkitehtuurin ensimmäisessä kerroksessa, niin data voidaan siirtää arkkitehtuurin jälkimmäiseen kerrokseen.

Siirrettäessä dataa arkkitehtuurin jälkimmäiseen kerrokseen, se mallinnetaan dimensionaalisesti pienemmiksi osajoukoiksi, joita kutsutaan Datamarteiksi. Nämä

Datamartit kuvastavat yrityksen eri liiketoimintaprosesseja. (Linstedt & Olschimke 2016, s.12). Dimensionaalinen mallinnus tarkoittaa sitä, että data jaetaan faktoihin, jotka kuvastavat erilaisia liiketoimintatapahtumia ja dimensioihin, jotka antavat näille tapahtumille kontekstin (Linstedt & Olschimke 2016, s.171-174). Esimerkiksi huonekaluliikkeen myyntitapahtumassa faktana voi olla myynnistä saatava rahasumma. Tähän rahasummaan liittyviä dimensioita ovat esimerkiksi myyty tuote, ajanhetki, myymälä ja asiakas. Dimensionaalista mallinnusta tehtäessä aiemmin käsitelty käsitelmä on todella hyvä apuväline. Käsitelmän avulla liiketoimintaprosessit on datan näkökulmasta helppo tunnistaa ja näiden perusteella mallintaa data dimensionaalisesti. Dimensionaalinen mallinnus on yleinen standardi datan mallintamisessa (Linstedt & Olschimke 2016, s.12).

Dataa halutaan yrityksissä usein tarkastella tietyistä näkökulmista osana tiettyä liiketoimintaprosessia. Datamartit ovat dimensionaalisesti mallinnettuja osajoukkoja yrityksen datasta, joita rakennetaan tarpeen mukaan kuvaamaan haluttua liiketoimintaprosessia. Relationaalisessa tietokannassa dimensionaalista mallinnusta kutsutaan tähtiskeemaksi (Linstedt & Olschimke 2016, s.172). Alla esimerkki aiemmin mainitusta huonekaluliikkeen myyntitapahtumasta kuvattuna tähtiskeemassa.



Kuva 4. Esimerkki tähtiskeeman mukaisesta mallinnuksesta

Tähtiskeemassa jokainen dimensio ja fakta ovat omia tietokantataulujaan. Faktataulussa on myyntitapahtuman rahasumma ja avainkentät yhdistämään tämä fakta sitä kuvaaviin dimensioihin. Tämänkaltaiselle tietomallille on helppo tehdä aggregoituja laskelmia eri dimensioiden suhteen ja se sopii tästä syystä hyvin datan mallinnustavaksi (Linstedt & Olschimke 2016, s.172). Datamarteissa oleva data on valmis loppukäyttäjien hyödynnettäväksi.

Tähtimalli sopii hyvin liiketoimintaprosessien tarkasteluun erilaisista näkökulmista ja siksi se sopii tietomallina hyvin strategisen päätöksenteon tueksi. Esimerkiksi huonekaluliikkeen myyntitapahtumia voidaan analysoida helposti eri ajankohtina,

asiakassegmenteittäin tai tuotekohtaisesti. Näille dimensioille voidaan kokeilla erilaisia yhdistelmiä ja tunnistaa myyntitapahtumien seasta yrityksen kannalta hyödyllistä tietoa. Tällaista tietoa voi olla esimerkiksi tietyn asiakassegmentin ostokäyttäytyminen.

Edellä mainitut vaiheet datan käsittelyssä toteutetaan lähes aina automaattisesti ajastamalla ETL-työkalut toteuttamaan kaikki datan käsittelyyn liittyvät vaiheet esimerkiksi kerran päivässä. ETL-prosessien yksityiskohtaiset toteutukset vaihtelevat kuitenkin todella paljon. Dataan päivitystiheyteen ja -tapaan vaikuttavat esimerkiksi yrityksen henkilöstön tietotarpeet, kyseessä olevan lähdejärjestelmän ominaisuudet ja ETL-prosesseista aiheutuvat kustannukset. (Kimball & Ross 2013, Luku 19) Yrityksen henkilöstöllä voi olla tarve saada ajankohtaista tietoa käytettäväkseen tiettyinä aikoina ja automaattiset ajastukset datan päivittämiselle tulisi tukea liiketoimintatarpeita. ETL-työkalujen tekemät kyselyt kuormittavat kuitenkin lähdejärjestelmiä ja voivat huonossa tapauksessa vaikuttaa näiden järjestelmien operatiiviseen käyttöön. Usein tehtävät päivitykset myös aiheuttavat yritykselle lisäkustannuksia. Näiden asioiden välille tulisikin löytää sopiva tasapaino, joka voi vaihdella yritysten ja käyttötapauksen välillä todella paljon.

Tässä kappaleessa kuvattu tietovarastoinnin hyödyntäminen on todella yleinen tapa käsitellä työssä aiemmin käsiteltyä strukturoitua dataa. Tämän lisäksi yrityksissä voi kuitenkin olla esimerkiksi myös aiemmin mainittua strukturoimatonta dataa tai dataa, jota halutaan saada hyödynnettäväksi reaaliaikaisesti. Näiden käsittelyssä voidaan osittain hyödyntää myös perinteistä tietovarastoa, mutta usein näiden käsittely vaatii tietovaraston lisäksi muita järjestelmiä kuten tietoaaltaita tai reaaliaikaisten datavirtojen käsittelyyn kehitettyjä ohjelmistoja. Näitä ei kuitenkaan tässä työssä tarkastella sen tarkemmin.

Koko yrityksen kattavaa tietovarastoa voidaan pitää hyvin keskeisenä osana yrityksen data-arkkitehtuuria erityisesti strategisessa päätöksenteossa. Tietovarasto ja sen ympärille kehitetyt järjestelmät ja työkalut keräävät, muokkaavat, yhdistävät ja jakavat yrityksessä olevaa tietoa päätöksentekijöiden tueksi. Seuraavassa alaluvussa käsitellään erilaisia tapoja, joilla dataa voidaan jakaa ja hyödyntää tietovaraston avulla.

5.3.5 Tiedon hyödyntäminen

Kun tekninen arkkitehtuuri on saatu toteutettua, niin lopuksi on aika rakentaa työkalut datan loppukäyttäjille. Datat loppukäyttäjillä tarkoitetaan yrityksen henkilöstöä, joka pyrkii käyttämään saatavilla olevaa dataa päätöksenteon tukena. Dataa voidaan

hyödyntää monella eri tavalla ja eri sidosryhmät saattavat tarvita dataa hyvin erilaisissa muodoissa hyödyntääkseen sitä tehokkaasti.

Tyypillinen käyttötarve yrityksen hallussa olevalle datalle on toistuva määrämuotoinen raportointi. Määrämuotoinen raportointi pysyy nimensä mukaisesti rakenteeltaan vakiona ja ainoastaan raporttien sisältämä data päivittyy uuden datan ollessa saatavilla. Tämänkaltaisessa raportoinnissa halutaan usein näyttää yritykselle tai jollekin sen toiminnolle tärkeitä suorituskyvyn mittareita ja näiden kehitystä ajan suhteen. Suorituskyvyn mittareita voivat olla esimerkiksi henkilöstön tyytyväisyys, varaston kiertonopeus ja liikevaihdon kasvu. Määrämuotoisten raporttien avulla päätöksentekijöillä on aina saatavillaan päivitetty tieto tärkeimmistä tunnusluvuista ja tietoa pystytään helposti hyödyntämään suunniteltaessa erilaisia toimenpiteitä. Toistuva määrämuotoinen raportointi on helppo automatisoida, koska se on rakenteeltaan muuttumatonta. Raportin tietojen päivittäminen pystytään toteuttamaan automatisoidusti yrityksen tietovarastosta ja raportin ylläpitämiseen ei mahdollisten korjausten ja jatkokehityksen lisäksi tarvita välttämättä muuta manuaalista työtä. Erilaista tietoa pystytään tietovaraston avulla helposti yhdistelemään ja tuottamaan päätöksentekijöille heidän kaipaamaansa tietoa. Markkinoilla on olemassa monia tämänkaltaiseen raportointiin kehitettyjä työkaluja, joiden käyttö yleistyy yrityksissä todella nopeasti.

Toinen tyypillinen käyttötarve datalle on sen saaminen data-analytiikkaan ja datatieteeseen erikoistuneiden työntekijöiden käyttöön. Data-analytiikassa ja datatieteessä ei keskitytä määrämuotoiseen raportointiin, vaan pyritään tutkimaan dataa ja löytämään sen seasta asioita, joilla yrityksen toimintaa pystytään kehittämään. Datan tutkimisessa voidaan hyödyntää esimerkiksi erilaisia tilastollisia menetelmiä, todennäköisyyslaskentaa, koneoppimista ja tekoälyä. Esimerkiksi huonekaluliikke voisi tutkia asiakas- ja myyntidatasta, kuinka eri asiakassegmenttien ostokäyttäytyminen vaihtelee. Tämän tiedon avulla asiakkaille pystytään esimerkiksi kohdentaa paremmin yrityksen tuotteiden mainontaa. Datan tutkiminen on usein kokeellista ja näihin kokeiluihin vaadittavat tietotarpeet saattavat vaihdella hyvinkin paljon. Määrämuotoiset raportit eivät todennäköisesti riitä kattamaan muuttuvia tietotarpeita. Tämänkaltaiselle käytölle sopiva ratkaisu arkkitehtuurin näkökulmasta on toteuttaa käyttäjille ohjelmallinen pääsy yrityksen tietovarastoon. Tällöin datan tutkimiseen erikoistunut henkilöstö voi omien tarpeidensa mukaan hakea tietovarastosta tarvitsemaansa dataa. Data-analytiikan ja datatieteen lisäksi ohjelmallinen pääsy tietovarastoon voi olla sopiva ratkaisu myös kertaluontoisten raporttien toteutuksessa. Datan tutkimiseen erikoistunut henkilöstö on usein myös osaavaa, joten pääsy suoraan tietovarastoon tai sen tarjoamaan rajapintaan voidaan todeta turvalliseksi.

Kolmantena esimerkkinä datan mahdollisista käyttötarpeista ovat muut yrityksen tietojärjestelmät. Tietovarastosta voidaan haluta tuoda dataa joidenkin operatiivisten järjestelmien käyttöön, jossa dataa hyödynnetään osana päivittäistä toimintaa ja päätöksentekoa. Tietovaraston integroiminen yrityksen muihin tietojärjestelmiin voidaan toteuttaa tarjoamalla integroitavalla tietojärjestelmälle ohjelmallinen pääsy tiettyyn osaan tietovarastoa, josta järjestelmä pystyy hakemaan tarvittavat tiedot. Yrityksessä voi olla käytössä operatiivisten järjestelmien lisäksi myös erilaisia analytiikkaan liittyviä järjestelmiä, joihin integraatioita voidaan myös toteuttaa. Näissä tapauksissa on hyvä kuitenkin muistaa yleiset data-arkkitehtuurin periaatteet ja pyrkiä noudattamaan yksinkertaista ja suoraviivaista arkkitehtuuria. Hajautettu tiedon käsittely ja integraatioiden suuri määrä voi johtaa monimutkaiseen arkkitehtuuriin ja moniin sen mukanaan tuomiin haasteisiin.

6. JOHTOPÄÄTÖKSET

Yritysten käytettävissä olevan tiedon määrän jatkuva kasvu luo yrityksille mahdollisuuksia hyödyntää tätä tietoa aiempaa enemmän päätöksenteon tukena. Digitalisaatio ja teknologian kehitys ovat tehneet tämän tiedon automaattisesta käsittelystä merkittävästi helpompaa ja kustannustehokkaampaa. Onnistuakseen tiedon hyödyntämisessä päätöksenteossa yrityksillä tulee olla toimiva data-arkkitehtuuri kaiken tämän tiedon hallinnalle. Operatiivisessa päätöksenteossa päätöksentekoa tukevat yleensä lähinnä operatiiviset tietojärjestelmät, mutta strategisessa päätöksenteossa yritys tarvitsee usein erillisen päätöksentekoa tukevan digitaalisen infrastruktuurin.

Strategisessa päätöksenteossa päätösten vaikutukset ovat yleensä yrityksen kannalta merkittäviä ja pitkäaikaisia. Strategisiin päätöksiin voidaan täten perustellusti käyttää enemmän aikaa ja resursseja, kuin operatiiviseen päätöksentekoon. Strategisissa päätöksissä hyödynnetty tieto on usein kerätty monesta eri lähteestä, tiedon käsittelyssä on hyödynnetty tilastollisia menetelmiä, tietoa on kerätty pitkältä ajalta ja tietoa on pyritty yhdistelemään ja aggregoimaan. Strategisen päätöksenteon tukena toimiva data-arkkitehtuuri pyrkii helpottamaan tämänkaltaisen tiedon hallintaa automatisoimalla ja digitalisoimalla datan keräämistä, auditointia, yhdistämistä, analysointia ja jakamista organisaation sisällä.

Toimiva data-arkkitehtuuri strategisen päätöksenteon tukena riippuu merkittävästi kyseessä olevasta yrityksestä. On olemassa kuitenkin yleisiä periaatteita, joita data-arkkitehtuurin kehityksessä tulisi ottaa huomioon. Näitä ovat esimerkiksi: datan laadun varmistaminen, datan turvaaminen, datan keräämiseen liittyvien säännösten noudattaminen ja master datan hallinta. Yksittäistä teknistä ratkaisua näiden periaatteiden toteuttamiseen on mahdotonta kertoa, koska yritykset ovat hyvin yksilöllisiä. Tyypillinen arkkitehtuuri sisältää kuitenkin keskitetyn paikan, johon dataa kerätään, missä sitä analysoidaan ja lopulta jaetaan yrityksen eri päätöksentekijöille. Erityisesti strategisessa päätöksenteossa tämänkaltaisen arkkitehtuuri on hyvin tyypillinen ja sen hyödyt ovat helposti ymmärrettävissä. Myös yrityksen päätöksistä hyödyntää eri tietotyyppisiä ja tietolähteitä on olemassa melko yleispäteviä ohjeita. Eri tietotyyppien ja tietolähteiden hyödyntämisessä voidaan pitää hyvänä lähtökohdana aloittaa yrityksen sisäisistä tietolähteistä ja niistä saatavasta strukturoidusta datasta. Sisäiset tietolähteet sisältävät yrityksen kannalta lähes täysin relevanttia tietoa ja esimerkiksi tiedon laatu ja alkuperä ovat yrityksellä hyvin hallinnassa. Strukturoitua dataa löytyy tavallisesti paljon yrityksen tietojärjestelmistä ja se on luonteeltaan helposti

käsiteltävää. Näiden ominaisuuksien lisäksi yrityksen sisäisten järjestelmien tallentama data on usein todella korkean arvon dataa. Sisäiset tietolähteet ja strukturoitu data luovat hyvän lähtökohdan tietolähteiden ja -järjestelmien näkökulmasta yrityksen data-arkkitehtuurille.

Yleisten periaatteiden lisäksi myös data-arkkitehtuurin kehitysprosessin voidaan katsoa olevan melko yleisesti hyödynnettävissä. Kehitysprosessissa pyritään ensin selvittämään data-arkkitehtuurin nykytila. Tämän jälkeen määritetään yrityksen päätöstentekijöiden tietotarpeet ja mahdolliset tavoitteet, sekä miten ja minkälaista dataa tulevaisuudessa haluttaisiin hyödyntää. Kokonaiskuvan saamiseksi ja dokumentaation tueksi kannattaa tehdä myös käsitemallinnusta yrityksen käytettävissä olevasta datasta ja sen välisistä suhteista. Käsitemalli auttaa merkittävästi myös arkkitehtuurin teknisessä toteutuksessa. Lopuksi kehitysprosessissa luodaan arkkitehtuurin tekninen toteutus. Prosessin vaiheet ovat usein kuitenkin päällekkäisiä ja niitä toteutetaan käytännössä ainakin osittain samanaikaisesti.

Teknisen toteutuksen alaluvussa kuvataan strategisen päätöksenteon tukena toimivaa tyypillistä keskitettyä arkkitehtuuria. Tällaisessa arkkitehtuurissa yrityksen käytettävissä oleva tieto pyritään erilaisten integraatioiden avulla tuomaan johonkin keskitettyyn järjestelmään. Hyvin tyypillinen tapa on kerätä haluttu tieto yhdestä tai useammasta tietokannasta koostuvaan tietovarastoon. Tietovaraston sisällä dataa yleensä yhdistetään ja muokataan helpommin käytettäväksi. Muokkauksessa datan laatua voidaan parantaa ja dataa usein mallinnetaan esimerkiksi dimensionaalisesti, jotta sen hyödyntäminen olisi loppukäyttäjille helpompaa. Loppukäyttäjille dataa voidaan lopulta jakaa esimerkiksi automaattisesti päivittyvien määrämuotoisten raporttien muodossa tai tarjoamalla ohjelmallinen rajapinta tietovarastossa olevaan dataan. Käyttötapaukset ovat kuitenkin hyvin yksilöllisiä ja niitä tulisi soveltaa yrityksen tietotarpeiden mukaan.

Data-arkkitehtuurissa voidaan siis arvioida olevan yleisiä prosesseja ja periaatteita, jotka auttavat yrityksiä toteuttamaan heille itselleen sopivan arkkitehtuurin strategisen päätöksenteon tueksi. Nämä prosessit ja periaatteet ovat kuitenkin melko yleisluontoisia ja todellisuudessa näiden prosessien ja periaatteiden mahdollisia toteutustapoja on olemassa lukemattomia erilaisia. Tässä työssä annetut esimerkit eivät siis välttämättä sovi sellaisenaan tietylle yritykselle. Yritysten erilaisten ominaisuuksien vaikutus sopivaan data-arkkitehtuuriin olisi hyvä aihe jatkotutkimukselle. Jatkotutkimuksessa voitaisiin käsitellä esimerkiksi yrityksen koon, toimialan ja strategian vaikutusta data-arkkitehtuurin kehittämiseen ja tämän avulla tehtävään strategiseen päätöksentekoon.

LÄHTEET

O.Abahmane, M.Binkkour, Strategic and Operational information support of decision making processes and systems, 2008, (Viitattu 02.12.2020), Saatavissa: <https://arxiv.org/abs/1502.01803>

BI-Survey, Using internal and external data for decision-making, (Viitattu 23.9.2020), Saatavissa: <https://bi-survey.com/data-sources>

Finto, tietotermit, tiedonhallinta, 13.10.2018, (Viitattu 21.9.2020), Saatavissa: <https://finto.fi/tt/fi/page/t7>

W.Gänswein, "Effectiveness of Information Use for Strategic Decision Making." Gabler Verlag, 2011. Web. (Viitattu 28.11), Saatavissa: <https://link-springer-com.lib-proxy.tuni.fi/book/10.1007%2F978-3-8349-6849-4>

S. Heinänen, Enfo: Master data management – laadukas data on liiketoimintasi perusta, 6.8.2020, (Viitattu 15.10.2020), Saatavissa: <https://www.enfo.fi/blogi/master-data-management-laadukas-data-liiketoimintasi-perusta>

W.H Inmon, D. Linstedt, and M. Levins, Data Architecture: A Primer for the Data Scientist. San Diego: Elsevier Science & Technology, 2019. Print. (Viitattu 22.9.2020), Saatavissa: <https://learning.oreilly.com/library/view/data-architecture-a/9780128169179/B9780128169162000012.xhtml>

J-P. Isson, M. Hwang, Unstructured Data Analytics: How to Improve Customer Acquisition, Customer Retention, and Fraud Detection and Prevention. Somerset: John Wiley & Sons, Incorporated, 2018. Print. (Viitattu 23.9.2020), Saatavissa: <https://learning.oreilly.com/library/view/unstructured-data-analytics/9781119129752/c02.xhtml>

R. Kimball, M. Ross. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley-Blackwell, 2013. Print. (Viitattu 26.11.2020), Saatavissa: [https://masterworkshop.skillport.com/skillportfe/assetSummaryPage.action?assetid=RW\\$6787:_ss_book:56391#summary/BOOKS/RW\\$6787:_ss_book:56391](https://masterworkshop.skillport.com/skillportfe/assetSummaryPage.action?assetid=RW$6787:_ss_book:56391#summary/BOOKS/RW$6787:_ss_book:56391)

C. Kimble, Knowledge Management, Codification and Tacit Knowledge, Information research, 28.5.2013, (Viitattu 21.9.2020), Saatavissa: <http://informationr.net/ir/18-2/paper577.html#.X2mMbmgaUm>

H. Laihonen, M. Hannula, N. Helander, I. Ilvonen, J. Jussila, M. Kukko, H. Kärkkäinen, A. Lönnqvist, J. Myllärniemi, S. Pekkola, P. Virtanen, V. Vuori, T. Yliniemi, Tietojohtaminen, 2013, (Viitattu 21.9.2020), Saatavissa: <http://urn.fi/URN:ISBN:978-952-15-3058-6>

D. Linstedt, M. Olschimke, Building a scalable data warehouse with data vault 2.0, Elsevier Inc, 2016

NodeGraph, 5 essential elements of Data Governance 24.6.2020 (Viitattu 15.10.2020), Saatavissa: <https://www.nodegraph.se/5-essential-elements-of-data-governance/>

NodeGraph, What is Data Quality? And why should you be thinking about it? 5.1.2019 (Viitattu 15.10.2020), Saatavissa: <https://www.nodegraph.se/what-is-data-quality/>

C. Pitt, AI in marketing: How to find the right data sources, Econsultancy, 16.9.2019, (Viitattu 23.9.2020), Saatavissa: <https://econsultancy.com/ai-in-marketing-how-to-find-the-right-data-sources/>

Precisely, Enterprise Data Warehouse (EDW), (Viitattu 18.10.2020), Saatavissa: <https://www.precisely.com/glossary/enterprise-data-warehouse>

D. Reinsel, J. Gantz, J. Rydning, The Digitization of the World From Edge to Core, 2018, (Viitattu 20.9.2020), Saatavissa: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

E. Salminen, Practical advice for writing publications, course material, TKT-9617 Scientific Publishing, Tampere University of technology, Nov 2009 (updated Aug 2012), (Viitattu 21.9.2020), 97 p. Saatavissa: http://www.cs.tut.fi/~ege/Misc/salminen_figures_styles_v14.pdf

P. Stiglich, Data Quality Pro: Necessity of Conceptual Data Modeling for Information Quality by Pete Stiglich (Viitattu 14.10.2020), Saatavissa: <https://www.dataqualitypro.com/blog/conceptual-modelling-for-data-quality>

Techopedia, Data Architecture, 26.1.2017, (Viitattu 20.9.2020), Saatavissa: <https://www.techopedia.com/definition/6730/data-architecture>