



**TURUN
YLIOPISTO**
Kauppakorkeakoulu

PILVIPALVELUT JA DATA-ALUSTAT LII- KETOIMINNASSA

Liiketoiminnan ja raportointiprosessin kehittäminen pilvipalveluiden ja data-alustojen avulla

Laskentatoimi ja rahoitus

Laatija:

Atte Sukari

Ohjaaja(t):

KTM Anu Ikonen-Kullberg

4.3.2024

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

Oppiaine: Laskentatoimi ja rahoitus

Tekijä(t): Atte Sukari

Otsikko: Otsikko

Ohjaaja(t): KTM Anu Ikonen-Kullberg

Sivumäärä: 72 sivua + liitteet 1 sivua

Päivämäärä: 4.3.2024

Yritysten raportointikehitys on jatkuvassa muutoksessa, ja pakollisten raporttien lisäksi yritykset pyrkivät hyödyntämään dataansa entistä tehokkaammin. Modernit pilvipalvelut ja data-alustat mahdollistavat yrityksille täysin uudenlaisia raportointimahdollisuuksia. Monesti näitä mahdollisuuksia tarkastellaan pelkästään lopputuloksen kautta. Datan hyödyntäminen on kuitenkin prosessi, joka vaatii uudenlaista ajattelua organisaation sisällä. Yritysten tulisi siirtyä enemmän prosessimaiseen ajatteluun ja kohti datan tuotteistamista.

Tämän tutkielman tavoitteena on selvittää, mitä pilvipalvelut ja data-alustat mahdollistavat johdon raportoinnille, mitä otetaan huomioon, kun suunnitellaan analytiikka-alustoja liiketoimintakäyttäjille, ja millainen prosessi raportoinnissa tulisi olla, kun mukana on moderneja analytiikka-alustoja. Tutkielman aikana käydään läpi datakehityksen prosessia ja sitä, mitä se vaatii liiketoiminnalta. Aihetta käsitellään liiketoiminnan näkökulman lisäksi myös teknisestä perspektiivistä. Tällä pyritään muodostamaan holistinen kuva siitä, mitä eri asioita yrityksen raportointiprosessiin ja analytiikkaan liittyy. Tutkimuksen lähestymistapa on kvalitatiivinen. Teoriaosuus perustuu tutkimuksiin perustuvaan kirjallisuuskatsaukseen. Empiirinen osuus muodostuu neljän erillisen yksilöhaastattelun kautta. Haastateltavilla on laaja työkokemus erilaisista rooleista datan ja raportoinnin kanssa toimimisen parissa.

Tutkimuksen tuloksista käy ilmi, että liiketoimintakäyttäjien osallistaminen raportointiin on haastavaa, vaikka käytössä olisi moderneja työkaluja ja menetelmiä. Liiketoimintakäyttäjien vähäinen osallistuminen heijastuu myös siihen, että datan validointiin käytetään yhä edelleen liian vähän resursseja, mikä voi heikentää raporttien luotettavuutta. Yksi keskeisimmistä löydöksistä liittyy siihen, että tärkeä ominaisuus, jota modernista analytiikasta ja raportoinnista haetaan, on reaaliaikaisuus. Modernit järjestelmät toimivat hyvin reaaliaikaisuuden kanssa, eikä isoja datamassoja nähdä haasteeksi. Enemmän haasteena nähtiin datan hyödyntäminen ja monimuotoisuus, kuten kuvat ja äänitiedostot.

Avainsanat: data, raportointi, analytiikka

SISÄLLYS

| | | |
|----------|--|-----------|
| 1 | Johdanto | 7 |
| | 1.1 Tavoite ja rajaus | 9 |
| 2 | Yritysten raportoinnin kehittyminen | 11 |
| 3 | Pilvipalveluiden perusteet ja niiden yhteys liiketoimintaan | 15 |
| | 3.1 PaaS ja roolitukset | 16 |
| | 3.2 Analytiikka-alustan määrittely ja implementointi | 20 |
| 4 | BI-kehitys organisaatiotasolla | 22 |
| | 4.1 Agile | 23 |
| | 4.2 Data mesh | 24 |
| 5 | Dataputket ja transformointi | 27 |
| 6 | Tiedon varastointi | 30 |
| | 6.1 Relaatiotietokannat | 31 |
| 7 | Visualisointi | 34 |
| | 7.1 Monitorointi | 37 |
| 8 | Empiria | 38 |
| | 8.1 Raportointiprosessi | 40 |
| | 8.2 Raportointiprosessin muutos | 43 |
| | 8.3 Modernit työkalut ja niiden vaikutus | 46 |
| | 8.4 Tiedon validointi | 49 |
| | 8.5 Tietoturva | 52 |
| | 8.6 Liiketoimintakäyttäjät | 55 |
| 9 | Johtopäätökset | 60 |
| | Lähteet | 69 |
| | Liitteet | 73 |

KUVIOT

| | |
|--|----|
| Kuvio 1 Azuren roolituksen arkkitehtuuri (Microsoft, 2024) | 17 |
| Kuvio 2 Data-alustan arkkitehtuuri mukailleen (Fugini ym. 2021) | 19 |
| Kuvio 3 DASE-arkkitehtuuri (Bandara ym. 2023, 195–212) | 20 |
| Kuvio 4 Data mesh (Montecarlodata, 2020) | 25 |
| Kuvio 5 Json-esimerkkidataa, joka sisältää yhden työntekijän tiedot | 28 |
| Kuvio 6 Json-data mallinnettuna tauluksi tietokannassa | 29 |
| Kuvio 7 Yksinkertainen relaatiotietokannan malli kahdesta taulusta: Customer ja Orders. 32 | |
| Kuvio 8 Esimerkki PowerBI-raportista | 34 |
| Kuvio 9 Tietokantataulu ilman visualisointia | 35 |
| Kuvio 10 Tietokantataulu visualisoituna | 36 |
| Kuvio 11 Esimerkki monitorointidashboardista | 37 |

TAULUKOT

| | |
|--|----|
| Taulukko 1 Haastateltavien tittelit ja työnantajat julkinen/yksityinen | 40 |
|--|----|

1 Johdanto

Raportointiprosessi on aihe, joka on jokaisen yrityksen tai organisaation agendalla jossain vaiheessa sen elinkaarta. Yrityksillä on usein myös lain ja säädösten velvoittama tarve raportoida esimerkiksi sen talouslukuja tai muita merkittäviä tapahtumia. Usein raportointia käsitellään kuitenkin melko pintapuoleisesti ja syvempi fokus on usein keskittynyt raportin loppusisältöön, eikä niinkään raportin rakentamisen koko elinkaareen. (McAfee & Brynjolfsson, 2012.)

Ongelma tämänkaltaisessa ajattelussa on se, että herkästi liiketoiminnan perspektiivi tulee mukaan aivan raportin alkupuolella, tai siinä kohtaa, kun tieto on kerätty, jalostettu ja on aika visualisoida tieto raportille. Näissä tapauksissa raportti ei välttämättä ole niin luotettava, kuin siinä tapauksessa, mikäli liiketoiminta olisi osallistutettu kaikkiin raportin rakentamisen osiin. Tiedon hyödyntäminen yleensä heikentyy, mikäli henkilöstöä ei osallisteta tarpeeksi tiedonkeruun ja jalostamisen prosessiin eli raportointiprosessiin. Tutkimukset yleensä keskittyvät, joko tekniseen näkökulmaan tai lähinnä loppusisältöön. Raportointi tulisi nähdä holistisena prosessina, jossa raportoinnin lopullinen kasaaminen ja visualisointi on vain yksi osa. (Gour, 2019.)

Forbesin artikkelin mukaan Yhdysvalloissa menetetään, joka vuosi arviolta 1,3 triljoonaa dollaria huonon datan vuoksi. Kyselyyn vastanneista yrityksistä 60% ei tiennyt paljonko heikko datan laatu todellisuudessa maksaa yrityksille. (Bansal, 2021.). Tämä kertoo karua kieltä siitä, että vaikka raportointi on yleistä ja jossain määrin trendikästä, ei datan laatuun, keräämiseen ja prosessointiin kiinnitetä läheskään samaa määrää huomiota. Dataa saatetaan myös kerätä, mutta sen validointiin tai laadunvalvontaan ei kiinnitetä huomiota. Tämänkaltaisesta datasta tehdyt raportit voivat olla yritykselle turhia, tai pahimmillaan vaarallisia. Yhteiskunnallisesta näkökulmasta tämä on todella pelottavaa. Raportointitarpeet ovat jatkuvassa kasvussa ja esimerkiksi ESG-raportointia povataan tulevan pakolliseksi entistä pienemmille yrityksille. Onko kuitenkaan kannattavaa tai mieluisaa luoda raporttia, mikäli sen taustalla olevaa dataa ei validoita tai siihen ei kiinnitetä huomiota? Kyseessä ei välttämättä ole siis tilanne, jossa yritys tarkoituksella tarjoaisi vääriä lukuja, vaan pikemminkin se, että yritys ei tiedä minkä laatuista dataa se tarjoaa. PwC (2022.) kertoo, että lukujen vääristeleminen omaksi edukseen saattaa tuoda yritykselle merkittävää mainehaittaa ja, että datan laatu on varmistettava. (PwC, 2022.)

Datan laadun varmistamisen lisäksi ja mahdollisten pakollisten raporttien, kuten ESG määrän kasvamisen lisäksi, myös tietoa on kasvavissa määrin saatavilla. Näin ollen myös raportointitarpeet kehittyvät ja kasvavat jatkuvasti. Aiemmin raportiksi on ehkä riittänyt muutama graafi jostain menneestä myynnin kehityksestä esimerkiksi kuluneella viikolla. Modernit analytiikka-alustat mahdollistavat johdolle myös uudenlaisen tarkkuuden ja perspektiivin raportointiin. Tällaisia ovat esimerkiksi isojen datamäärien hyödyntäminen tai reaaliaikainen raportointi. Aiemmin myös raporttien muuttaminen oli usein todella työlästä ja vaati paljon resursseja. Ylipäätään näiden raporttien rakentaminen oli todella monimutkaista ja työlästä. Nykyaikaiset alustat ovat osana mahdollistamassa kustannustehokasta, joustavaa ja monipuolista raportointia, joka kattaa muuttuvat tarpeet. (Patel, 2020.)

Muuttuvat tarpeet voivat kattaa myös täysin uusia raportointeja. Yksi esimerkki on Digian (2023) luoma tekoäly, joka pyrkii ennustamaan jääkiekon SM-liigan voittajan kauden alussa. Tämä tekoäly hyödyntää modernia data-alustaa ja pyrkii raportoimaan todennäköisyyksiä sille, mikä joukkue tulee voittamaan SM-liigan. (Digia, 2023.). Samankaltaista kehittynyttä analytiikkaa ja raportointia nähdään myös Yhdysvalloissa NHL:ssä. Amazon Web Serviceä hyödyntävä koneoppimismalli pyrkii raportoimaan eri pelitilanteista, miten hyvä kyseinen tilanne oli kokonaisuutena ja mikä sen maaliollettama oli. Tämänkaltaiset raportointimahdollisuudet rikkovat kokonaan sen käsityksen, että raportoinnin tulisi keskittyä talouslukuihin maksimissaan vuoden ajalta. Sen sijaan ajatus siitä, mitä ylipäätään voidaan raportoida, muuttuu radikaalisti. Yritykset voisivat esimerkiksi tarkastella maaliollettaman tavoin jokaista myyjän transaktiota ja laskea odotusarvon mahdolliselle ostokselle, tai oikealle tuotteelle. (ks. Amazon, 2023.)

Toisaalta kun tämänkaltaisia raportteja usein tarkastellaan ajatellen pelkästään lopputuotetta, eikä niinkään matkaa siihen. Sama ajatus pätee tekoälyn tarjoamiin trendeihin ja mahdollisuuksiin. Tekoälyä hyödynnetään, mutta usein sen kokonaisvaltainen toteutus jää hämärän peittoon. Artikkeleissa usein puhutaan mahdollisuuksista ja haasteista, joita tekoälyn implementointi tuo tullessaan. Tietoturvaa tuodaan myös esille, mutta usein senkin näkökulma jää liiketoiminnan kannalta epäselväksi. Saako tekoälyä käyttää tiedostojen tarkasteluun vaiko ei. Tekoälyn suosio on kasvanut valtavasti ChatGPT:n ja muiden isojen kielimallien myötä. Suurin osa hyödyistä tulee asiakassuhteiden hallinnan kanssa ja tarkemmin sanottuna viestinnästä. Haittapuolena sen sijaan nähdään asioita kuten:

- Riippuvuus teknologiasta
- Nykyinen osaaminen ja toisaalta sen mahdollinen heikentyminen
- Misinformaatio
- Tekoälyn harhat
- Negatiivinen vaikutus asiakassuhteisiin

Huomioitavaa tässä on se, että tekoälyn harhat tai puolueellisuus on melko alhaalla listalla verrattuna teknologiseen riippuvuuteen. (Forbes, 2021.)

Yksi pohdittava asia tähän on se, että miten riippuvaisia yritykset ovat jo nykyisestä teknologiasta. ERP:t, raportointijärjestelmät, sähköpostit, raportoinnin esitystyökalut ja muut yritysmaailmassa keskiössä olevat ratkaisut. Esimerkiksi Microsoft määrittelee, että sen palveluissa käytettävät tekoälymallit ovat paitsi lakien ja säädösten mukaisia, niin yritysten data pysyy vain yrityksille nähtävinä. Tämä yhdistettynä siihen, että yritykset jo käyttävät näiden isojen palveluntarjoajien osaamista ja kapasiteettia ovat iso syy siihen, miksi tietoturvaongelmista puhuminen tekoälyn kohdalla on osittain jopa ylikorostunut. Tämä ei tarkoita, että tietoturva olisi turha aspekti, mutta enemmänkin perspektiivi ja perustus yrityksen it-infrastruktuurissa. (Microsoft, 2023.)

1.1 Tavoite ja rajaus

Tutkielman tavoite on selvittää mitä modernit pilvipalvelut ja analytiikka-alustat mahdollistavat johdon raportoinnille ja miten näitä uusia mahdollisuuksia voidaan hyödyntää päätöksenteossa. Lähtökohtana tutkimukselle on se, että kysymyksiä pohditaan liiketoiminnan kannalta teknisestä näkökulmasta. Mitä nämä palvelut käytännössä mahdollistavat ja mitä ne vaativat. Pelkät mahdollisuudet eivät siis riitä, vaan tarkastelussa on myös vaatimukset ja käyttötapaukset. Tutkielman aikana käsitellään myös hyvää arkkitehtuuria ja millä tavalla yritysten kannattaisi lähteä hyödyntämään näitä palveluita. Tätä tarkastellaan paitsi teorian, mutta myös asiantuntijahaastatteluiden kautta. Tämänkaltaisia tutkimuksia samankaltaisella näkökulmalla on verrattain vähän. Saatuja tuloksia voidaan näin ollen soveltaa laajasti liiketoiminnan tai teknisten

toteutusten kautta. Pysin tutkimuksessa tarjoamaan teknistä näkökulmaa ja sen tuomia mahdollisuuksia liiketoimintaan. Tutkimuskysymykset ovat:

- Mitä pilvipalvelut ja data alustat mahdollistavat johdon raportoinnille?
- Mitä otetaan huomioon, kun analytiikka-alustoja suunnitellaan liiketoimintakäyttäjille?
- Millainen prosessi raportoinnissa tulisi olla, kun mukana on moderneja analytiikka-alustoja?

Työn edetessä aihetta käsitellään ensin teorian kautta. Teorian sisältö tarkastelee pilvipalveluita ja analytiikka-alustoja. Pääasiallinen fokus on Microsoftin tarjoamassa ekosysteemissä ja sen osissa, mutta tulokset ovat yleistettävissä myös muihin palveluntarjoajiin. Microsoft on valittu siitä syystä, että fortune 500 yhtiöistä yli 95 prosenttia käyttää Azurea, joka on Microsoftin tuottama ja ylläpitämä pilvipalvelu (Microsoft, 2024.)

2 Yritysten raportoinnin kehittyminen

Pilvipalvelut ja erilaiset analytiikka-alustat ovat kiinteä osa nykypäivän liiketoimintaa. Datalähtöinen ajattelu ohjaa toimintaa ja päätökset pyritään tekemään entistä perustuen faktatietoihin, kuin tunteisiin tai yksinkertaistuksiin. Liiketoiminnan kehittyessä uusia raportointitarpeita tulee entistä enemmän ja näin ollen myös yrityksen teknologiatoimintaan tulee uusia ulottuvuuksia, kuten tietovarastointi, data governance, edistynyt analytiikka ja monitorointi. Toimintojen kasvaessa myös monimutkaisuus lisääntyy ja erilaisten migraatioiden tarve nostaa päätään. Aiemmin riittäneet henkilökohtaiset taulukot, eivät välttämättä palvele yrityksen tarpeita riittävän hyvin. Datan laatu voi olla heikko, laskentateho ei riitä, kyberturvallisuus ja tehokkuus eivät välttämättä ole sillä tasolla, joka vastaisi yrityksen tarpeita. (Bendre ym. 2015,2000–2003.)

Tarkastellessa datan volyymia, jota yrityksellä on käytössä, huomataan, että datan määrä on kasvattanut merkittävästi viimeisessä kymmenessä vuodessa. Terminä käytössä on vakiintunut Big Data. Big datan määritelmänä on yleensä volyymi, datan nopeus, vaihtelevuus, totuudenmukaisuus ja arvo. Etenkin kolme ensimmäistä tarjoavat isoja haasteita yritykselle, jonka datan hyödyntäminen on alkeellisella tasolla. Perinteiset taulukkotyökalut, kuten Excel eivät kykene työstämään suuria datamassoja, puhumattakaan niiden analysoinnista. Exceliä voidaan edelleen monissa sovelluksissa käyttää liiketoimintakäyttäjien eräänlaisena front-endinä tai tarkastelutyökaluna. Tässä tarkoitetaan sitä, että monimutkaiset laskennat ja muut prosessointia vaativat tehtävät, kuten aggregoinnit hoidetaan tietokannoissa ja taulukko-ohjelmien tehtäväksi jää sille soveltuvan datamassan tarkastelu. (Bendre ym. 2015,2000–2003.).

Voluumin lisäksi nopeus tuo ongelmia. Data voidaan tuoda ylätasolla kahdella tavalla. Joko datan lataaminen tehdään erissä (batch), jolloin dataa siirretään esimerkiksi joka maanantai yhden kansion verran. Erissä prosessointi sopii hyvin tapauksiin, jossa reaaliaikaista tietoa ei tarvita ja keskittyminen on enemmän isoissa datamäärissä, joita analysoidaan. Mikäli tarvitaan reaaliaikaista dataa, voidaan hyödyntää striimausta. Datan striimaus tarkoittaa sitä, että data saadaan reaaliaikaisesti ja viive on joko todella pieni, tai sitä ei ole. Vaihtelevuus on yksi big datan ominaisuuksista. Perinteiset taulukkotyökalut kykenevät tarkastelemaan esimerkiksi CSV-tiedostoja tai xml-tiedostoja. Entäpä jos data on kuvia, tai äänitiedostoja? Taulukointityökalut, tai edes

monet perinteiset business intelligence -työkalut, kuten Tableau tai PowerBI pysty itseksensä näitä prosessoimaan. Tällöin voidaan hyödyntää moderneja pilvipalveluita ja erilaisia analytiikka-alustoja. (Berisha ym. 2022, 24–24.)

Näiden palveluiden helppokäyttöisyys ja skaalautuvuus tekee sen, että niiden käyttö ei ole rajattu pienellä käyttäjäkunnalla, joilla on valtavat resurssit. Yritykset voivat aloittaa käytön yleensä todella pienellä budjetilla ja kasvattaa tai laskea sitä tarpeiden mukaan. Yksi iso etu on myös se, että yritysten ei tarvitse itse hallita käytössään olevaa infraa ja laskentatehoa. Valitsemansa pilvipalvelu hoitaa tämän osuuden. Tietoturva, tietoverkot ja muut haasteet ovat helposti hallittavissa. Palveluiden avulla voidaan myös varautua erilaisiin katastrofitilanteisiin, kuten datan korrupioon, tai konesalipaikalliseen katastrofiin, kuten tulvaan tai myrskyyn. Fyysisten palvelimien komponenttien hajoaminen ei myöskään ole ongelma, sillä sekin osuus on pilvitarjoajan vastuulla. Tämä ei kuitenkaan vapauta yritystä kaikesta vastuusta, vaan esimerkiksi tietoturvan kanssa tulee silti olla tarkkana. (Maresova ym. 2017, 521–533.)

Puhuttaessa pilvipalveluiden ja analytiikka-alustojen tarjoamista mahdollisuuksista yritykselle on väistämättä tarkasteltava yritysten valmiuksia tämänkaltaisille palveluille. Modernit palveluntarjoajat pyrkivät helppokäyttöisyyteen, sekä tarjoavat laajat dokumentaatiot siitä, miten tuotteita tulisi käyttää. Usein kuitenkin tämänkaltaisten ympäristöjen luominen, suunnittelu, sekä ylläpito vaatii taitoja, joita etenkin pk-yrityksillä ei itsellään välttämättä ole. Tällöin voi olla mielekästä turvautua ulkopuoliseen apuun esimerkiksi konsultointia tarjoavilta yrityksiltä, tai rekrytoida ulkopuolinen henkilö ylläpitämään tarvetta. Nykyaikaiset alustat tarjoavat helppokäyttöisiä aloitusratkaisuja kaikkensuuruisille yrityksille ja budjeteille. Toisaalta on hyvä huomioda, että mikäli yrityksen sisäinen data on arkaluontoista, voi olla syytä hyödyntää ammattilaisen palveluita ainakin alkuun pääsemiseksi. Pilvipalvelut ja analytiikka-alustat eivät ole immuuneita tietovuodoille, ja vaikka täydellinen järjestelmä kehitettäisiin, on käyttäjä edelleen ketjun heikoin lenkki. (Hasanah, 2021.)

Tietovuotojen ja alkuun pääsemisen lisäksi tulee yrityksen tarkastella omien työntekijöidensä valmiuksia omaksua uuden järjestelmän ominaisuudet. Monesti tämä tarkoittaa jonkinasteisia lisäkoulutuksia ainakin osalle henkilöstöstä. Näiden lisäkoulutusten avulla yritys voi varmistua siitä, että uutta järjestelmää käytetään mahdollisimman tehokkaalla ja oikealla tavalla. Näin saadaan liiketoiminta enemmän

mukaan raportin luomiseen. Mitä helpompi analytiikka-alustaa tai työkalua on käyttää, sitä useampi ihminen yrityksessä sitä hyödyntää. Kyse ei välttämättä ole raa'asta kahtiajaosta esimerkiksi data-tietelijöiden ja johtoportaan välillä, vaan pikemminkin siitä, että mahdollisimman monella työntekijällä olisi mahdollisuus ja insentiivi tutkia yrityksen dataa helposti ja turvallisesti. (Hasanah, 2021.)

Yleensä moderni raportointi vaatii toimiakseen teknisen puolen ja liiketoiminnan välille jonkinlaisen välikappaleen. Usein liiketoiminta-analytytikot tai liiketoimintakehittäjät täyttävät tämän aukon. Näiden ihmisten tehtävänä on vähentää siilautumista eri osastojen välillä ja pyrkiä luomaan dialogi eri tekijöiden välille, jolloin liiketoiminta ymmärtää mahdollisuudet ja rajoitteet, sekä tekninen toteutus ymmärtää tarkoituksen ja lopputuotteen. Usein näiden liiketoiminnan ja teknisen toteutuksen välimaastossa toimivien henkilöiden tehtävänä on toimia tulkkina kahden maailman välillä. Tämä vaatii työntekijältä, sekä teknistä osaamista, liiketoimintanäkemyistä ja substanssiosaamista siitä, mitä sillä hetkellä ollaan tekemässä. (Hasanah, 2021.)

Perinteisiä raportteja ovat monet taloudelliset raportoinnit, kuten tuottojen ja kulujen raportointia, tai maksuvalmiuden eri osia. Talouslukujen raportointi tarkasti vaikuttaa merkittävästi yrityksen toimintaan. Mikäli esimerkiksi maksuvalmiuden luvut eivät pidä paikkaansa, voi se olla yritykselle kohtalokasta. Toisaalta hyvin tehdyn taloudellisen raportoinnin on nähty olevan korrelaatiossa yrityksen menestykseen (ks. Mirouniuc ym. 2015.)

Mitä sitten on hyvä raportointi ja minkälaisia piirteitä siinä on? Hyvälle raportille on asetettu laajasti erilaisia periaatteita, joita se noudattaa. Raportin on oltava esimerkiksi relevantti, tarkka ja sen on vastattava oikeaan kysymykseen. Paperilla nämä kuulostavat helpolta, mutta mitä tapahtuu, kun yrityksen datamäärä kasvaa alati ja ainoa tapa raportoida on taulukko-ohjelmistot, joissa jokainen luo itse raporttinsa? Yrityksen tiedon luotettavuus ei tarkoita pelkästään sitä, että yritys ei pyri aktiivisesti väärentämään dataansa, vaan myös sitä, että johdolle esitettävät raportit ovat oikeita. Näin ollen johdon raportointiin ja niiden luontiin tulisi kiinnittää yhtä iso huomio, kuin finanssipuolen raportteihin. (Forbes, 2023.)

Raporttien lisäksi analytiikka-alustat mahdollistavat erilaisten visualisointien käytön. Näiden avulla yritykset voivat reaaliaikaisesti seurata liiketoimintansa ydinkysymyksiä ja erilaisia mittareita. Visualisointien rakentamiseen on tarjolla kasapäin erilaisia

ohjelmistoja, kuten PowerBI ja Tableau. Näiden avulla yritykset pystyvät rakentamaan juuri itselleen sopivan raportin. Tämänkaltaiset raportointimallit on usein helppo jakaa esimerkiksi toimiston seinällä olevassa televisiossa. (Chartio, 2021.)

3 Pilvipalveluiden perusteet ja niiden yhteys liiketoimintaan

Pilvipalveluiden yleistyminen voidaan verrata karkeasti aikaan, jolloin sähkö yleistyi. Omista generaattoreista siirryttiin kohti isompia jaettavia voimaloita, joita useampi ihminen pystyi hyödyntämään. Vertailuna voidaan myös käyttää liikkumismuotoja. Pitkien matkojen siirtymiseen voidaan käyttää kaikkea omasta autosta lentoyhtiöiden lentokoneisiin. Pilvipalveluiden perusajatus on sama. Ne tarjoavat isoja palvelinkeskuksia ja palveluita, joita useamman käyttäjät voivat hyödyntää. (Sandholm & Dongman 2014, 1–10.)

Ennen julkopilven yleistymistä yritysten täytyi itse omistaa tarvittava laskentateho. Usein tämä oli kallis investointi, joka vaati paitsi oikeat kontaktit, mutta myös merkittävää osaamista serverien ja tietokantojen hallinnasta. Yksi pilvipalveluiden merkittävimmistä kustannushyödyistä tulee siitä, että palvelut ovat skaalautuva. Käyttäjä voi itse päättää paljonko milläkin hetkellä tarvitsee laskentatehoa. Joinain päivinä tehoa voi tarvita merkittävästi enemmän ja joinain taas vähemmän. Näin ollen yritykset maksavat vain siitä laskentatehosta ja resursseista, joita ne käyttävät. (Sandholm & Dongman 2014, 1–10.)

Julkopilven suosio kasvoikin juuri siitä, että aiemmin palvelut, jotka olivat saatavilla, vain isoille yrityksille, olivat nyt käytettävissä, jopa start-upeille. Erilaiset demot oli mahdollista saada pystyyn merkittävästi nopeammin pilvessä, kuin, että itse ostaisi fyysisen serverin (Sandholm & Dongman 2014, 1–10.)

Johdon raportointi on perinteisesti nähty perinteisesti lähes pelkästään raportointina, joka on pakollista esimerkiksi tilintarkastusta varten. Tämänkaltainen raportointi on yleensä ollut hyvinkin staattista ja pyrkinyt lähinnä raportoimaan historiatietoja lukuina tai yksinkertaisina visualisointeina. (De Langhe&Punthoni, 2023.)

Analytiikka-alustojen ja raportointityökalujen kehittyessä, myös raportointitarpeet ovat muuttuneet ja kehittyneet samalla. Raportointi on kehittynyt historiatiedon staattisesta raportoinnista dynaamiseen ja ennustavaan raportointiin, jonka tarkoitus ei ole tyydyttää viranomaisten tarpeita tilintarkastusta varten, vaan esimerkiksi ennustaa asiakkaiden ostokäyttäytymistä. Tämänkaltainen raportointi vaatii tuekseen pelkän taulukkolaskentatyökalun ja laskentatoimen ammattilaisen lisäksi muutakin, kuten

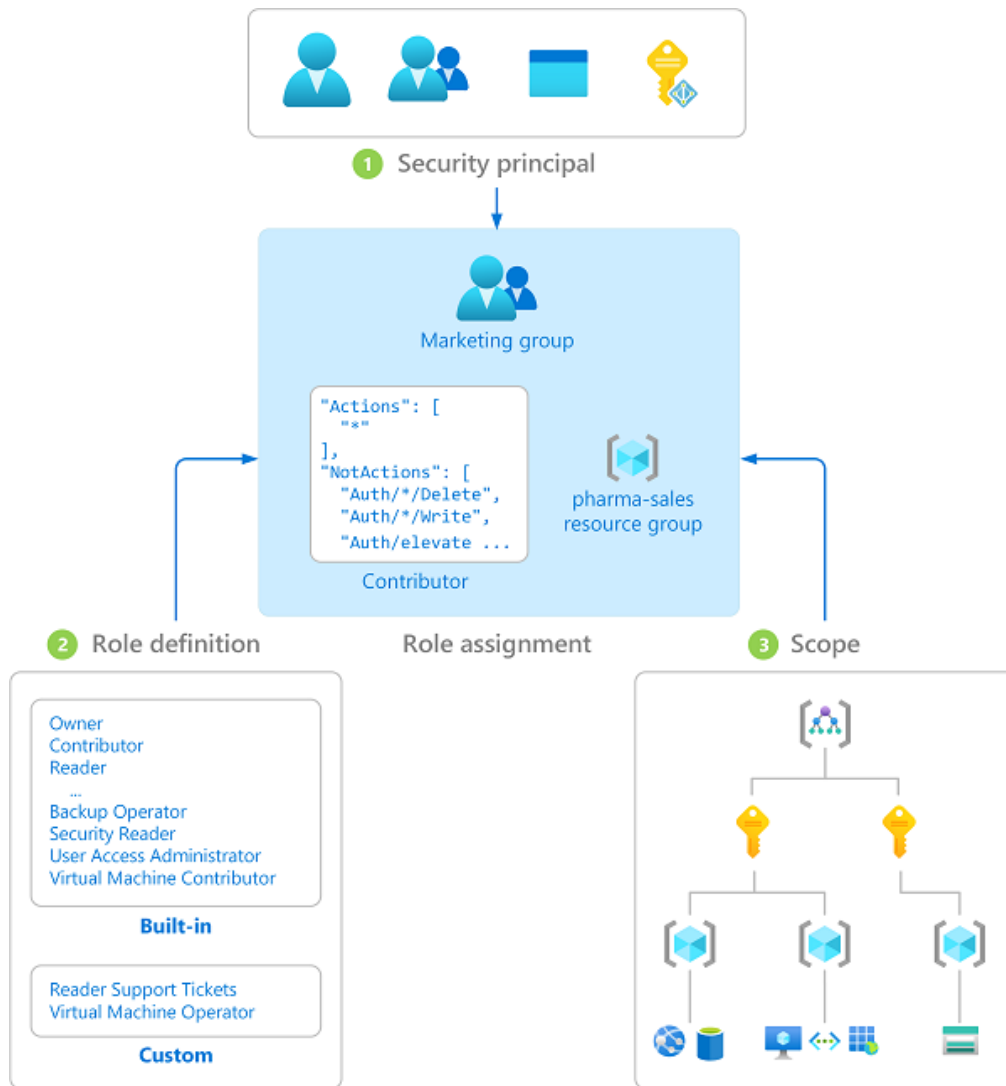
esimerkiksi infrastruktuurin data-alustalle, BI-kehittäjän tai tiimin ja jonkinlaisen prosessikuvauksen raportoinnista. (De Langhe&Punthoni, 2023.)

Kehittyneet työkalut mahdollistavat liiketoiminnalle uusia mahdollisuuksia raportointiin, joita ei ennen edes ajateltu. Tekoäly ja koneoppiminen ovat avainasemassa, kun aletaan puhumaan ennustavasta raportoinnista. Ennustavan raportoinnin lisäksi on mahdollista saada reaaliaikaista tietoa suoraan raporteille ja visualisoida se. Tästä enemmän kappaleessa 7. Toisaalta raportoinnin siirtyessä enemmän siihen, että data on kaikille saataville, niin mahdollistaa tämä datan uudenlaisen tutkiminen myös laskentatoimen työntekijöille. Datan tutkiminen ja ad-hoc -raportointi ei ole enää pelkästään teknisten BI-kehittäjien vastuulla, vaan liiketoiminta voi suoraan tutkia dataa itse. Esimerkiksi erilaiset maksuvalmiuden raportit tai myynnin seuranta ovat helpommin seurattavia ja tutkittavia, kun liiketoimintakäyttäjät pääset itse tutkimaan dataa. (ks. De Langhe&Punthoni, 2023.)

3.1 PaaS ja roolitukset

Microsoft Azure on Microsoftin tarjoama pilvipalvelu. Sitä voidaan käyttää esimerkiksi virtuaalipalvelinten alustana tai yleisesti kehitysalustana. Usein puhutaan IaaS (Infrastructure as a Service) ja PaaS (Platform as a Service). Siinä missä alkutilanteessa yritysten täytyi itse ostaa ja luoda resurssinsa, infransa, sekä koko palvelunsa, niin Azure tarjoaa monia valmiita ratkaisuja. Nämä ratkaisut kattavat esimerkiksi laajan alan tietokantoja, analytiikka-alustoja, sekä niihin liittyviä komponentteja. Azuren avulla yrityksen ei siis tarvitse välttämättä itse luoda monimutkaista koodia, vaan sen voi käyttää valmiita ratkaisuja, mikäli se niin haluaa. (Microsoft, 2023.)

Kuten monet muut pilvipalvelut Azure tarjoaa salaisuuksien, kuten salasanojen hallintaan työkaluja. Salaisuuksien lisäksi palvelussa pystyy hallitsemaan käyttäjien identiteettejä ja niiden käyttöoikeuksia, sekä palveluiden verkotuksia. Näiden avulla yritykset kykenevät hallitsemaan ja monitoroimaan kenellä on pääsy eri palveluihin. Azuressa on monia eri tapoja muodostaa pääsynhallintastrategia. Yksi peruseriaatteista on rooliperusteinen käyttäjänhallinta eli RBAC (Role Based Access Control). Rooliperusteissa hallinnassa (Kuvio 1.)



Kuvio 1 Azuren roolituksen arkkitehtuuri (Microsoft, 2024)

kullekin käyttäjälle määrätään tietyt oikeudet. Toisella käyttäjällä voi esimerkiksi olla pelkkä lukuoikeus tiedostoihin, kun toinen käyttäjä taas pystyy poistamaan ja lisäämään tiedostoja. (Microsoft, 2023.)

Roolit itsessään eivät välttämättä tarjoa riittävää suojausta esimerkiksi tietokantojen taulutasolla, vaan käytetyissä tietokannoissa ja analytiikka-alustoissa voidaan määrätä henkilöille pääsy vain tiettyihin tauluihin. Yleensä roolitukset hoidetaan etenkin laajassa mittakaavassa jakamalla käyttäjät ryhmiin.

Roolitus yrityksen toiminnassa voi näyttää esimerkiksi seuraavalta:

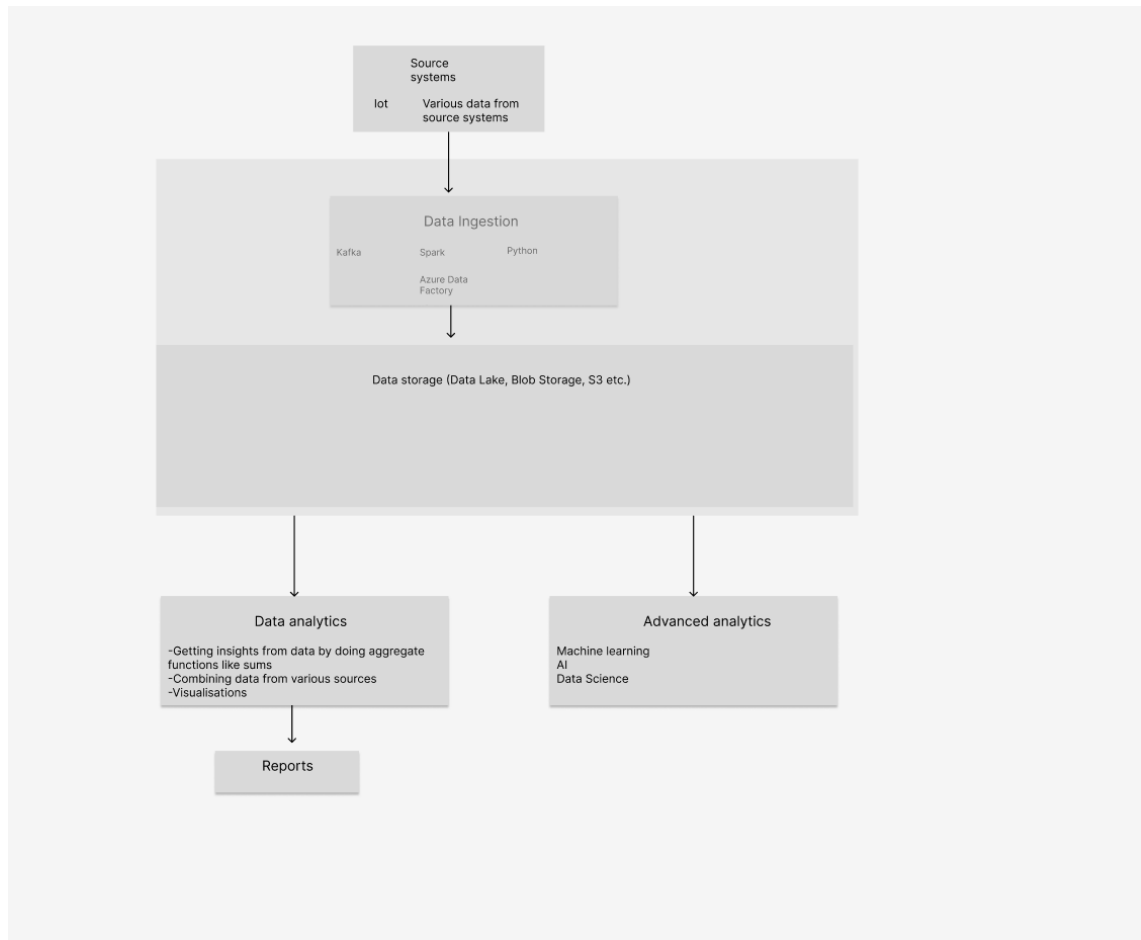
- Kirjanpitäjillä on lukuoikeus vain tiettyihin tietokantoihin. He näkevät vain tarpeellisen, mutta eivät näe kaikkia resursseja, eivätkä pysty jakamaan muille käyttäjille rooleja.

- Johtajat voivat nähdä laajasti analytiikka-aineiston ja raportoinnin. He pystyvät tarkastelemaan yrityksen kaikkia lukuja, mutta eivät välttämättä pysty poistamaan tietoja tai jakamaan oikeuksia

Esimerkkitilanteessa yritys on rakentanut infransa sillä tavalla, että yksi ihminen ei ole vastuussa kaikesta, tai pysty yksin päättämään kaikkea. Tämä liittyy vahvasti yritysten sisäiseen kontrolliin. Segregation of Duty (SoD) pyrkii mitigoimaan riskin paitsi väärinkäytöksiin, mutta myös virheisiin. Yrityksen huomioidessa oikeanlaisen ja järkevän rooliin se voi parhaimmillaan välttyä isoilta katastrofeilta, jotka voisivat pahimmillaan ajaa yrityksen konkurssiin. (Kim ym. 2002, 165–212.)

Yleensä puhutaan vähäisimpien oikeuksien periaatteesta, eli kullakin toimijalla on vähäisin määrä oikeuksia, joita tarvitsee työnteossa. Mikäli henkilön tarvitsee nähdä vain markkinoinnin dataa ei ole syytä, että pääsy annetaan myös finanssipuolen dataan. (Microsoft, 2024.)

Lähtöjärjestelmät voivat olla monenlaisia, kuten ERP-järjestelmiä, tietokantoja, sensoreita tai jotain muita järjestelmiä, jotka tarjoavat dataa, jota yritys haluaa hyödyntää. Analytiikka-alustan ajatuksena on tuoda kaikki tämä data yhteen paikkaan, jossa se on käytettävissä koko organisaatiolle. Tämänkaltaiseen tehtävään liittyy yleensä integraatiot eli lähdejärjestelmien datan saaminen yhtenäiseen lokaatioon, jossa tieto säilötään (Kuvio 2.).



Kuvio 2 Data-alustan arkkitehtuuri mukailten (Fugini ym. 2021)

Monissa moderneissa tietokannoissa ja ERP-järjestelmissä on esimerkiksi valmiita työkaluja yhteuden muodostamiseksi lähdejärjestelmän ja yhtenäisen tiedonvarastointiratkaisun välillä. Organisaation ei siis välttämättä tarvitse ohjelmoida ja konfiguroida monimutkaista integraatiota, vaan yksinkertainen klikkailu graaffisella käyttöliittymällä saattaa riittää. Yksi tämänkaltaisen työkalu on Azuren Logic apps. Logic appsin avulla yritys pystyy graaffisen käyttöliittymän kautta integroimaan eli tuomaan tietoa vanhasta järjestelmästä Azureen ja sitä kautta hyödyntämään dataa erilaisten resurssien avulla. Integraatiot kuitenkin vaativat aina tarkkaavaisuutta ja pohdintaa, jotta ne toimivat oikein. (Microsoft, 2024.)

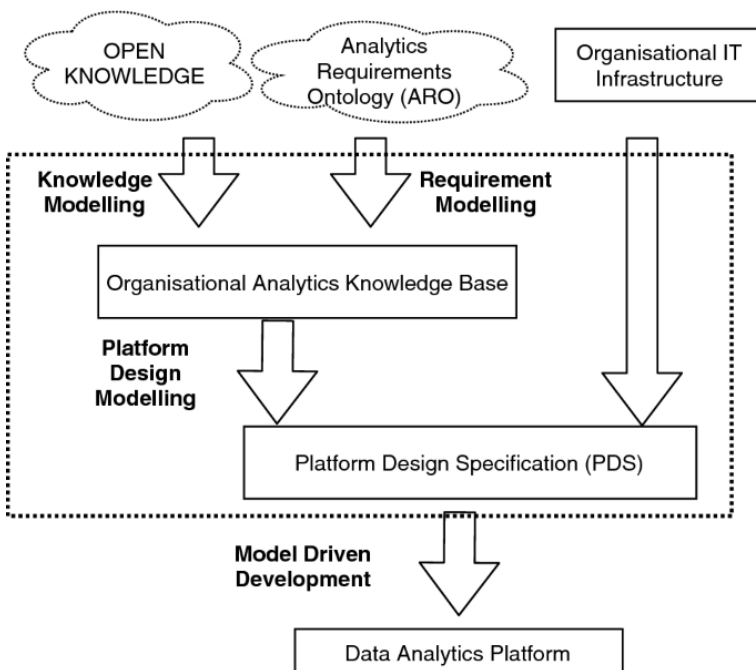
Datan säilytykseen on useita eri vaihtoehtoja. Tiimin on tärkeää sopia, että miten dataa säilötään ja minkälaista dataa aiotaan ylipäättään säilöä. Tämä on tärkeä osa prosessia monesta syystä. Tietovarastointi ensinnäkin mahdollistaa useiden eri datalähteiden yhdistämisen, mutta myös isojen datamassojen kyselyn ja hallinnan. Liiketoiminnan kannalta tämä on äärimmäisen kriittistä. Mikäli halutaan edetä kohti edistynyttä

analytiikkaa, kuten koneoppimista ja tekoälyä dataa on oltava yhtenäisesti, paljon ja samassa paikassa helposti saatavilla. (Alexopoulos ym. 2014.)

Liiketoiminnan on tässä kohtaa tärkeä tuoda ilmi, minkälaista loppuraporttia toivotaan ja minkälaisia käyttömahdollisuuksia datalle halutaan tulevaisuudessa. Datan säilytys ei ole välttämättä täysin yksinkertaista siinä suhteessa, että kaikki tieto tuotaisiin samaan paikkaan ja prosessi olisi valmis. Tämän jälkeen tieto pitää vielä saada muotoon, jossa liiketoiminta pystyy sitä hyödyntämään. Tietoaltaan ja tietovaraston käsitteet käydään tarkemmin läpi kappaleessa 5.1. (Alexopoulos ym. 2014.)

3.2 Analytiikka-alustan määrittely ja implementointi

Analytiikka alustan määrittelemiseen voidaan käyttää DASE-mallia (Data Analytics Solution Engineering). Tämän mallin (kuvio 3.) avulla voidaan rakentaa analytiikka-alusta ja se antaa raamit sille, mikä oikeastaan on analytiikka-alusta. Mallin avulla pyritään tunnistamaan toistuvia operatiivisia analytiikkatarpeita ja rakentamaan ratkaisuja niihin. Malli (Kuvio 3.) korostaa asiantuntijoiden tietoa ja yhteistyötä yrityksen sidosryhmien välillä. Tämä vaatii yrityksen johdolta näkemystä siitä, mikä yrityksen it-strategia on ja mitä tarpeita sekä vaatimuksia tämä strategia sisältää. (Bandara ym. 2023, 195–212.)



Kuvio 3 DASE-arkkitehtuuri (Bandara ym. 2023, 195–212)

Yrityksen tulee tarkastella sitä mitä lopullisia tavoitteita raportoinnille ja datan hallinnalle on. Mitkä kaikki mittarit ovat tärkeitä ja mitä asioita ylipäättään halutaan mitata. Analytiikka-alusta voi myös mahdollistaa tulevien kassavirtojen ja markkinan ennustamisen aiempaa tarkemmin. Täten voi nousta uusia raportointitarpeita. Raportointiin liittyy myös tulevan datan ja historiadatan hallinta. Datan elinkaaren hallinta on tärkeää loppuraportin paikkansapitävyyden kannalta, sillä huono datan laatu tekee raportista pahimmillaan haitallisia. (Bandara ym. 2023, 195–212.)

Mikäli yritys on alaytiikkapolullaan alkuvaiheessa, voi siirtyminen analytiikka-alustalle tuoda myös paljon uusia tarpeita. Modernit analytiikka-alustat mahdollistavat myös kokeilujen tekemisen aiempaa helpommin, sillä data on yhtenäisessä paikassa ja oikeanlaisessa muodossa. Samalla alustalla voi siis parhaimmillaan muodostaa uusia koneoppimisen implementaatioita, sekä data-analytiikan analyyseja, että perinteisestä raportointia. Tämä tuo datan lähemmäs liiketoimintaa ja helpottaa liiketoiminnan käyttäjien aktiivista innovointia. (Bandara ym. 2023, 195–212.)

Etenkin innovointien näkökulmasta analytiikka-alustan käyttöönotto usein vähentää osastojen välistä siiloutumista ja helpottaa liiketoiminnan osallistumista myös raportointiputken alkupäähän. Tällöin toimintoja voidaan tehostaa esimerkiksi siten, että raskaat aggregoinnit, kuten summat tai ryhmittelyt toteutetaan ennen raporttia. Tämä helpottaa myös datan auditointia ja mitigoi riskiä väärälle datalle, kun rikastettu data saadaan suoraan tietokannasta, eikä sitä tehdä erikseen käyttäjien raporteille. (Bandara ym. 2023, 195–212.)

4 BI-kehitys organisaatiotasolla

Perinteinen Business Intelligence eli BI rakentuu usean tekijän, kuten raportoinnin, tiedon varastoinnin ja data governancen varaan. Tämänkaltaisessa mallissa liiketoiminta tunnistaa itse tarpeensa, jonka jälkeen se ottaa yhteyttä henkilöön tai tiimiin yrityksen sisällä, joka voi toteuttaa tämän raportointitarpeen. Yleensä henkilöt tällaisessa prosessi erotellaan liiketoimintakäyttäjän ja “power userin” eli teknisen ammattilaisen välille. Tekninen ammattilainen on yleensä teknologiaintensiivinen työntekijä, joka toimii raportin rakentajana toimeksiantoa vastaan. Tämä prosessi vaatii yleensä paljon aikaa ja työtunteja. Se vaatii myös ymmärrystä liiketoiminnan puolelta, että minkälainen raportointi on kannattavaa tai mahdollista. (Pałys ym. 2023, 795–803.)

Nykyaikaisessa organisaatiossa tämä on muuttumassa enemmän self-service BI:n eli SSBI:n suuntaan. SSBI:llä tarkoitetaan tilannetta, jossa liiketoimintakäyttäjä tuodaan lähemmäs raportointia. Tämä tuo selkeänä etuna esimerkiksi innovoinnin lisääntymisen ja siiloutumisen vähenemisen teknisen ammattilaisen ja liiketoimintakäyttäjän välillä. (Pałys ym. 2023, 795–803.)

Siirtyminen SSBI:hin ei tule ilman haasteita. Datan laadun varmistaminen saattaa olla vaikeaa, mikäli oikeita prosesseja ei implementoida. Tietoturva-aspektit ovat myös kysymys, sillä SSBI:n yksi periaate on se, että liiketoimintakäyttäjillä tulisi olla oikeuksia laajasti erilaiseen dataan. Tietoturvan ja laadun lisäksi ongelma on teknisissä taidoissa ja ymmärryksessä. SSBI yrittää luoda innovointia sen kautta, että liiketoiminnalla on laajempi näkemys ja käyttöoikeus dataan. Tämä ei kuitenkaan auta, mikäli liiketoiminnalla ei ole osaamista ja tietotaitoa datan hyödyntämiseen. (Pałys ym. (2023), 795–803.)

Ratkaisuna tähän on esitetty Managed Self-Service BI:tä eli MSSBI. Inmon (2015) vertaa SSBI:tä siihen, että lapsille annetaan sormivärejä ja oletetaan, että näistä kasvaa huipputaitelijoita. MSSBI:n idea on antaa työntekijöille oikeat työkalut, koulutus ja mahdollistajat siihen, että yrityksen BI on korkeatasoista ja kaikille saataville olevaa. Inmon (2015) myös selostaa, että self-service BI on häilyvä termi, koska IT joutuu yleensä aina lataamaan ja valmistelemaan datan, tarjoamaan pääsyn siihen ja varmistamaan, että kaikki toimii kuten pitää. Tämä kaikki tapahtuu ennen analysointiprosessia. (Inmon ym. 2015.)

4.1 Agile

BI:n siirtyessä yhä enemmän pois siitä, että liiketoiminta pelkästään ilmoittaa raportointitarpeet ja tekniset työntekijät toteuttavat nämä tarpeet, tulee myös työskentelytapojen muuttua. Yksi keskeisimpiä ongelmakohtia käytännössä kaikessa teknisessä kehityksessä on ollut tilanne, jossa valmiin tuotteen saaminen kestää todella kauan. Yleensä lopputuote ei ole sitä mitä on haluttu. Tämä voi johtua useasta syystä. Raportin kehitys ei välttämättä ole epäonnistunut, vaan raportointitarpeet ovat yksinkertaisesti muuttuneet. Miten tällaiseen tilanteen tulisi varautua? (Inmon ym. 2015.)

Agile-metodien ajatuksena on se, että asiakkaalle ei tarjota täysin valmista tuotetta, jossa on kaikki ominaisuudet toiminnassa ja valmiina. Sen sijaan tarjotaan MVP (Minimum Viable Product) eli minimivaateet täyttävä tuote. Raportoinnin kannalta tämänkaltainen tilanne voi olla esimerkiksi uusi raportointijärjestelmä, jonka tarkoituksena on tuottaa 5 eri raporttia. Asiakkaalle tarjotaan heti ensimmäinen raportti, eikä odoteta, että muut valmistuvat. Toisaalta välttämättä yhtäkään raporttia ei odoteta loppuun asti ennen sen tarjoamista asiakkaalle. MVP määritetään datatuotekohtaisesti asiakkaan kanssa ja se on täysin riippuvainen tilanteesta, tapauksesta ja sille asetetusta viitekehuksesta. (The Scrum Culture, 2015.)

Agilesta on olemassa useampia eri versioita ja muotoja. Yksi vaihtoehto on Scrum. Scrumin perusidea on se, että kehitys jaetaan sprintteihin eli prosesseihin, jotka kestävät yleensä kaksi viikkoa. Näille kahdelle viikolle sovitaan tehtävät ja tehtävät pyritään suorittamaan näiden kahden viikon aikana. Suunnittelupalaverit pidetään aina sprintin alussa palaverina, johon osallistuu kaikki tiimin jäsenet. Suunnittelupalaverin lisäksi sprintin aikana voidaan pitää esimerkiksi päiväpalavereita (daily) tai viikkopalavereita (weekly). Näiden tarkoituksena on käydä läpi, miten eri tehtävät edistyvät, tai mistä syystä ne eivät edisty. Ideana tässä on se, että mikäli jokin asia ei edisty, niin se käy ilmi mahdollisimman nopeasti, jolloin pullonkauloilta kehityksessä vältytään. Palaverien tarkoituksena on pullonkaulojen välttämisen ja tunnistamisen lisäksi pyrkiä allokoimaan resursseja mahdollisimman tehokkaasti. Esimerkiksi mikäli joku tiimin jäsen on jatkuvasti ylityöllistetty tai jollakin liian vähän tekemistä, voidaan resursseja allokoida tehokkaammin. (The Scrum Culture, 2015.)

Palaverien lisäksi sprinttien lopussa järjestetään yleensä palaveri, jossa käydään läpi kulunut sprinttiä. Tarkoituksena on käydä läpi sen aikana tapahtunutta kehitystä,

onnistumisia ja ilmi käyneitä haasteita. Ideana tässä taustalla on se, että tiimin sisällä avoimesti keskusteltaisiin, mitä hyvää, huono tai kehitettävää toiminnassa on. Oikeanlaisen palaverin pitäminen ei ole yksinkertaisesta. Tilasta on luotava sellainen, jossa jokainen uskaltaa ja haluaa sanoa oman mielipiteensä riippumatta siitä, onko se kehityskohde vai onnistumisen hehkuttamista. (The Scrum Culture, 2015.)

Agilekehitys ei toimi ilman oikeanlaista tiimiä. Kehittäjien lisäksi tiimi vaatii projektipäällikön, Scrum Masterin ja tuoteomistajan. Roolit eivät välttämättä ole kaikki erillisillä toimijoilla, mutta toimintamalli pysyy silti samana. Scrum masterin toiminnan tehtävä on käydä läpi tulevia tehtäviä ja toisaalta niitä, jota ei vielä ole ehditty tehdä eli backlogia. Backlogin funktio on esittää niitä tehtäviä, joita tiimin ajatuksena on suorittaa projektin aikana. Scrum masterin vastuulla on sopia tiimin kesken missä ja missä sprinteissä näitä tehtäviä suoritetaan. (The Scrum Culture, 2015.)

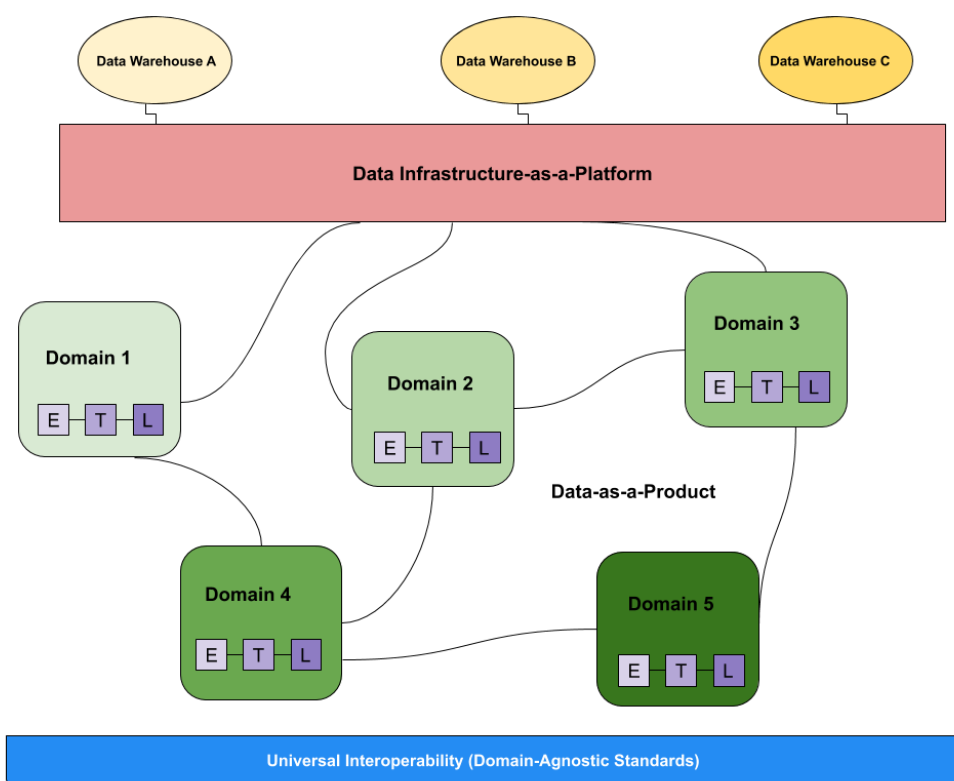
Agile ei ole oikotie onneen, eikä se itsessään ratkaise kaikkia kehitykseen liittyviä ongelmia. Onnistuneeseen kehitykseen on silti pohdittava esimerkiksi sitä, mikä on mvp missäkin tuotteessa. (The Scrum Culture, 2015.)

4.2 Data mesh

Data mesh on trendaava käsite ja sen ajatus sitoutuu pitkälti MSSBI:n periaatteisiin. Data mesh tarjoaa ratkaisun tilanteeseen, jossa datan käsittely ja toimitus on keskitetty. Perinteisessä keskitetyssä mallissa liiketoiminta nähdään ikään kuin loppuasiakkaana, jolle datatiimi kerää, varastoi ja jalostaa dataa. Ongelmaksi tämä muodostuu siinä kohtaa, kun data ei olekaan enää liiketoiminnan sivutuote, jota voidaan hyödyntää tai sitten ei. Informaation hyödyntäminen on nykypäivänä pakollinen osa liiketoimintaa. (Machado ym. 2021.)

Data meshin ideana on poistua keskitystä prosessista. Vaikka datan säilytys itsessään olisi keskitetty, ei se tarkoita sitä, että tiimien tulisi olla. Data meshissä toimintaperiaatteena on alueellinen ja hajautettu omistajuus, sekä arkkitehtuuri. Yksi selkeä etu tässä on se, että liiketoiminta siirtyy lähemmäksi tietoa. Tämä kuulostaa tutulta MSSBI:stä, jossa myös pyritään saamaan data kaikkien saataville ja liiketoiminnan, sekä kehittäjien tulisi toimia yhdessä, eikä erikseen. Samoin data mesh korostaa itsepalvelun periaatetta, jossa infra ja data on saatavilla kaikille sitä tarvitseville. (Montecarlo data, 2020.)

Liiketoiminnan ja BI-kehityksen yhdistyessä muodostetaan domaineja, jolloin dataa ei käsitellä pelkästään informaationa, vaan datatuotteena. Näille datatuotteille tulisi löytyä aina omistaja. Yksi etu mitä data mesh tällä tavoin saavuttaa on se, että tuoteomainen ajattelu antaa datalle jo lähes itseisarvona tuotemaisia ominaisuuksia, kuten laskennallisuuden, laadun ja käyttötarkoituksen. Tämänkaltaiset ominaisuudet sovitaan yleensä jo datatuotteen suunnitteluvaiheessa. Tällä tavoin vältetään tilanteelta, jossa tiedonkeruu ja jalostus on musta laatikko liiketoiminnalle, eikä sitä ymmärretä tai sen laatua ja toimintaa valvota. Data meshissä ylätasolla yhdessä tai useammasta tietovarastosta muodostetaan infra, josta etl-prosessien avulla domain-tiimit rakentavat tietotuotteita ks. Kuvio 4. (Montecarlodata, 2020.)



Kuvio 4 Data mesh (Montecarlodata, 2020)

Data mesh, SSBI ja MSSBI ovat hyviä esimerkkejä siitä, mihin suuntaan tiimirakenne on menossa. Data halutaan yhä vahvemmin tuotteistaa ja summittaisesta big datan hyödyntämisestä halutaan tilanteeseen, jossa lopullisella tietotuotteella on omistaja. Tämä omistaja ei pelkästään ymmärrä lopullista tuotetta tai raporttia, vaan kokonaisuuden, jonka avulla tämä tuote rakennetaan. Tämä ei tarkoita sitä, että tuoteomistajan rooli olisi ymmärtää teknisesti kaikki tai liiketoiminnan näkökulmasta jokainen perspektiivi, vaan ennemmin kokonaiskuvan hahmottaminen tuotteen ympärillä. (Montecarlodata, 2020.)

Data meshin yksi vahvuus tulee sen domain rakenteesta perinteisen BI:n heikkouksiin kuuluu vaikeus tai jopa mahdottomuus tehdä ad-hoc kyselyitä ja analyyseja dataa vastaan. Tämä johtui yksinkertaisesti siitä, että liiketoiminnalla ei ollut näkyvyyttä raportointiprosessiin ja koko tietotuotteen käsitys oli vieras. SSBI ja MSSBI sen sijaan tunnistaa tarpeen ad-hoc kyselyille, mutta ei tarjoa suoranaista vastausta siihen, miten datan laatua ja sen osuuksia tulisi hallita. Data mesh sen sijaan tarjoaa melko suoran viitekehyksen sille, millä tavalla eri tiimien tulisi rakentua datatuotteiden ympärille ja miten eri tiimit voivat hyötyä muiden domain tiimien datatuotteista. (Montecarlo data, 2020.)

5 Dataputket ja transformointi

ETL tulee sanoista Extract, Transform, Load. Sillä tarkoitetaan prosessia, jossa lähtödata jostain lähdejärjestelmästä noudetaan, muutetaan käytettävään muotoon ja ladataan tallennuspaikkaan. ETL-prosessi voi tapahtua myös eri jaksoissa, kuten ELT, mutta peruseriaate pysyy samana. (Karagiannis ym. 2013, 927–945.)

Alkulähteenä toimii jonkinlaiset lähdejärjestelmät. Nämä järjestelmät voivat olla esimerkiksi ERP-järjestelmistä tai sensoreita, joista halutaan dataa raporteille. Yksi etu tällaisessa, on tilanne, jossa yrityksellä on useita eri lähdejärjestelmiä ja niistä saatavaa data halutaan yhdistää. Esimerkiksi mikäli yritys laskee tuotteiden kannattavuutta tilanteessa, jossa erp-järjestelmä näyttää pelkästään myynnit mutta ei laatukustannuksia ei saada tarkkaa dataa. Yritys voi etl-prosessilla yhdistää eri lähteistä tulevaa data tietovarastoon ja luoda raportin, joka kertoo tuotteiden katekenteen aiempaa tarkemmin. (ks. Karagiannis ym. 2013, 927–945.)

Yleensä etl-prosessit kehitetään ja ne automatisoidaan toimimaan tietyllä syklillä. Myyntidata voidaan päivittää esimerkiksi reaaliaikaiseksi, mutta edellä mainitut laatukustannukset voidaan laskea kerran päivässä tai viikossa. Tämänkaltaisen optimointi vaatii paitsi tietojenkäsittelyn osaamista, mutta myös liiketoiminnan osallistumista. Yleensä tiimissä on tai siihen voidaan implementoida liiketoiminta-analyttikon tittelillä toimiva henkilö. Liiketoiminta-analyttikon toimii ikään kuin linkkinä teknologiatimin ja liiketoiminnan välillä. Tämänkaltaista ideaa on puhuttu esimerkiksi data mesh -paradigman avulla. Yritysten täytyy tehdä ETL-prosessista osa raportointia ja tuotteistaa data. (Ks. Montecarlo data, 2020) Yritysten täytyy kyetä hyödyntämään teknologiaa tehokkaalla tavalla. Pelkkä napin painaminen harvoin riittää, vaan täytyy ymmärtää kattavasti liiketoimintaympäristö ja sille sopivat prosessit. (Lubwama, 2019.)

ETL-prosessien ymmärtäminen vaatii myös ymmärrystä ja näkemystä lopputuotteesta. ETL prosessin tärkeys perustuu osittain myös siihen, että muovaa liiketoimintadataa tarkemmaksi ja yksityiskohtaisemmaksi. Sen tarkoitus on olla osa prosessia, jossa luodaan syväluotaava kuva organisaation tietovarastoon virtaavasta datasta. ETL nähdään usein pelkästään automaation osana ja pakollisena keinona siirtää tai sisällyttää dataa. ETL on kuitenkin paljon, muuta ja siihen liittyy paljon teknisiä, mutta myös

liiketoimintapuolen aspekteja, jotka on tärkeä ottaa huomioon prosessia suunniteltaessa. (Lubwama, 2019.)

Herkästi pohdittavaksi tulee, että miksi transformaatiota edes tarvitaan. Miksi yritykset eivät hyödynnä raakadataa sellaisenaan? Kysymys on hyvä ja vastaus siihen ei ole täysin yksioikoinen. Raakadata on yleensä nimensä mukaisesti jalostamatonta dataa suoraan lähteestä. Sen tiedostomuoto saattaa olla vaikeasti hyödynnettävä tai se sisältää myös ylimääräistä dataa, jota liiketoiminta ei tarvitse, kuten eri prosessien metatietoja. Tästä syystä raakadataa yleensä puhdistetaan, jäsenellään ja siistitään siten, että sen hyödyntämisestä tulee mieluisaa. Esimerkkinä tästä on rajapinnoista usein saatava json-data (Kuvio 5).

```
"employees": [  
  {  
    "id": "1",  
    "first_name": "John",  
    "last_name": "Doe",  
    "email": "john.doe@example.com",  
    "gender": "Male",  
    "salary": "5000",  
    "years_of_experience": "3"  
  },  
]
```

Kuvio 5 Json-esimerkkidataa, joka sisältää yhden työntekijän tiedot

Tämä data on ihmiselle luettavassa muodossa, mutta mikäli näitä objekteja on paljon, on niiden läpikäyminen työlästä ja vertailu hankalaa. (Sajida & Ramakrishna 2015.)

ETL-prosessin ensimmäinen vaihe, eli extract alkaa sillä, että tämä data luetaan lähdejärjestelmästä. Tämän jälkeen sille tehdään tarvittavat transformaatiot eli esimerkiksi määritetään kullekin kentälle tietotyytit. Tämä tehdään siksi, että kun raporttia lähdetään rakentamaan, niin eri lähteistä tulevat tietotyytit eivät välttämättä suoraan tue toisiaan, jolloin niiden yhdistely ei onnistu. (Sajida & Ramakrishna 2015.)

Load prosessissa tämä tieto ladataan esimerkiksi tietokantaan. Tässä tietokantana on avoimen lähdekoodin PostgreSQL-tietokanta (kuvio 6.).

| | id text | first_name text | last_name text | email text | gender text | salary integer | years_of_experience integer |
|---|------------|--------------------|-------------------|----------------------|----------------|-------------------|--------------------------------|
| 1 | 1 | John | Doe | john.doe@example.com | Male | 5000 | 3 |
| 2 | 2 | Jane | Doe | jane.doe@example.com | Female | 6000 | 2 |
| 3 | 3 | Jim | Beam | jim.beam@example.com | Male | 7000 | 5 |

Kuvio 6 Json-data mallinnettuna tauluksi tietokannassa

Kuvasta näemme, että yhden työntekijän tiedot ovat nyt siististi yhdellä rivillä. Tämänkaltaista dataa on helpompi tarkastella suurempia määriä ja toisaalta epäoleellisia kenttiä voidaan myös tiputtaa pois. Koko tämän prosessin ei tulisi olla pelkästään teknisen puolen vastuulla, vaan liiketoimintaa on tärkeä osallistaa aina kun sille on tarve. Hyvänä esimerkkinä tästä voi olla se, että kaikista näistä tiedoista tarvitaan pelkästään osa, jolloin taulua voidaan muokata. (ks. Sajida & Ramakrishna 2015.)

Idea näissä tietokantatauluissa on se, että lopuksi yhdestä tai useasta taulusta muodostetaan näkymä, jota halutaan hyödyntää raportoinnissa. Tätä näkymää voidaan suoraan hyödyntää raportointityökalujen, kuten PowerBI:n avulla. Näin ollen raportilla näkyvä tieto on peräisin suoraan tietokannasta, jota päivitetään ja valvotaan. Tämä aiheuttaa sen, että mikäli raportteja on useampia, niin voidaan saada varmuus siitä, että käytetty data on oikeaa ja se tulee oikeasta lähteestä. Verrattuna siihen, että jokainen työntekijä lataisi tiedot ensin omalle työasemalleen ja sen perusteella loisi raportin, on tämä merkittävä parannus datan laadun varmistuksessa, tietoturvassa ja ylipäätään koko raportointiprosessissa. (Sajida & Ramakrishna 2015.)

6 Tiedon varastointi

Yrityksen siirtyessä edistyneeseen raportointiin tulee tietokannat väistämättä eteen. Tietokantatyyppejä on useita erilaisia ja niihin sisältyy kaikkiin hyviä ja huonoja puolia. Yrityksen on tunnistettava minkälaista dataa se kerää ja miten se aikoo sitä hyödyntää. Monesti hyppäys saattaa tuntua suurelta, mikäli aiemmin jokainen työntekijä on käyttänyt esimerkiksi omaa taulukkolaskentatyökaluaan datan tarkasteluun. Yleensä tässä kohtaa yritys joko itse kouluttaa tai palkkaa asiantuntijan auttamaan tietokannan pystyttämässä, sekä ylläpidossa. (Kimball & Margy 2013.)

Tietokannan käyttäminen mahdollistaa sen, että yritys siirtyy hyödyntämään yhtenäisestä tietovarastoa. Tämän lisäksi datalle määritetään usein muitakin kriteereitä, kuten datan omistaja ja erilaisia laatuvaatimuksia, sekä mittareita. Nämä eivät välttämättä ole sellaisia tietoja, jotka näkyisivät suoraan raportoinnilla vaan enemmän testejä, joilla lähtödatan laatua valvotaan. (Kimball & Margy 2013.)

Puhuttaessa tiedonkäsittelystä ja sen varastoinnista termit menevät herkästi sekaisin, sillä käsitteissä on eroja. Tietoallas on yleistynyt sitä mukaan, kun big datasta on puhe lisääntynyt. Tietoaltaalle ladattava data on usein täysin prosessoimatonta. Tietovarastolla sen sijaan oleva data on usein puhdistettu ja järjestetty siten, että sitä pystytään kyselemään tai se on mallinnettu jollain tavalla. Yrityksillä voi olla erikseen tietoallas, jossa tämä raakadata säilötään. Tietoaltaalle tuleva data tulee yleensä suoraan lähdejärjestelmistä, kuten ERP-järjestelmistä. Tämän jälkeen tieto siirretään tietovarastoon. Siirron aikana data voidaan organisoida ja prosessoida siten, että se on tietovarastolla oikeassa muodossa. Tietoaltaalle yleensä ladataan kaikki data lähdejärjestelmistä ja sen prosessointi tapahtuu vasta siinä kohtaa, kun se siirretään varastoon. (Khine ym. 2018.)

Tietoallas, esimerkiksi Azure Data Lake on yleensä liiketoimintakäyttäjälle melko näkymätön osa. Liiketoimintakäyttäjälle tarjotaan se data, joka on valmiina oikeassa muodossa analyyseja varten. Sen sijaan esimerkiksi data-tieteilijöille voidaan tarjota jonkinlainen suora pääsy tietoaltaalle. Yksi syy tähän voi olla datan tutkiminen tai esimerkiksi koneoppimismallien tekeminen, jotka vaativat huomattavasti enemmän dataa, kuin liiketoimintakäyttäjien tarpeet keskimäärin. (Kimball & Margy 2013.)

Tietovarastolla terminä tarkoitetaan järjestelmää, jossa säilytetään keskitetysti dataa yhdestä tai useasta järjestelmästä. Näiden järjestelmien pääasiallinen tehtävä on tarjota ja säilöä dataa oikeassa ja strukturoidussa muodossa. Nämä toimivat usein ETL-prosessien load-vaiheen päätepisteenä. Esimerkkinä tietovarastosta on Azure Synapse. Microsoft kuvailee tätä palvelua skaalautuvana yrityskäyttöön suunnattuna tietovarastona, jossa käyttäjä voi kysellä dataa valitsemallaan tavalla. Tämänkaltaiset järjestelmät kestävät suuriakin datamääriä ja niissä on mahdollista skaalata laskentatehoa tilannekohtaisesti. (Microsoft, 2024.)

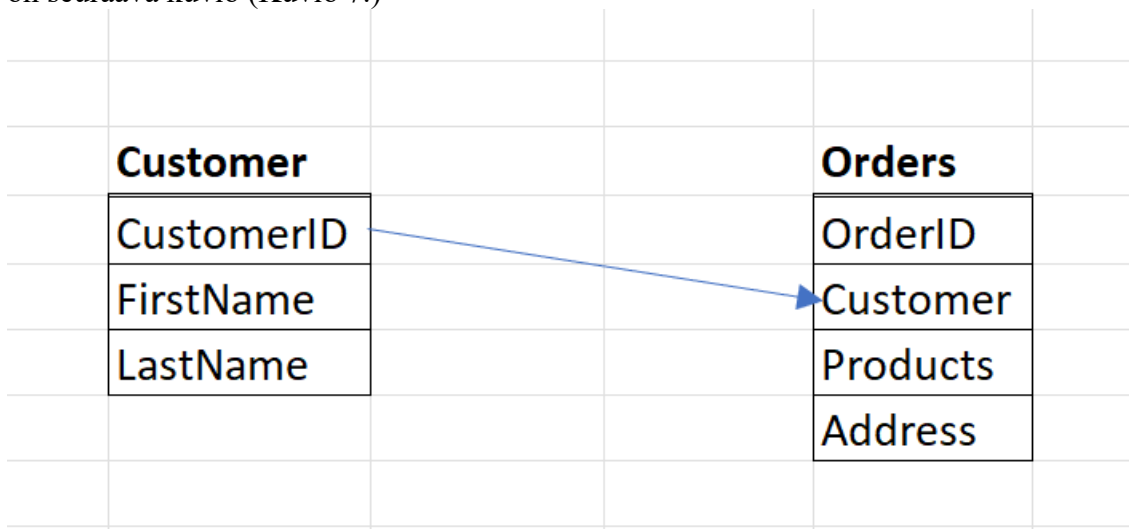
6.1 Relaatiotietokannat

Relaatiotietokannat ovat tietokantoja, jotka varastoivat strukturoitua dataa. Strukturoitu data on yleensä tekstimuodossa ja se on varastoitu riveiksi ja sarakkeiksi. Esimerkkinä strukturoidusta datasta on csv-tiedosto (Comma Separated Value). Strukturoidun datan lisäksi data voi olla semi-strukturoimatonta, tai strukturoimatonta. Semi-strukturoitu data ei sisällä skeemaa ja sisältää useita attribuutteja. Ei strukturoitu data on yleensä videoita tai äänitiedostoja. (Praveen ym. 2020, 67-69.)

Relaatiotietokannat ovat käteviä siinä mielessä, että datalle on määritetty jokin skeema ja sillä on jonkinlainen merkitys. Skeemalla tarkoitetaan usein tietokannan kuvailevaa rakennetta. Se voi kertoa esimerkiksi eri taulujen relaatioita tai jonkin taulun sarakkeiden määrän, niiden nimet ja tietotyypin. Tietokannan skeeman määrittely auttaa usein myös datan validoinnissa ja sen hallinnassa. Mikäli johonkin sarakkeeseen pitäisi aina tulla jokin arvo, niin siihen ei voida asettaa dataa, jossa tätä arvoa ei ole. (Praveen ym. 2020, 67-69.)

Relaatiokantojen avulla yritykset voivat paitsi helposti tehdä erilaisia hakuja tietokantaan SQL:n (Structured Query Language) avulla, mutta myös tunnistaa erilaisia relaatioita eli suhteita. Relaatiotietokantojen rakentaminen vaatii kuitenkin työtä, osaamista ja ylläpitoa. Yksi suurimmista hyödyistä hyvin ylläpidetylle relaatiokannalle on sen luotettavuus ja yksinkertaisuus. Yrityksen raportit voidaan suoraan kytkeä tietokantaan ja tieto voidaan tuoreuttaa niin usein, kun sillä on tarvetta. Tietokanta myös mahdollistaa sen, että tiedolle muodostuu yksi totuudenlähde. Näiden avulla yrityksillä on mahdollisuus hyödyntää dataa ja kätevästi tutkia sitä. (Manning, 2015.)

Tietokantojen rakenteita ja rakennustapoja on monia. Suurin osa malleista kuitenkin keskittyy siihen, miten tietokannan tieto on jaoteltu. Relaatiotietokannoissa tieto yleensä jakaantuu näkymiin ja tauluihin. Taulut ovat perinteisempi tapa sisällyttää dataa. Niiden suurin etu on se, että niihin voi suoraan ladata dataa. Haittapuolina kuitenkin on heikompi pääsynhallinta ja nopeasti kehkeytyvä monimutkaisuus. Sen sijaan näkymissä pääsynhallinta on helpompaa ja eri tietolähteistä koostuva näkymä voi olla hyvinkin monimutkainen, mutta näkyä loppukäyttäjälle kuitenkin yksinkertaisena, mikä pelkkää dataa tarkastellaan. (Flesca ym. 2018.) Yksinkertainen kuvaus tietokantojen relaatioista on seuraava kuvio (Kuvio 7.)



Kuvio 7 Yksinkertainen relaatiotietokannan malli kahdesta taulusta: Customer ja Orders.

Tässä kuviossa pyritään havainnollistamaan kahden taulun välistä relaatiota. Customer-taulussa sarakkeina ovat CustomerID, FirstName ja LastName. Orders-taulussa sen sijaan on OrderID, Customer, Products ja Address. Tehdyn relaation perusteella voidaan päätellä, että CustomerID on sama kuin Customer. Tämä mahdollistaa sen, että nämä taulut voidaan yhdistää tai niistä voidaan tehdä kyselyitä CustomerID:n tai Customer-kentän perusteella. Yleensä tämänkaltaisissa tietokannoissa tietokanta voitaisiin mallintaa esimerkiksi dimensioiden eli tähtimallin avulla. Dimensiomallinnuksessa tietokanta mallinnetaan dimensio- ja faktatauluihin. Jaottelu menee siten, että dimensiotaulussa on asiat, joiden mukaan tietoa haetaan tai pyritään jaottelemaan. Esimerkiksi Kuvio 4. Customer-taulu olisi todennäköisesti Dimensio-taulu ja Orders faktataulu. Tämä voidaan ajatella siten, että Orders-taulu sisältää tiedot jokaisesta yksittäisestä CustomerID:n tai asiakkaan tilauksesta. (ks.Flesca ym. 2018.)

Yksi huomioonotettava seikka, kun tarkastellaan tietokantoja ja niiden mallinnusta on tiedon granulariteetti eli millä tasolla tietoa tarkastellaan. Esimerkissä voidaan tarkastella

tietoja hyvinkin tarkasti ja antaa jokaiselle tilaukselle oman OrderID kentän. Näin ollen jokaisen asiakkaan tilauksia voidaan tarkastella erillisenä transaktiona. Tämä parantaa mahdollisuutta tiedon hyödyntämiseen, mutta myös lisää kertyvän datan määrää. Tietoa myös tarkastella esimerkiksi siten, että rivejä summataan keskenään. Näin ollen jokaisella asiakkaalla on yksi OrderID, mutta esimerkiksi tuotteiden määrä tai muut sarakkeet voivat päivittyä. Tämä vähentää kerrytettyä dataa, mutta herkästi lisää monimutkaisuutta ja vähentää analytiikan tarkkuutta. Mikäli haluttaisiin tarkastella, että mitä tuotteita yleensä ostetaan yhdessä, ei tämä enää onnistu ainakaan samalla tarkkuudella, sillä kaikki ostot ovat saman OrderID:n takana. (ks.Flesca ym. 2018.)

Granulariteettia pohdittaessa on tärkeää selvittää, että mikäli tietoa yhdistellään, niin sen tarkastelun taso on sama. Lopullisen datan laatu kärsii merkittävästi tai se voi olla täysin väärin, mikäli eri tarkastelutason dataa yhdistellään keskenään, eikä tätä oteta huomioon. Tätä voi välttää tapahtumista asettamalla erilaisia ehtoja tai sääntöjä tietokantarakenteisiin, kuten avaimia tai sääntöjä ja testejä. (Dong ym. 2023, 34-48.)

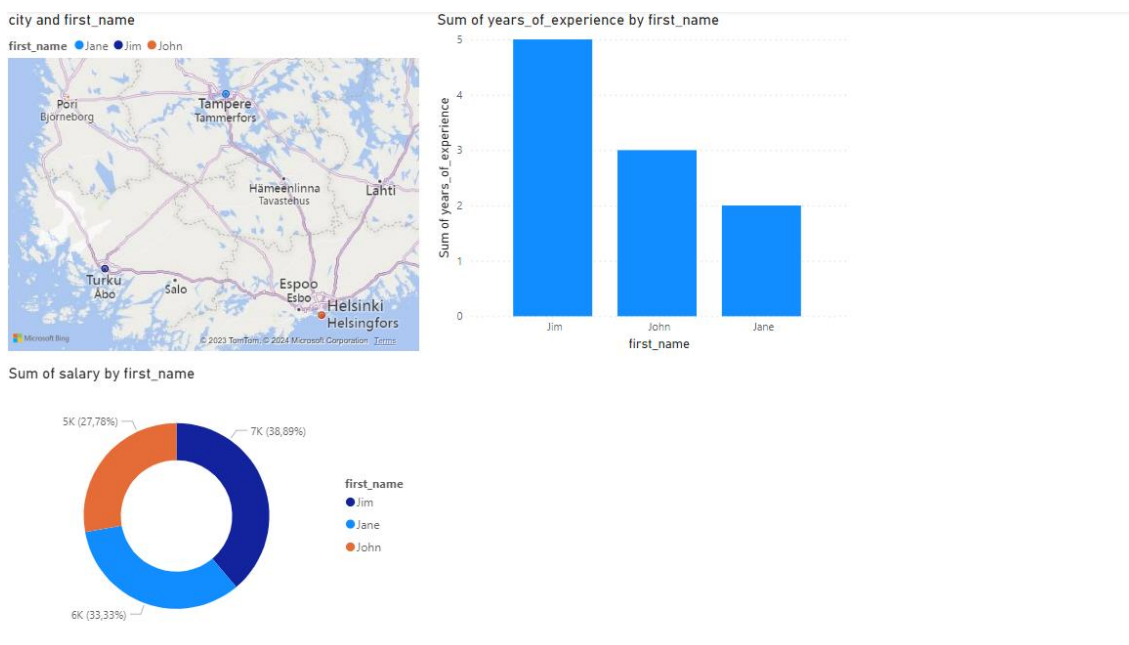
Tietokannat itsessään eivät yleensä natiivisesti tarjoa mahdollisuutta raportoida kattavasti niiden sisältämää dataa. Tämän lisäksi ETL-prosessin tulee olla oikeanlainen, jotta uutta dataa virtaa oikeassa syklissä tietokantaan ja siellä olevat taulut sekä näkymät päivittyvät. Kuitenkin nämä taulut ja näkymät eivät ole parhaassa muodossa, mikäli datasta halutaan saada jonkinlaista raporttia. Tarvitaan siis tauluja tai näkymiä, jotka rikastavat ja yhdistävät fakta- ja dimensiotauluja. Raportointitaulut tai näkymät sisältävät usein tietoa useista eri näkymistä, sekä aggregointeja, jotka helpottavat raportointia. Yleensä ongelma on se, että ihmiset eivät ole hyviä päässään summaamaan ja tarkastelemaan useita rivejä data. Nämä raportointitaulut tai näkymät tarvitsevat siis tuoda sellaiseen muotoon, jossa ihmiset voivat niitä tarkastella. Yleensä tämä tapahtuu visualisoimalla. (Chen ym. 2007.)

7 Visualisointi

Tietovaraston rakentamisen jälkeen yrityksellä tulisi olla yksi selkeä totuudenlähde raporteilleen. Raportissa esitettävät luvut ovat käyneet läpi tarkasti laaditun prosessin, joka toimii tehokkaasti, sekä laadukkaasti. Näin ollen saatuun dataan voidaan luottaa paremmin, kuin tilanteessa, jossa jokainen raportti olisi rakennettu itse ladatusta ja puhdistetusta datasta. (Dong 2023, 34-48.)

Minkälaisia asioita modernin analytiikka-alustan avulla pystytään raportoimaan ja tutkimaan, mikä aiemmin oli haastavaa tai mahdotonta? Yhtenäistetty tietovarasto mahdollistaa sen, että tavallisempien raporttien, kuten myyntimäärän lisäksi voidaan esimerkiksi koneoppimisen tai tekoälyn avulla kehittää kokonaan uusia ja ennustavia raportointimalleja. Tekoälyn avulla yritys voi analysoida esimerkiksi saamiensa palautteiden sävyä ja pyrkiä parantamaan palvelunsa laatua saatujen tulosten perusteella. Tämänkaltaisen datan kerääminen, laadun varmistaminen ja hyödyntäminen olisi ilman modernia tietovarastoa ja analytiikka-alustaa todella työlästä, sekä viriheherkkää. (Soni ym. 2020.)

Raporttia rakentaessa tulee pohtia tarkasta, minkälaista dataa halutaan näyttää. Raportissa on mahdollista tuoda ilmi dataa melko rajattomin vaihtoehdoin. Visualisointikeinoja ja tapoja on useita. Esimerkiksi aiempi kuvio (Kuvio 6) taulu, johon on lisätty henkilöiden kotikaupungit voisi raportissa näyttää seuraavalta (Kuvio 7).



Kuvio 8 Esimerkki PowerBI-raportista

Tämänkaltainen raportti saa tietonsa suoraan tietovarastosta. Mikäli raportoinnissa tarvitaan, voi tämä raportti olla täysin reaaliaikainen tai se voidaan päivittää tietyn ajan välein. Raportti päivittyy aina uusimpaan ja tarkimpaan tietoon automaattisesti. Tämä mahdollistaa yritykselle paremman raportoinnin ja helpottaa reagointia muuttuviin tilanteisiin. (ks.Linstedt & Olschimke, 2015.)

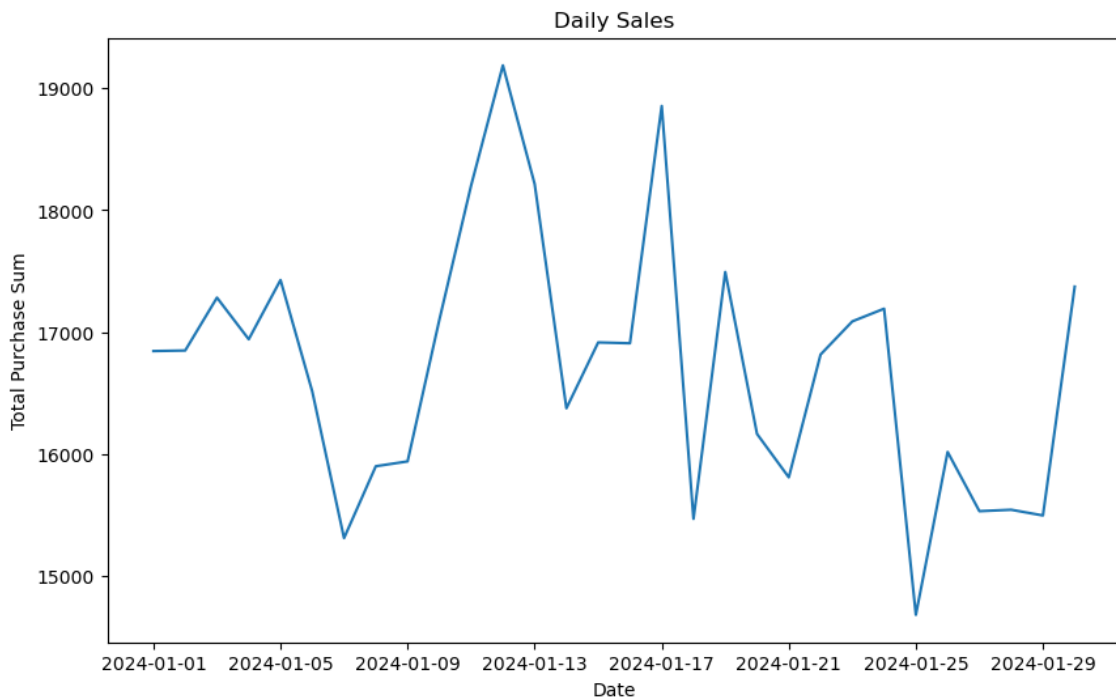
Datan visualisoinnilla tarkoitetaan jonkinlaista graafista esitystä datasta. Johdon raportoinnissa ja tietokannoissa tämä yleensä liittyy siihen, että raportointitaulu halutaan tuoda sellaiseen muotoon, jossa sitä on helppo tulkita ja tarkastella. Esimerkiksi mikäli taulu sisältää paljon rivejä ja sarakkeita ei ihminen siitä yleensä saa itsekseen syvällistä tietoa ulos. Esimerkiksi tietokantataulu (Kuvio 8).

| OrderID | PurchaseSum | Date |
|---------|-------------|------------|
| 1 | 55.332537 | 2024-01-29 |
| 2 | 71.803747 | 2024-01-09 |
| 3 | 60.673574 | 2024-01-28 |
| 4 | 54.943435 | 2024-01-29 |
| 5 | 42.941825 | 2024-01-02 |
| ... | ... | ... |
| 9996 | 55.494235 | 2024-01-23 |
| 9997 | 40.317982 | 2024-01-06 |
| 9998 | 76.084523 | 2024-01-18 |
| 9999 | 3.354955 | 2024-01-15 |
| 10000 | 81.543933 | 2024-01-02 |

Kuvio 9 Tietokantataulu ilman visualisointia

Tässä taulussa on 10 000 riviä ja kolme saraketta, OrderID, PurchaseSum, sekä Date. Mikäli tätä taulua tarkasteltaisiin ilman visualisointia ei siitä välity ihmiselle juurikaan tietoa, sillä datamäärä on niin suuri. (ks.Chen ym. 2007.)

Kuvasta käy oikeastaan vain ilmi rivimäärä ja se, että ostosten summat ja päivät vaihtelevat. Mikäli kuviota visualisoidaan jollain tavalla, on tilanne täysin toinen. Kuviossa 10



Kuvio 10 Tietokantataulu visualisoituna

näkyä selkeästi se, miten myynti vaihtelee kuluva kuukauden aikana. Myynnit on summattu päiväkohtaisesti. Kuun puolessavälissä näkyä piikkiä, kun taas alku ja loppupuolella pientä laskua. (ks.Chen ym. 2007.)

Dataa voidaan tarkastella useasta eri syystä. Visualisoitua dataa voidaan käyttää nykytilanteen hahmottamiseen tai uuden Asian tutkimiseen. Kuvio 9:n perusteella voisi vähittäiskauppa esimerkiksi pohtia sitä, miten myyntiä voidaan lisätä kuun alku- ja loppupäässä. Toisaalta resursseja voidaan myös allokoida tehokkaammin aikoina, jolloin myynti on heikompaa. Tämänkaltainen visualisointi on usein tarkoitettu sidosryhmille, kuten johdolle tai keskijohdolle. Visualisointi on selkeä ja helppo ymmärtää. Analysoinnin jälkeen voidaan pohtia juurisyytä sille, miksi myynti on heikkoa tiettyinä päivinä. (ks.Chen ym. 2007.)

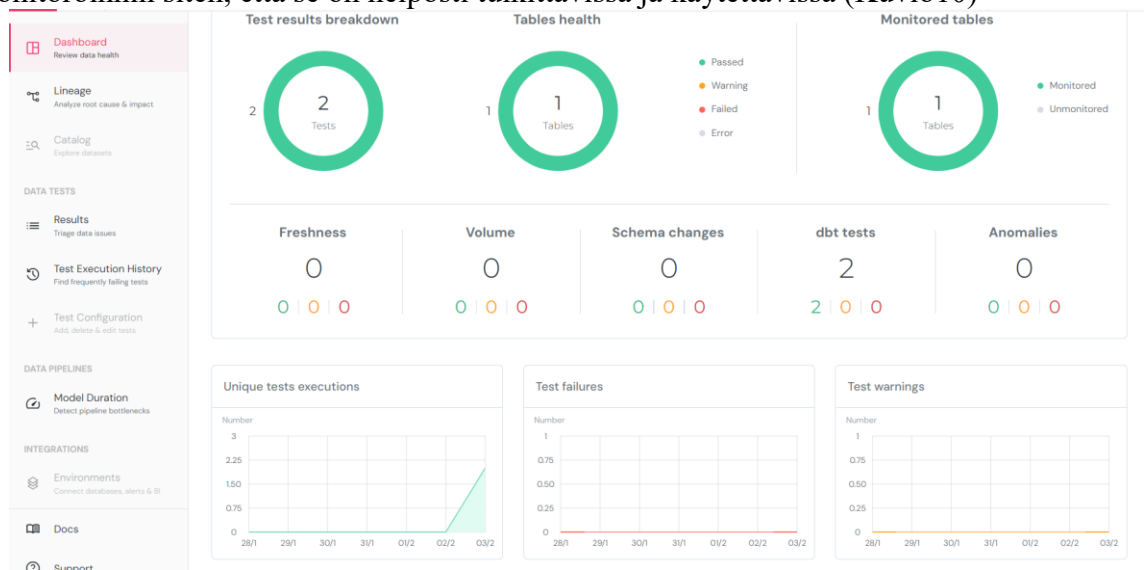
Mikäli data tarkastellaan enemmän tutkivasta näkökulmasta, on tämä paljon yksilöllisempää. Tutkiminen ei välttämättä suoranaisesti tarkoita sitä, että tieto vietäisiin raportille, mutta sen avulla voidaan saada analysoitua uusia tuloksia. Yleensä tämänkaltaiset visualisoinnit saattavat sisältää monimutkaisia lämpötilakarttoja tai muita vaikeammin tulkittavia graafeja. Tällaiset visualisoinnit toteutetaan yleensä tietovaraston kerroksilla, jotka sijaitsevat ennen raportointitauluja, jopa raakadatalla. Yleensä syy on

se, että tämänkaltainen tutkiva visualisointi vaatii suuren määrän tietoa, jota ei ole välttämättä kannattavaa kerätä jatkuvasti. (Chen ym. 2007.)

7.1 Monitorointi

Tietokannoista raportille saatava data ja sen etl-prosessi ovat raportin osia, jotka vaativat monitorointia. Raporttiprosessia suunniteltaessa tulisi selvittää, mitkä ovat sellaisia asioita, jotka ovat äärimmäisen liiketoimintakriittisiä. Tämänkaltaisia aspekteja voivat olla esimerkiksi liiketoiminnan avainluvut, kuten myynti ja kate. Nämä ovat avainlukuja, joiden laadusta ja saatavuudesta on oltava äärimmäisen varmoja. Tämä tarkoittaa sitä, että jonkinlaisia testejä on suunniteltava laadun varmistamiseksi. Toisaalta vähemmän kriittiset raportit vaativat ehkä heikompaa valvontaa. Miten valvonta ja monitorointi toteutetaan? (Linstedt & Olschimke, 2015)

Monet alustat, kuten Azure, tarjoavat omiin ratkaisuihinsa valmiina erilaisia monitorointiratkaisuja. Nämä ratkaisut auttavat suoraan tarkastelemaan datan laatua ja siinä tapahtuvia muutoksia valittujen mittareiden avulla. Syy tähän on se, että mikäli raportoinnissa halutaan käyttää tietokantaa, joka toimii ”yhtenä totuutena”, niin tulee varmistua, että tietokannassa oleva data on kuranttia. Yleensä tällaiset palvelut tarjoavat monitoroinnin siten, että se on helposti tulkittavissa ja käytettävissä (Kuvio10)



Kuvio 11 Esimerkki monitorointidashboardista

Kuvion raportti on elementary-nimisen työkalun muodostama raportti Postgres-tietokannassa, jossa on yksi taulu. Tällainen monitorointi auttaa yritystä paitsi tehostamaan ja optimoimaan raportointia, mutta myös valvomaan sen tietoturvaan, sekä parantamaan luotettavuutta. Yhdessä nämä luovat koherentit kokonaisuuden, joka auttaa luomaan raportoinnista sellaista, joka tukee liiketoimintaa ja jonka laatuun voi luottaa. (ks.Linstedt&Olschimke, 2015.)

8 Empiria

Tutkielman tavoitteena on ymmärtää, millä tavoin modernit analytiikka-alustat ja pilvipalvelut, vaikuttavat johdon raportointiin ja minkälainen on hyvä raportointiprosessi. Kysymykset pohjautuvat teoriaosuuteen ja niiden avulla pyritään luomaan holistinen ja syväluotaava kuva, joka auttaa vastaamaan tähän kysymykseen.

Tutkielma toteutettiin laadullisena tutkimuksena puolistrukturoituna teemahaastatteluina. Haastateltavat löytyivät tutkijan kontakteista eri it-alan yrityksistä. Haastateltavat työskentelevät kaikki eri organisaatioissa. Organisaatiosta yksi on julkinen, yksi yksityinen osakeyhtiö ja kaksi pörssiyritystä. Otannassa pyrittiin haastattelemaan niin raportin käyttäjiä, raportin luoja ja analytiikka-alustan rakentajia. Tällä tavoin haastatteluista pyrittiin saamaan kokonaiskuva yrityksen tai organisaation raportointiprosesseja ja saamaan vastaus tutkimuskysymyksiin. Haastateltavia rajasi tutkimuskysymysten ja substanssin lisäksi seuraavat asiat:

- Työtehtävä
- Käytettävä teknologia

Nämä rajaukset muodostuvat tutkimuskysymysten avulla. Luotettavien ja mielekkäiden tulosten saamiseksi oli yksi kriteeri haastateltavalle työtehtävä. Työtehtävän tulee olla raportointiin liittyvä, sekä kokemusvuosiltaan sellainen, että voidaan olettaa tietotaidon lisäksi myös kumuloitunutta hiljaista tietoa, sekä muita parhaita käytänteitä, joita ei välttämättä teorian puolelta suoraan löydy. Työtehtävän lisäksi rajaus muodostui käytettyyn teknologiaan. Tämä rajaus tapahtui melko itsestään tutkimuskysymyksen kautta. Tutkimuskysymysten kannalta ei ole mieleistä haastatella toimijaa, jonka raportointitavat ja työkalut ovat vanhentuneita. Rajaus tapahtui tässä keskusteluiden ja empirian vertaamista teoriaan. Mikäli näiden perusteella voidaan todeta, että raportoinnissa on moderneja piirteitä, niin myös kohta 2 täyttyy.

Haastateltaviin otettiin yhteydenotto, joko sähköisesti tai perinteisen keskustelun ohessa. Haastatteluun suostuville sovittiin jokaiselle erillinen ajankohta, jolloin haastattelu suoritetaan. Haastattelut suoritettiin Microsoft Teamsin välityksellä. Haastattelun aikana tuotiin ilmi, että kaikki haastattelut ovat anonyymejä. Haastatteluista kerätty data säilöittäisiin luottamuksellisesti ja hyvää tutkielmatapaa noudattaen. Tutkijan pohdinnan

mukaan, anonymiteetti on paitsi tärkeä osa tutkimustulosten kannalta, mutta nimien ja organisaatioiden julkaiseminen ei toisi tutkimukseen lisäarvoa.

Haastatteluja varten tutkija teki haastattelurungon. Tämän haastattelurungon tarkoitus on tuoda haastateltavien mielipiteet ja vastaukset mahdollisimman tehokkaasti esille, kuitenkin johdattelematta. Haastateltaville ei toimitettu kaikkia kysymyksiä, vaan pelkät teemat, joihin varautua. Tällä tavoin haastateltavat eivät valmistaudu liikaa haastattelua varten, jolloin tulokset ovat mahdollisimmat aitoja ja täten tutkimustulokset parempia.

Haastattelut litteroitiin Teamsin tekoälyn avulla sanasta sanaan. Jokaiselle haastateltavalle annettiin mahdollisuus saada litteroitu aineisto itselleen. Haastattelu oli jaoteltu teemoihin seuraavasti:

- Raportointiprosessi ja sen arviointi
- Modernit analytiikka-alustat niiden hyödyt
- Modernit analytiikka-alustat niiden riskit
- Laadunvalvonta raportointiprosessissa

Nämä teemat käytiin läpi haastattelun aikana. Teemoja käsiteltiin kysymyksin, joiden avulla pyrittiin saavuttamaan syvällistä tietoa ja havaintoja, joita haastateltavilla on ollut. Jaottelu on toteutettu siten, että haastattelun alussa kysymykset keskittyvät raportointiprosessin ja sen arviointiin. Tässä pyritään saamaan kokonaisvaltaista käsitystä ja tietoa siitä, millä tavalla raportointiprosessi yleensä etenee, mitä vaiheita siinä on ja millä tavalla prosessia arvioidaan. Toisena teemana on analytiikka-alustojen hyödyt. Tässä osiossa pyritään keskittymään siihen, mitä analytiikka-alustoja ylipäätään nykyään käytetään ja minkälaisia ominaisuuksia niissä on. Seuraavassa osiossa keskitytään riskeihin, joita tämänkaltaisten alustojen käyttöön liittyy. Riskeissä käydään läpi esimerkiksi tietoturvaa ja pohdintaa siitä, että alusten käyttöönoton helpottuessa, mitä asioita yrityksen tai organisaation tulee silti ottaa huomioon.

Riskien jälkeen haastatteluissa pohditaan ja käydään läpi laadunvalvontaa raportointiprosessissa. Esimerkiksi uudet tekoälyratkaisut eivät ole mieluisia, mikäli datan laatu on heikko tai sitä ei valvota. Datan valvonta on usein sivuroolissa, joten tarkoitus on syventyä myös käytännön ratkaisuihin ja prosesseihin. Haastattelut analysoitiin näiden teemojen alla ja löydöt sitoutettiin raportointiin.

Haastatteluiden aikana fokus pidettiin liiketoiminnan ja teknisen osaamisen välillä. Haastatteli antoi viitekehyksen haastateltavalle, jonka sisällä vastaukset annettiin. Tutkimuksen mielekkyyden kannalta haastateltavia ei lähdetty liikaa rajoittamaan, vaan vastaukset pyrittiin saamaan mahdollisimman autenttisesti ja hyvää tutkimustapaa noudattaen.

Haastattelut toteutettiin käsitellen haastattelun alussa sitä, minkälaista on hyvä analytiikka ja BI, josta johdettiin kysymys, että millä tavoin ala on muuttunut viime vuosina tai historiansa aikana. Raportointiprosessia pohdittaessa käytiin läpi myös mitä liiketoiminta saa tästä prosessista, sekä mitä pitää ottaa huomioon, kun luodaan analytiikka-alustaa liiketoimintakäyttäjille. Haastattelun aikana pohdittiin myös mitä mahdollisuuksia ja haasteita modernit analytiikka-alustat tuovat. Tämän lisäksi haastatteluissa käytiin läpi tietoturva ja käsiteltiin sitä, millä tavoin tietoturva näkyy pilvipalveluissa, sekä millä tavoin asiasta nykyään keskustellaan.

| | Työnimike | Julkinen vai yksityinen työnantaja |
|----|-----------------------------------|------------------------------------|
| H1 | Lead Developer | Yksityinen |
| H2 | Lead Developer | Yksityinen |
| H3 | Asiantuntija, Data ja analytiikka | Julkinen |
| H4 | Senior Consultant | Yksityinen |

Taulukko 1 Haastateltavien tittelit ja työnantajat julkinen/yksityinen

8.1 Raportointiprosessi

Haastattelun alussa käytiin läpi sitä, minkälainen ylipäätään on hyvä raportointiprosessi ja BI, sekä mitä se sisältää. Tutkija pyrki myös omilla jatkokysymyksillään selvittämään tarkemmin, millä tavoin haastateltava kokee prosessin, sekä mitä ominaisuuksia tämä näkee siinä olevan.

Jokainen haastateltava mainitsi jossain muodossa sen, että raportointiprosessin lähtökohtana tulee olla jokin liiketoiminnan tarve ja kaiken lähtevän siitä, että tämä tarve kartoitetaan. Mitä oikeasti halutaan ja miten sitä aiotaan käyttää, olivat yleisiä teemoja,

jotka toistuivat haastattelun aikana. Tämä vastuu nähtiin olevan, sekä liiketoiminnalla, että teknologisilla osaajilla.

“No joo se no tilaajan pitäisi pitäisi speksata se se liiketoiminnallinen tarve, että mitä informaatiota se haluaa.” (H3)

Teknologisen osaamisen lisäksi nähtiin tärkeäksi kokonaiskuvan ymmärtäminen ja kartoittaminen. Kokonaiskuvalla kaikki painottivat yhteistyötä eri sidosryhmien välillä. Eri sidosryhmien välinen kommunikaatio auttoi välttämään tilannetta, jossa raportointiprosessi tai raportointi luodaan vain sen takia, että sellainen prosessi pitää olla.

“Joo joo tota no siis mun mielestä kun niinku hyvä olisi jos voidaan lähteä jotenkin liikkeelle siitä, että mikä se tarve on että mikä se niin kun asia on mitä sillä yritetään ratkaista ja mitä sillä halutaan, mitä halutaan että se niinku raportti näyttää käytännössä.” (H1)

Haastattelujen aikana tuli ilmi se, että raportointiprosessi ei ole staattinen kokonaisuus vaan päinvastoin hyvinkin dynaaminen. Raportointiprosessi ei välttämättä ole lainkaan loppuun hiottu heti alussa, vaan se muotoutuu tekemisen myötä. Tämä edellyttää järjestelmältä paitsi joustavuutta, mutta myös resilienssiä muuttuvia dataja ja tarpeita kohtaan. Nämä muuttuvat datat ja tarpeet käsittivät paitsi lähdejärjestelmien muutokset, mutta myös muuttuvat liiketoimintatilanteet ja perinteiset raportointitarpeet. Modernin raportointiprosessin on oltava, sekä vakaalla pohjalla, mutta myös kykenevä muuntautumaan tarvittaessa.

“Raportointiprosessi niin sehän on semmoista niinku vähän niinku jatkuvaa kehittämistä, että sen takia se ei saa olla sellainen niinku pistekohtainen vaan vaan tota noin niin sellainen että siellä on niinku laajasti dataa tarjolla ja ja tota noin niin asiat on niinku tehty tehty silleen että niitä on helppo ylläpitää ja ja korjata ja muuntaa ja ja ja sitten jos jos tota noin niin raportit prosessi edellyttää että jotain uutta lisätietoa tarvitaan sinne niin se ei sitten.” (H2)

Puhuttaessa hyvästä raportointiprosessista, jokainen haastateltava mainitsi yhtenä vahvana vaatimuksena ja tekijänä raportin reaaliaikaisuuden tai ainakin mahdollisimman lähelle sitä olevan teknisen ja liiketoiminnallisen toteutuksen. Reaaliaikaisuudessa nähtiin tuovan paitsi mahdollisuuksia, mutta myös haasteita. Käytännön tasolla vaatimus

liittyi pitkälti siihen, että liiketoimintaympäristö on muuttunut nopeammaksi ja reagointiaika eri tilanteisiin on lyhentynyt.

Tutkijan kysyessä jatkokysymyksiä kävi selkeästi ilmi, että suurin osa nykyisistä organisaatiosta uudistaessaan raportointiaan modernimpaan pitää reaaliaikaisuutta erittäin tärkeänä ominaisuutena. Muutaman tunnin tai jopa päivän viive ei enää riitä, vaan tuore tieto pitää olla saatavilla heti.

“Vedetään kaikki tota asiakkaat minkä kanssa raporteista on puhuttu niin aina kaikki haluaa että ne on käytettävissä heti eikä niinku tyyliin huomenna tai tunnin päästäköön ja, että raportit mitä, mitä siitä, mitä tästä tehdään iin pitää olla käytettävissä mahdollisimman nopeasti.”(H4)

Hyvän raportointiprosessin kannalta nähtiin myös tärkeäksi se, että raportti on helppo ylläpitää. Helpon ylläpidon lisäksi myös raportointiprosessin kehittämisen ja analytiikka-alustan käytön tulisi olla mahdollisimman helppoa ja suoraviivaista. Toisaalta muutosten tekeminen ei saisi olla liian aikaa vievää, sillä dataa tarvitaan jatkuvasti nopeammalla tahdilla. Liiketoiminnan tarpeet nähtiin lisääntyneeksi ja muutokset enemmän sellaisiksi, että niitä tulee varmasti. Tämä toisaalta pätee Agilen periaatteisiin, jossa pyritään saamaan MVP (Minimum Viable Product) ulos mahdollisimman nopeasti, jonka jälkeen jatkaa kehitystä.

“Me ei laiteta koko arkkitehtuuria tai koko prosessia uusiksi sen sen pienen muutoksen takia, vaan se saadaan niin kuin niin kun ne muutkin uudetkin tiedot sieltä mukaan. Tai jos me tarvitaan se tieto vähän niinku nopeammalla tahdilla tai tai jotain muuta vastaavaa, niin niin se tavallaan se prosessi on rakennettu niin robustiksi että se tota noin niin asiat saa tota noin niin organisoitua vaikka vaikka muutoksia tulee koska niitä tulee”(H2)

Tutkija huomasi haastatteluissa, että yhteinen näkemys raportin, analytiikka-alustan ja koko prosessin helppokäyttöisyydestä oli sekä teknisemmällä henkilöllä, kuin enemmän liiketoimintaosaajien näkökulmasta. Monesti johto arvostaa yksinkertaisuutta myös raporteilla, eikä välttämättä monimutkaisia kaavioita, tai mallinnuksia. Tämä voi olla jossain määrin ristiriidassa sen kanssa, että dataa ja tietoa on kasvavissa määrin, mutta raportit halutaan mahdollisimman yksinkertaisina. Välttämättä tämä ei ole huono asia, vaan kyse on enemmänkin siitä, mikä on yksinkertaista. Tiedon visualisointi tai sen

esittäminen myös korostuu. Johto todennäköisesti ei tarkastele dataa tai raporttia, mikäli se on esitetty huonosti, tai liian monimutkaisesti. Tämä voi tarkoittaa esimerkiksi sitä, että monimutkaisesta analytiikasta esitetään lähtökohtaisesti pelkkä tulos. Se missä tekniset ihmiset joutuvat siivoaman enemmän raportejaan, on myös johdon väistämättä tultava vastaan siinä, minkälaista dataa osataan tulkita raporteista.

“Ja sitten varmaan jos ajattelee niinku raportointia ihan ylimmän johdon näkökulmasta niin sitten semmoinen että niitä saa niinku helposti ettei tarvitse osata oikeastaan mitään että ne tulee ikään kuin viran puolesta. Sitten sinne johtoon jotain yhteenvetoraportti ja muita tai sitten joku dashboard josta johto voi käydä katsomassa yksinkertaisia liikennevaloja.”(H4)

Yksi haastateltavista toi kuitenkin ilmi sen, että vaikka analytiikka-alusta olisi rakennettu miten helpoksi ja hyvin toimivaksi, ei liiketoimintaa silti välttämättä kiinnosta sen toiminta, vaan pelkästään lopputulos eli raportti. Liiketoiminnan näkökulmasta tämä on ymmärrettävää, sillä raportti nähdään osittain esimerkiksi jonkin järjestelmän sivutuotteena, eikä itseisarvona. Raporttien tekemistä ei nähty ydinosaamisena, vaan erillisenä osana, ja jopa “likaisena työnä”. Toisaalta haastatteluiden aikana kävi ilmi myös se, että IT-osasto ei voi yksin toteuttaa raportointiprosessia tai muutoksia siihen, vaan vaatii jatkuvaa kommunikaatiota ja yhteiseloä liiketoiminnan kanssa.

“Jep sitten voi niinku ehkä miettiä sitä, että joku muu voi olla parempi. Tekee niitä niitä tota sitä varsinaista ikään kuin likaista työtä siellä taustalla ja sitten itselle jää aikaa. Sitä yrittää selvittää mihin se niinku liittyy bisneksen näkökulmasta.”(H4)

8.2 Raportointiprosessin muutos

Tutkija pyrki tutkimuksen aikana myös kartoittamaan millä tavoin raportointiprosessi ja siihen liittyvät analytiikka-alustat ovat muuttuneet haastateltavien työuran aikana. Pyrkimyksenä oli käsitellä teknologioiden lisäksi myös muuttuvia liiketoiminnan tarpeita ja mahdollisuuksia.

Haastateltavilla oli erilaisia näkemyksiä siihen millä tavalla prosessi itsessään oli muuttunut ja millä tavoin käytettävissä olevat työkalut olivat kehittyneet. Kuitenkin kaikilta yhtenäisenä vastauksena kävi ilmi se, että siirtyminen pois omista servereistä ja palvelimista kohti pilveä on hyvässä vauhdissa. Pilvisiirtymää ei sen sijaan nähty itseisarvona, vaan hybridistrategiaa myös korostettiin, jossa yrityksellä on oman fyysisen

serverin lisäksi pilvipalvelu käytössä. Pilvisiirtymä nähtiin myös yhtenäistävän toimintamalleja ja muuttavan esimerkiksi ETL-mallin enemmän ELT-malliin, jossa data ensin ladtaan jonnekkin alustalle ja sitten transformoidaan.

“Niin kun aikaisemmin tuli vielä niinku on premiratkaisuja tosi paljon enemmän vastaan, että kaikki se niinku datansiirto tehtiin on premis maailmassa ja työkalut oli ehkä vähän erilaisia että tehtiin jotain SSIS-paketteja ja tota transformaatiota ja ehkä niinku se malli oli just se niinku enemmän ETL malli, että eikö transform ja sitten load tyyppisesti ja sitten tota niin sitten nykypäivänä kun ollaan enemmän siirretty tuonne pilvien alustoille niin sitten on usein se malli että lähdejärjestelmistä vaan niin kun napataan se data sellaisena minä se on ja sitten siirretty enemmän tämmöiseen ELT malliin, että extract load ja transform vasta sitten sen jälkeen kun ollaan saatu se sinne niinku pilvin alustalle se data.”(H4)

Haastateltavilta tuli myös havaintoja siitä, että monet asiat ovat kuitenkin pysyneet melko samana vuosien aikana. Tietokantojen toiminta nojaan edelleen vahvasti relaatorakenteisiin ja sen lainalaisuudet pätevät edelleen uusissakin tietovarastointiratkaisuissa. Myöskään datamäärien ei nähdä välttämättä kasvaneen itseisarvona kaikissa toteutuksissa, vaan nekin ovat pysyneet osittain samana. Lähdejärjestelmät ovat edelleen toiminnoiltaan samankaltaisia, vaikka ne tarjoisivat uusia ratkaisuja.

“ Tää on ihan samanlaista kun mä oon aloittanut tietovarastointinnit niin 2004 vuonna niin niin eihän tää nyt itse asiassa siitä ole hirveästi hirveästi muuttunut muuttunut ne perusasiat että että tota noin niin useimmissa tapauksissa ei ole välttämättä edes data volyymit kasvanut tuommoisessa niinku raportoinnissa juurikaan ja ne asiat ne on ne taloushallinnon myynnin hankinnan, varasto logistiikan järjestelmät mistä sitä tietoa haetaan” (H2)

Muutoksina nähtiin kylläkin se, että esimerkiksi kehitys-, testi- ja tuotantoympäristöt tulivat erillisinä mukaan analytiikka-alustoille ja raportointiprosessiin. Näiden periaatteiden avulla saatiin paremmin toteutettua jatkuvaa kehitystä ja jatkuvaa integrointia.

“Se on tietysti hyvä asia ja tota tietoon pääsee helpommin käsiksi, koska se on siellä niinku pilvessä saatavilla ja jaettavissa muille organisaation ulkopuolisille käyttäjille

että tota noin niin sitten jos me ajatellaan niinku hallinnan kannalta tai näin niin niin ei siellä ole aika ihan aikaisemmin niinku tämmöistä tietovarastointi ympäristössä niin ei välttämättä ollut mitään kunnollisia kehitys-, testi- ja tuotantoympäristöjä.”(H2

Muutokset ovat koskeneet myös käyttäjiä. Se missä ennen analytiikka kuului vain pienelle osalle työntekijöitä, on vastuu ja valta nykyään jakautunut. Analytiikka-alustat tehdään liiketoimintakäyttäjät mielessä pitäen ja “black-box” tyyppistä ajattelua, jossa tietovarasto on musta laatikko, jonka toiminnasta tietää vain asiaan pitkälti vihkiytynyt työntekijä on siirrytty kohti tilannetta, jossa liiketoimintakäyttäjät ovat iso ryhmä nykyisiä käyttäjiä.

“Että musta silloin joskus ennen vanhaan kaksituhattaluvun alussa niin ne käyttäjät oli ikään kuin niitä tietokanta asiantuntijat jotka sitten teki niitä vaikka jotain staattisia raportteja sitten johdolle siitä mitä ne sai. Mut nytten niinku käyttäjät on enempi sitten suoraan jo liiketoiminta ihmisiä.”(H4)

Eroja nähtiin myös työkaluissa ja niiden tarjoamissa mahdollisuuksissa. Aikaisemmin työkalut ovat nähty enemmän rajoittavina tekijöinä tai ainakin niiden ominaisuudet on koettu tällaisiksi. Haastatteluiden aikana kävi ilmi, että moderneilla analytiikka-alustoilla ja pilvipalveluilla on tarjota käytännössä rajattomasti eri resursseja ja palveluita. Kyse on siis enemmän budjettikysymyksestä ja oikeasta työkalun valinnasta, eikä niinkään siitä, että jotakin asiaa ei ole mahdollista tehdä. Tärkeää on myös pohtia se, mikä työkalu tilanteeseen on järkevin ja helpoin hallita.

“Niinku analysoimaan ja yhdistelemään ja sitten on työkaluja on suuri määrä on eri BI työkaluja ja sitten vähän muitakin missä niitä voi käyttää. Ja niinku nyt voi nyt on se tilanne että aina kun joku käyttötapaus tulee niin voi valita aina millä työkalulla ja kannat siinä tehdä kun ennen se oli silleen että mietittiin että onko mitään miltä voi tehdä ja sitten ehkä oli joku ja sitten se tehtiin sillä mutta nykyään se on aina niin päin että on todella monta eri vaihtoehtoa joista pitää valita että mikä niistä on se hallittaviin”.(H3)

Muutoksia nähdään myös siinä, millä tavoin raporttia kehitetään. Ohjelmistokehityksestä tuttuja piirteitä, kuten versionhallinta ja agile ovat molemmat asioita, joita nykyään hyödynnetään pitkälti myös raportinkehityksessä. Tämän nähdään lisäävän raportin luotettavuutta, prosessin joustavuutta, sekä yleisesti parempaa kokemusta koko raportointitiimille. Yksi iso muutos on tullut siihen, että kehittäjiltä vaaditaan entistä

enemmän taitoja. Yhden taidon osaaminen hyvin tai todella hyvin ei enää välttämättä riitä, vaan on osattava useita taitoja ja kehityttävä ajan trendien mukana sopivissa määrin.

“Tavallaan siihen siihen kehittämiseen on tullut ehkä enempi sellaisia niinku sovelluskehityksestä tuttuja käytäntöjä piirteitä, että että tota on ehkä ehkä lähentynyt lähentänyt vähän sitä sitä maailmaa mitä sovelluskehitys on. Miten se sovelluskehitystä tehdään ja ja tota noin niin niinku kehittäjän näkökulmasta, niin niin niin kun semmoinen muutos on tietysti se, että että että se ei ole pelkästään se sql-server pömpeli minkä ympärillä ollaan vaan ja sql kieli vaan tota noin. Kehittäjän pitää niinku osata paljon muutakin muutakin tuota noin niin asioita mitkä liittyy siihen siihen sen niinku ympärilleen. Eli tota noin niin vaikka pythonia on hyvä osata ja ja ja devopssia ja tota noin niin kaikkia muita palveluita jotka joita kautta sitten sitä tietoa integroidaan esimerkiksi.”(H2)

8.3 Modernit työkalut ja niiden vaikutus

Modernit analytiikka-alustat, pilvipalvelut ja raportointityökalut vaikuttavat merkittävästi siihen, minkälaista raportointia yritys voi tehdä, tai ylipäätään minkälainen raportointi olisi optimaalista missäkin tilanteessa. Tutkija pyrki haastatteluiden aikana kartoittamaan haastateltavien näkemyksiä siitä, minkälaisia hyötyjä he kokivat näistä työkaluista olevan liiketoiminnalle ja millä tavoin he konkreettisesti ne näkivät.

Yksi suurimmista hyödyistä moderneissa työkaluissa liittyi siihen, että reaaliaikaisen datan lisäksi ne mahdollistavat nopean kehityksen. Kehitystarpeita tulee jatkuvalla syötöllä, joten myös työkalujen tulisi olla sellaisia, että ne kykenevät hallitsemaan tämänkaltaisen tarpeen.

“Että kun käyttäjä käyttää jotain raporttia, niin sehän ei siitä näe että miten se on tehty vaan se vaan näkee että miten se toimii sillä sillä hetkellä, niin se ei sinänsä ole väliä hänelle, että kuinka työläästi se on tehty jos se toimii. Mutta sitten kun nykyään niitä tarpeita tulee hirveätä vauhtia joka suunnasta niin siinä se on sitten aika tärkeää se nopeus että kuin helposti niitä pystytään tekemään kun mitä nopeammin pystytään tekemään, sitä enemmän pystytään tekemään ja sitä enemmän saadaan informaatiota sitten.”(H3)

Konkreettisia hyötyjä myös nähdään siitä, minkälaisia raportteja pystytään tuottamaan. Nykyaikaisissa tietovarasto- ja tietoallasratkaisuissa tiedon säilytys ja laskentateho on

eriytetty erilliseksi. Käytännössä tämä tarkoittaa sitä, että tieto on fyysisesti jossain muodossa esimerkiksi tietoaaltaalla ja tarpeen tullen käynnistetään tietokone, jossa on tarvittava määrä laskentatehoa tämän tiedon tutkimiseksi tai raportoimiseksi. Modernit ratkaisut kykenevät myös analysoimaan esimerkiksi kuvia ja videoita, jotka aiemmin olivat huomattavasti vaikeampia tai jopa mahdottomia toteuttaa. Myös erilaiset aikasarja-analyysit ovat moderneilla alustoilla parempia toteuttaa alustan eri alueisiin tapahtuvan optimoinnin takia.

“Kun meillä on se storage-layer siellä erikseen, niin käytännössä se mahdollistaa niin kun myös kaiken niin kun muun tyyppisen datan tuomisen, että on se sitten kuvaa tai jotain muuta. Toki sitten se niinku että miten sen tyyppistä dataa hyödyntää raportoinnissa, niin se on sitten eri asia että mikä on järkevää ja mikä ei mutta tota, mutta se on niinku mahdollista ja tota just esim nää aikasarja asiat on ehkä niinku aikaisemmin se ei ole ollut ihan se mihin niinku joku sql tietokantaa esimerkiksi on optimoitu.”(H1)

Haastatteluiden aikana korostuu myös se, että historiatietojen tutkimisesta pyritään siirtymään enenevässä määrin kohti ennustavaa raportointia, jossa pyritään analytiikan avulla ennustamaan esimerkiksi tulevia trendejä. Ennustettavuuden lisäksi yhtenäinen tietovarasto mahdollistaa sen, että aiemmin ad-hoccina tapahtuvat raportoinnin ja laskennat erilaisissa taulukkolaskentaohjelmissa siirtyvät enemmän tietovaraston puolelle, jolloin ne tulevat kaikille saataviksi, eikä pelkästään yhden ihmisen työasemalle. Etuja tässä on paljon. Tiedon laatu on parempaa ja sen validointi helpompaa. Myös avoin raportointi parantaa innovointia yrityksen sisällä.

“Ehkä ennustamiseen liittyviä asioita pystyy toteuttamaan semmoisella analytiikalla, että ne on niinku sellaisia asioita, mitä sitten aikaisemmin on joutunut sitten joku joku asiantuntija sitten ehkä pähköilemään tai tekemään jossakin tavallaan sen järjestelmän ulkopuolella jollain excelillä tai jollain muulla ja ja tota noin niin ja sitten niinku jalostaa sitä tietoa, että että asioita voi automatisoida.”(H2)

Yksi selkeä ja konkreettinen suuntaus mitä haastatteluiden aikana käytiin läpi, oli se, että liiketoiminta tuodaan entistä lähemmäs raportointia. Tällä ei välttämättä tarkoiteta sitä, että liiketoiminnan olisi pakko tehdä omat raporttinsa, vaan enemmänkin se, että mahdollisuus datan tutkimiseen ei ole pelkästään analytiikan parissa työskentelevillä ihmisillä, vaan kaikille, joilla on oikeus tarkastella dataa.

“No siis varmaan ehkä semmoinen niinku toinen trendi suuntaus on on mikä on, niin on se niinku datan mikä se oikea termi on demokratisointia sillain että niinku se tuodaan lähemmäksi sitä myös sitä business käyttäjää ja melkein kaikki työkalut ainakin yrittää tehdä sitä helpommaksi että myös niinku bisneskäyttäjät pääsee lähemmäksi niin koska kuitenkin niinku asiat vie aikaa asioiden toteuttaminen vie aikaa niin yritetään ainakin niinku vähentää sitä että sitten ollaan niinku kiinni siitä jostain tietystä.” (H1)

Puhuttaessa Big Datasta ja ylipäätään siihen liittyvistä haasteista yleinen näkemys oli se, että Suomessa hyvin harva yritys hyödyntää todellista Big Dataa. Haasteena ei niinkään nähty olevan iso datamäärä, vaan enemmänkin useammat datatyypit, kuten erilaiset taulukkomuodot, kuva ja videot. Näiden kaikkien eri datatyypien monimuotoisuus tuo vaikeutta paitsi niiden hyödyntämiseen ja varastointiin, mutta myös raportointiin. Laskentateho nähtiin samankaltaisena kuin varastoinnissa yleisesti, eli sitä voidaan lisätä käytännössä rajattomasti ja isoin rajoittava tekijä on budjetti.

“Ja sitten tota no mä nyt mä en nyt tiedä että onko tää nyt sitten suomen mittakaava että en mä nyt tiedä että missään firmassa vielä ihan hirveän big dataa on on niinku tällainen rehellisesti riippuu toki sitten niinku että mistä aspekteista niinku big dataa katsotaan yks on toki se niinku datan määrä, mutta toinen tää spekti on se niinku datan monimuotoisuus ja mun mielestä itse asiassa sen monimuotoisuus on se että mikä tuo enemmän haasteita. On se määrä itsessään, että laskentatehoa saadaan toistaiseksi ainakin vielä niinku lisättyä monen firman tarpeisiin ihan riittävästi ja ja niin kun jotain orkesterointi syklejä voidaan vetää tiukemmalle, jotta niinku saadaan sitä dataa siirrettyä ja näin edespäin, että niinku sitten enemmän mistä ne haasteet tulee, on just ehkä sitä datan monimuotoisuudesta ja miten sitä dataa saadaan. Itse asiassa hallittua, että niinku kenellä on oikeudet.” (H1)

Hyvinä puolina nähdään myös uusien trendien tunnistaminen ja yleisesti ajan tasalla pysyminen yrityksen toiminnasta. Vanhemmissa järjestelmissä kankeus ja hitaus on aiheuttanut sitä, että informaation lopputulos on ollut lähes aina reaktiivista. Jokin tieto on saatu, johon on reagoiduttu, koska asia on jo tapahtunut. Sen sijaan modernien järjestelmien tuoma reaaliaikaisuus ja helppokäyttöisyys mahdollistaa sen, että tulevaisuutta voidaan pyrkiä ennakoimaan.

“No vaikka tyylisiin jonkun laskutuksen kehittyminen tai sitten jos katsotaan jostain no ulkopuolelta jotain isompia trendejä, että miten mihin suuntaan on menossa vaikka vaikka talous tai joku toimiala tai muuta.”(H4)

Modernit analytiikka-alustat ja pilvipalvelut luottavat pitkälti keskitettyihin tietovarastointiratkaisuihin, joissa lähdejärjestelmien tiedot kerätään yhteen paikkaan, josta niitä voidaan hyödyntää. Tämänkaltainen tiedon kerääminen vaatii aina teknisten prosessien lisäksi myös tietoa ja informaatiota siitä, minkälaista dataa tietoaaltaalle, tai varastoon ollaan tuomassa. Näin saavutetaan tilanne, jossa loppukäyttäjän ei tarvitse manuaalisesti yhdistellä ja kerätä eri tietoja, vaan ne löytyvät helposti yhdestä lokaatiosta. Tällöin tiedon laatu paranee ja herkkyyks virheisiin vähenee. Usein tämänkaltainen lähestymistapa on vaikeaa tai jopa mahdotonta.

“On se, että että saadaan semmoista dataa näkyväksi, mitä ei ehkä niinku ole pystytty tuoda näkyväksi ja tuoda sellaista dataa. Yksi sellaista dataa yksi mikä on ehkä eri järjestelmissä ja jos sun pitäisi jotenkin niinku manuaalisen keinoin saada joku yleiskuva jostain tilanteesta, niin se tarkoittaisi sitä että sä käyt jotenkin tekee jotain dataa eksperttejä manuaalisesti ja sitten sä ehkä pyörittelet niitä excelissä ja saat jonkun graafin aikaiseksi.”(H4)

8.4 Tiedon validointi

Datan validoinnilla tarkoitetaan tiedon oikeellisuuden tarkistamista. Tämä on osa raportointiprosessia. Tiedon validoinnista kysyttäessä tutkija pyrki kattavasti tarkastelemaan sitä, mitä aspekteja tähän kuuluu. Hyvin nopeasti muodostui näkemys jokaisen haastateltavan datan perusteella, että vaikka tietoa validoidaan sitä tehdään edelleen todella vähän. Yksi haastateltava mainitsi, että hänen mukaansa yritykset ovat vasta viime vuosina heränneet siihen, että vaikka tietojärjestelmät olisivat ajan tasalla, niin niissä oleva data saattaa olla laadultaan todella heikkoa. Yksi selkeä aspekti tähän on myös se, että yritykset ovat ymmärtäneet datan validoinnin olevan tärkeä osa, joka vie aikaa ja resursseja, kuten mikä tahansa muukin liiketoiminnan prosessi.

“Siihen ollaan herätty nyt monessa paikassa tällä hetkellä, että sanotaan joitain vuosia sitten niin datan laatuun niinku heräsi ensimmäiset edistyneet firmat. Mutta nyt tällä hetkellä niinku tuntuu että, monessa paikassa aletaan ymmärtää, että että meidän data ei olekaan hyvässä kunnossa ja että se vaikuttaakin aika paljon kaikkeen ja että sen

kunnostaminen ja ylläpito vaatiikin oikeasti merkittävää panostusta. Vähän niinku, että datak on asetti niin sellainen ajattelu niin nyt alkaa alkaa valuu tonne kentälle, että sitten aletaan ymmärtää mitä se tarkoittaa. Muutamia vuosia sitten edistyneimmät sanoi sitä, mutta muut eivät ymmärtäneet, että se tarkoittaa niin. Niinku mun mielestä tai mun kokemuksen perusteella niin semmoista datan laadun mittarointi ja tehdään vielä tosi vähän että sitä pitäisi ja kannattaisi tehdä ja kehittää.”(H3)

Yksi ongelma validoinnissa on se, että lähdejärjestelmät eivät itsessään kontrolloi niistä saatavaa tietoa juurikaan. Tämänkaltainen erhe saattaa aiheuttaa sen, että loppuraportilla näkyvät graafit ja luvut eivät pidä paikkaansa, jolloin liiketoiminta tekee päätöksensä virheelliseen tietoon perustuen. Toisaalta ongelmana on myös se, että selkeitä prosesseja laadunvalvonnalle ei ole. Vaikka pinnalla olevat asiat, kuten Data Mesh tai muut liiketoimintaa osallistavat aspektit raportointiprosessissa ovat osa modernia tiedonkäsittelyä, ei näistä huolimatta teknisen toteutuksen ja liiketoiminnan välinen kommunikointi tapahdu saumattomasti.

“Eli tota noin niin sitten kun jotain asioita menee rikki niin sitten sitten menee tai jotain asioita näkyy raportilla että täällä on on tota noin niin virheellistä tietoa niin jaa nää loppujen lopuksi niinku hirveän paljon ei tehdä semmoista niinku datan validointia ja ja se onkin mun mielestä sellainen mitä mitä kannattaisi tehdä paljon enempi näissä dataprojekteissa että että.Tietolähteitä voi olla sitten niinku tosi iso määrä ja ja tieto voi siellä tieto lähteessä olla. Voi olla järjestelmiä jotka ei niin tarkkaan kontrolloi sitä niin kun tiedon syöttöä niinku arvoja mitä syötetään sinne tietoa että sieltä kuitenkin niinku voi puuttua tietoa ja näin pois päin”. (H2)

Haastavan aspektin datan validointiin ja yleisesti laadunvalvontaa tekee myös se, että datamäärät ovat suuria ja niiden valvonta vaatii automatiikkaa. Tämänkaltaisia automaattisia työkaluja on, mutta niiden hyödyntäminen on vielä melko vähäistä.

“Kyllä siis mä varmaan tota no ihan hirveästi en ole vielä nähnyt niin kun tai siis työkaluja on olemassa, mutta hirveästi en ole niinku toteutuksia nähnyt niistä vielä niinku reaali maailmassa itse, mutta jos niin kun puhutaan siitä että dataa on paljon, niin se luultavasti tarkoittaa sitä että ei sitä dataa laatua voida valvoa manuaalisesti, jotenka silloin tarvitsee löytää niitä myös niinku niitä prosesseja siihen että”.(H1)

Toinen selkeä haste on tutkimuksen aika selkeästi esiin tullut havainto siitä, että vaikka liiketoiminta on enenevässä määrin mukana analytiikan käyttöönotossa ja sen suunnittelussa, niin parannettavaa silti on. Tiedon validointi ja sen oikeellisuus liittyy usein siihen, että liiketoiminnan tai jonkinlaisen tuoteomistajan tulisi ilmoittaa selkeät ominaisuudet tai raja-arvot tietyille datalle. Esimerkiksi tietystä järjestelmästä saatava tieto tulisi aina olla positiivisia kokonaislukuja. Tämänkaltaiset validoinnit jäävät hyvin usein haastatteluiden mukaan melko vähäiseksi osaksi koko kokonaisuutta.

“Ja mutta siinä niinku tarvitaan sitä niinku bisneksen inputtia siihen, että miten me voidaan tietää, että onko se data oikealla laista, että välttämättä jos on joku joku henkilö joka tulee jotain koodaamista sinne tekemään niin se voi päätellä siitä datasta jotain ja tehdä jotain tulkintoja. Mutta voiko se olla? Voidaanko olla varmoja että se on oikein mitä niinku tulkitsee sillain jos ei ole sitä niinku data-osaamista niin niin tota siinä niinku vaaditaan sitä yhteistyötä ja niinku tota kommunikointia siitä, että löydetään ne että mitä mitä ne asiat on, mitä pitää oikeasti sitä datan laadun laadusta tsekata.”(H1)

Liiketoiminnan näkökulmasta tilanne tuntui käytännössä samalta, eli ristiriitaa kokemuksesta ei ollut. Osallistuminen prosessiin koettiin melko vähäiseksi ja selkeää syytä tähän ei ollut. Raportoinnin tulokset otettiin annettuna, vaikkakin mainittiin, että osa henkilöistä on asemassa, joissa dataa voidaan kyseenalaistaa. Selkeää prosessia tai toimintatapaa siihen, että miten tämän kyseenalaistamisen tulisi tapahtua ei käynyt ilmi.

“Luulen, että se on aika vähäistä, että sitten kun tavallaan se järjestelmä on pystyssä ja sieltä saadaan sitä dataa, niin varmaan paljon otetaan niinku annettuna että OK, tämmöistä on. Mutta kyllä mullakin on niinku silleen muistissa, että sitten on tiettyjä tiettyissä paikoissa on ihmisiä jotka niinku kyseenalaistaa sitä sitä tietoa jota saa ja se on tietysti ihan tervettä että että tota ne nekin voi olla virheellisiä sitten loppujen lopuksi ne.”(H4)

Kommunikointi ei ole ollut ainoa syy datan validoinnin heikkoudelle, tai sen puuttumiselle. Enemmän kyseessä on tilanne, jossa liiketoiminta ei ole tarpeeksi investoitunut raportointiprosessiin, tai jos on, niin mukana olo tapahtuu liikaa raportin visualisoinnissa ja kuvioissa, eikä niinkään alkupäässä, jossa data tuodaan yhtenäiseen lokaatioon. Toisaalta mikäli puutteita havaitaan, niin niitä on yleensä lähdetty tutkimaan. Yhtenä mainittavana asiana on liiketoiminnan intuitio. Liiketoiminta tietää minkälaista

dataa järjestelmistä tulee ulos ja raportille. Tämänkaltainen näkemys pitäisi kyetä kommunikoidaan selkeästi, mikäli datan validointi halutaan automatisoida.

“No siis varmaan semmoisissa tilanteessa missä tulee niinku nähdään selvästi. Nyt on joku ongelma niin sitä aletaan niinku pureutumaan siihen että mistä tää johtuu niin sitten sitten helposti voidaan myös lähteä niinku tutkimaan sitä että onko tää lähtödata edes edes niinku kuranttia, että ehkä se on enempi just sen tyyppisissä tilanteissa että että on näky löydetään jotain selittämätöntä tai semmoista mitä ei haluta haluttaisi nähdä niin sitten sitä lähdetään tutkimaan.”(H1)

Selkeä näkemys siitä, että pelkkään intuition validoinnissa ei voi luottaa käy myös ilmi haastatteluista. Mikäli vika huomataan vasta siinä kohtaa, kun tietoa aletaan hyödyntämään esimerkiksi johdon raporteissa, voi olla jo liian myöhäistä. Monitorointi koetaan siis pakolliseksi, mutta yleisesti siihen panostetaan liian vähän.

“Mutta pelkästään sen (intuition) varaan ei missään nimessä voi jättää, koska virheitä voi olla vaikea huomata. Sieltä saattaa olla, että luvut vääristyy jonkun datavirheen vuoksi, mutta sitä ei ymmärretä ja sitten tehdään johtopäätöksiä virheellisen datan perusteella niin sen takia mun mielestä on todella tärkeitä, että rakennetaan noita datan laadun mittarointi ratkaisuita.”(H3)

8.5 Tietoturva

Haastatteluiden aikana pyrittiin tarkastelemaan myös kriittisesti modernien analytiikka-alustojen ja pilvipalveluiden aiheuttamia mahdollisia haasteita, mahdollisuuksia ja ominaisuuksia liittyen tietoturvaan ja pääsynhallintaan raportoinnissa. Tietovarastoissa dataa kerätään yhtenäiseen lokaatioon, jolloin tietoturva muodostuu erittäin tärkeäksi aspektiksi. Kaikilla ei välttämättä tule olla pääsyä kaikkeen dataan, ja osa datasta, kuten henkilötiedot voidaan joutua muuttamaan tunnistamattomaan muotoon. Haastatteluiden aikana kysyttiin myös siitä, mitä yritysten tulisi ajatella mahdollisista tietoturvauhista liittyen julkipilveen ja suuriin toimittajiin ylipäättään.

Pääsynhallinta nähtiin haastateltavien kesken kriittisenä osana modernia raportointiprosessia. Yhteinen tietovarasto ei saa vaarantaa tietoturvaa. Ratkaisuna tähän tuli monien palveluntarjoajien suosima rooliperusteinen pääsynhallinnan malli, jossa käyttäjät jaetaan rooleihin ja näiden roolien pääsynhallinnan oikeuksia tarkastellaan.

Tämän lisäksi yhtenä tärkeänä aspektina nähdään se, että kaikki tieto ei ole suoraan saatavilla, vaan jollain tapaa maskeerattu rivi- tai sarakekohtaisesti.

“No se se ehkä on aika asia, joka riippuu aika paljon siitä teknologiasta mikä käytetään. Mutta siis ainahan siinä tulee ottaa huomioon niinku pääsynhallinnan asiat. Kyllähän niinku aika paljon nykypäivänä mennään semmoiseen niinku rooliin pohjautuvaan ja sitten niinku usein on jonkinlaista niinku tämmöistä row level ja column level security ja mitä ehkä tarvitsee siihen niinku sen lisäksi sitten.”(H1)

Tiedon suojaaminen nähdään tärkeäksi osaksi raportointiprosessia. Raporteilla saattaa olla yrityksen toiminnan kannalta kriittistä tietoa, joka ei saa päästä väärin käsiin. Yritys voi myös pitää hallussaan henkilötietoja, kuten ip-osotteita tai henkilöturvatuksia. Tämänkaltainen data on luonnollisesti tärkeää yrityksen toiminnan kannalta, mutta välttämättä jokaisen henkilön ei tarvitse nähdä kokonaisuudessaan näitä kaikkia tietoja jokaisesta transaktiosta.

“Joo eli eli tota noin niin tietysti että se tieto pitäisi suojata koko koko niinku tavallaan sen sen matkan aikana. Eli eli tota noin niin niinku sieltä lähtee sitä integroinnin aikana ja ja tota noin niin jos se on jossain väli-storagessa ja ja sitten sitten kun se päätyy sinne loppupisteeseen niin ja ja tota noin niin ja sitten kun siellä on on se raportti tai joku muu.”(H2)

Teknisen toteutuksen lisäksi haastattelija pyrki fokuoitumaan siihen, minkälaisia asioita yritysten tulisi ottaa huomioon ylipäätään valitessaan palveluntarjoajaa. Haasteita tässä on useita. Ensitöikseen yrityksen tulisi pohtia omia kyvykkyyksiään ja vaatimuksia järjestelmälle. Halutaanko enemmän SaaS (Software as A service) vaiko PaaS (Platform as a Service)? Yhtenä suurempana käytännön erona on se, että valmiiksi rakennettu palvelu on helpompi ottaa käyttöön, mutta usein tarjoaa vähemmän räätälöityjä mahdollisuuksia yritykselle. Sen sijaan alustat tarjoavat enemmän mahdollisuuksia, mutta vaativat yritykseltä itseltään usein enemmän, kuten teknisesti taitavia työntekijöitä. Tutkija pyrki kysymään haastateltavia näkemyksiä ja mielipiteitä näihin eroihin ja siihen, miten ne vaikuttavat yrityksen raportointiprosessiin.

“Niissä on paljonkin eroja (Saas vs PaaS), mutta tota ehkä kun puhutaan noista niinku ehkä kun. No jos otetaan niinku PaaS ratkaisusta niin se usein tarkoittaa sitä, että itse tarvitsee konfiguroida ihan niitä resursseja mitä luodaan ja voi konfiguroida verkkoja ja

hallinnoida että miten asiat on speksattu ynnä muuta ja sitten taas jos mennään enemmän tällöinen niinku power bi tyypisiin ratkaisuihin jotka on selkeästi saa sitten saa ratkaisuja niin sitten sulla on niinku tenantti-tason asetuksia, että ne on enemmän semmoisia asetuksia kuin että sä voisit niinku hyvin täsmällisesti konfiguroida sitä ympäristöä en nyt tiedä, että osaanko sitä paremmin selittää, mutta siinä on toisaalta sitten niinku vähemmän speksatattavaa vähemmän hallittavaa, mutta se tarkoittaa myös vähemmän kontrollia, että siinä pitää sitten tehdä se niinku arvio siitä, että kuinka luotettu se alusta on, luotetaanko heidän tietoturva speksien ynnä muihin, että se on niinku enemmän sitten tällöinen.”(H1)

Asia, jota yritykset pohtivat ylipäätään uusien alustojen käyttöönotossa on riskit, joita ne sisältävät. Nämä riskit ovat osittain yleisiä, kuten tietoturvallisuus, mutta myös alakohtaisia, kuten työkalun tai alustan valmius kyseisen alan tulevaisuuteen. Tietoturvaspekti tulee kuvioihin siinä kohtaa, kun pohditaan mitä dataa ylipäätään voidaan viedä pilveen. Suurin osa moderneista analytiikka-alustoista on kuitenkin pilvipohjaisia ja vaikka voidaan määrittää datan lokaatio suurimmassa osassa palveluissa nähdään tämä silti riskinä.

“No kyllä toi ihan ihan perus kysymys on pilven kanssa on monissa organisaatioissa se että mitä dataa saa laittaa pilveen ja mitä ei kun. Se edelleen nähdään niin semmoisena isona uhkana. Niinku monet lähtökohtaisesti näkee sen niin, että pilveen datan laittaminen on merkittävä riski siihen entiseen verrattuna, eli että ne on jossakin omalla excelin kulmalla”.(H3)

Lainsäädäntö on yksi aspekti siinä, mitä dataa pilveen voidaan laittaa. Isoimmilla palveluntarjoajilla käyttäjä voi itse päättää hyvinkin tarkasti missä data säilötään. Missä maassa, missä konesalissa, millä koneella ovat kaikki asioita, joita yritys voi halutessaan päättää.

Haastatteluiden aikana kävi kuitenkin hyvin ilmi se, että pilvipalveluita ei nähdä tietoturvana, vaan enemmänkin kyse on siitä, että käyttäjät ovat itse myös vastuussa tietoturvasta, eivätkä modernit palvelut välttämättä takaa automaattisesta tietoturvaa.

“Se on yks ja sitten voi olla joissain maissa lainsäädäntökin ihan estää sitä. Vaikka se ei ihan välttämättä ole niinku järjellä perusteltavissa, että miksi. Mut näin voi olla. Mutta Suomessa nyt ei ole onneksi nyt ihan niin silleen ongelma se, mutta niinku yritykset itse

estää itseään tekemästä niitä monesti mun käsityksen mukaan. Nyt on tämmöinen peruskysymys varmaan jonkun aikakauden varmaan vielä sen parissa ollaan, mutta niinku he veikkaisin että joskus 5 tai 10 vuoden kuluttua ihan hatusta heitettyinä, niin ehkä se alkaa olla sitten ohi se keskustelu, että voiko pilveen laittaa dataa vai ei, että ehkä ne alkaa siihen mennessä olla sitten ne vakiintunut. Käytännöt ja riskit ja hyödyt niin hyvin kaikkien tiedossa, että ehkä siitä ei sitten tarvitse enää niin keskustella. Mutta tässä vaiheessa se on varsinkin varsin yleinen keskustelunaihe. ”(H3)

Ylipäättään riski ison palveluntarjoajan, kuten Microsoftin kanssa toimiminen nähtiin vähäiseksi etenkin tietoturvan kannalta. Yritysten täytyy pohtia sitä, millä tavoin ne käsittelevät dataansa ja miten sitä säilövät. Spekulaatiivinen pohdinta siitä, että voiko tämänkaltaiseen isoon korporaatioon luottaa on haastateltavien mielestä erikoista. Usein firmat saattavat jo ennestään käyttää pääsynhallintaan esimerkiksi Microsoftin palveluita ja työasemiensa käyttöjärjestelminä Windowsia. Pohdinta siitä, että uskaltaako palveluntarjoajaan tässä kohtaa luottamaan ei ole tehokkainta mahdollista resurssienkäyttöä.

”Mä en kyllä nyt oikein osaa osaa niinku sanoa että millaisia riskejä liittyy microsoftin käyttöön. Esimerkiksi että se nyt on sellainen, että jos ei jos ei luota microsoftiin, niin kyllä elämä on sitten aika vaikeata nykypäivän.”(H3)

8.6 Liiketoimintakäyttäjät

Lähtökohta liiketoimintakäyttäjille on yleensä se, että substanssiosaaminen on enemmän liiketoiminnassa, kuin teknisissä taidoissa. Tämänkaltaisessa tilanteessa myös analytiikan on mahdollistettava tehokas ja järkevä ratkaisu raportointiprosessille.

Liiketoimintakäyttäjät haluavat helpon pääsyn dataan ja luotettavuutta sille, että data on oikealaista. Yleensä dataa halutaan jollain tapaa yhdistellä ja visualisoida raporteille, joten myös kyvykkyudet tämänkaltaiselle toiminnalle ovat tärkeitä. On myös hyväksyttävä, että liiketoimintakäyttäjät eivät ole homogeeninen ryhmä. Kyse on pitkälti yksittäisen käyttäjän motivaatiosta opetella uutta järjestelmää ja halukkuutta tutkia dataa.

”Joo no liiketoiminta käyttäjiä tosiaan on hirveän eri tasoisia että on niitä jotka ei halua koskea mihinkään dataan pitkällä tikullakaan. Ne ehkä käyttää jotain raporttia, koska ne saa sitten hyödyllistä tietoa, mutta ne ei halua koskea mihinkään. Sitten on niitä joita

kiinnostaa itsekin sorkkia sitä dataa ja löytyy vähän osaamistakin siihen jotain niinku yksinkertaista kysely kautta ohjelmointiosaamista niin.” (H3)

Helppokäyttöisyys on yksi aspekti, kun halutaan rakentaa toimiva analytiikka-alusta, joka palvelee raportoinnin tarpeita. BI-tuotteet ja prosessit, joissa liiketoimintaihmiset voivat itse ad-hoc -tyylillä tutkia dataa ja rakentaa omia mallejaan ovat asioita, joista aiemmin voitiin vain haaveilla. Siksi on äärimmäisen tärkeää, että liiketoimintahenkilöiden kyvykkyydet ja halut tutkia dataa ovat tiedossa. Mikäli palvelua halutaan hyödynnettävän, on sen opettelu oltava ainakin perusteisiin melko helppoa. Mikäli palvelu koetaan liian vaikeaksi ei sitä yksinkertaisesti käytetä, vaikka se olisi hyödyllinen. Selainpohjaisuus ja mahdollisimman vähäinen avun tarve käyttöönottovaiheessa ovat avainasemassa siinä, että uusia työkaluja todellisuudessa implementoidaan käyttöön.

“Just mainitsin sen, että ei tarvitse asentaa mitään koneelle vaan selaimen kautta voi käyttää näitä pilvi-työkaluja niin kyllähän se on helppo saada. Sitten tällöinen käyttäjä mukaan kun postataan vaan linkki sillä teamsissa että mene tuonne ja alla tekemään kyselyä ja katso mitä löydät sen sijaan että sanoo että käypä tuolta jostakin pyytämässä joku asennus sun koneellasi niin sitten ne siinä vaiheessa on jo että OK en jaksa ei kiinnosta.” (H3)

Liiketoiminnan osallistaminen raportointiprosessiin enenevissä määrin on mahdollisuus, mutta myös haaste. Keskimäärin liiketoimintakäyttäjät eivät ole teknisesti niin taitavia, joten on huolehdittava rajojen asettamisesta. On tärkeää, että raportteja voi luoda itse ja dataa voi tutkia, mutta esimerkiksi tietojen poistaminen tai pysyvän vahingon aiheuttaminen on rajattava pois käyttäjiltä. Yksi tapa toteuttaa tämä on luoda liiketoimintakäyttäjille valmiita datasettejä ja rajata pääsy vain niihin. Tällöin liiketoiminta ei näe raakadataa, jossa on mahdollisesti turhia sarakkeita, tietoja mihin heillä ei ole pääsyä tai muita vastaavia ominaisuuksia.

“Niin käytännössä se on niin kun tulee semmoiset asiat kuin pääsyn hallinta, että käyttäjillä on pääsy niihin patoihin mitä he tarvitsee eikä sitten sillain että pääset johonkin muuhun dataan sitten niinku asiat että sitten jos on niitä NS niinku valmiita datasettejä joita saa hyödyntää niin ei sitten ole semmoista että joku voi vahingossa poistaa ne vaan että niinku tällöisiä hyvin niinku tietyllä tavalla yksinkertaisia asioita, mutta minkä asioiden kanssa pitää olla sitten niinku tarkempuna” (H1)

Kyseessä on siis tasapaino sen välillä, että rajoitetaan riittävästi, jotta turvallisuus ja hyvät käytänteet toteutuvat, mutta ei liikaa, jolloin innovointi ja koko alustan perusidea ad-hoc-raportoinnista vesittyy. On tärkeää mahdollistaa ja motivoida siihen, että raportointiprosessi on kokonaisuus, johon kuuluu useampi työntekijä, eikä pelkkä IT-osasto. Viestintä, raportointiprosessi ja analytiikka on oltava sillä mallilla, että niiden käyttäjillä on selkeä ohjeistus ja varmuus siitä mitä he voivat tehdä ja mitä heiltä odotetaan.

“Niin tota sitten niin kun asioiden mahdollistaminen, mutta kuitenkin sillä, että kenenkään ei tarvitse pelätä, että hitsi jos mä teen täällä jonkun virheen tai silleen, niin se on mun mielestä asia joka voi helpottaa sitä niin kun. Business- tasonkin käyttäjien toimintaa ja toki sitten jos on niinku vähän että niinku jos pystyy löytää sen että mitä työkaluja he haluaa käyttää.” (H1)

Liiketoimintakäyttäjien taso on myös haastatteluista saatavan datan perusteella vaihteleva, mutta myös todella teknisesti taitavia käyttäjiä on löytynyt. Yhtenä syynä tähän on nähty se, että modernit työkalut ovat huomattavasti aikaisempia helpompia. Usein myös taulokkolaskentaohjelmat, kuten Excel ovat näyttäneet ihmisille, että datan parissa työskentely ei välttämättä ole niin vaikeaa ja pelottavaa, kuin voisi kuvitella.

“Excel on opettanut ihmiset tavallaan semmoinen niinku käsittelee sitä dataa, että sitten se ei niinku enää ole niin pelottavaa, koska kaikki liiketoiminta ihmiset käyttää exceliä työkseen niin se tavallaan sitten sen datan näkeminen jossain power hidasta tai muussa niin ei ole enää niin iso hyppäys sitten sitten kuitenkaan.” (H4)

Aiemmin myös työkalut on nähty hyvin yksiulotteiseksi siten, että niiden käyttäminen vaatii paljon osaamista, tai sitten ne ovat melko turhia. Modernit ratkaisut raportoinnissa, kuten PowerBI ovat ratkaisuja, joita voi käyttää hyvinkin vähäisellä kokemuksella ja ymmärtää sen toimintaa. Työkalu kuitenkin tarjoaa syvyyttä ja mahdollisuuksia myös vaativammallekin tehokäyttäjälle.

“No varmaan joidenkin ihmisten on niinku syytä opetella jotain tällöisiä niinku raportointijärjestelmiä käyttämistä muutakin kun tavallaan sen lopullisen raportin käyttäjän että kyllä se mitä mulla on ollut asiakkaita niin kyllä siellä on niinku ne on ihan tavallisia liiketoiminta ihmisiä jotka on ottanut niinku power bi haltuun tai jotain

tämmöistä, että me ei tarvitse siihen omaa työhön sitä mihin paljon excelin jää sitten pieneksi. Että kun dataa on tarpeeksi, niin excelillä ei tee enää mitään.”(H4)

Haastatteluista käy myös ilmi, että analytiikka-alustan suunnittelussa on tärkeä huomioda se, että minkälaisia käyttäjiä organisaatiolla on? Onko organisaatiolla tehokäyttäjiä, kuten datatieteilijöitä tai data-analyyytiikoita, vai ovatko käyttäjät puhtaasti liiketoimintahenkilöitä. Tämänkaltainen jaottelu vaikuttaa monesta syystä. Tehokäyttäjät, kuten datatieteilijät vaativat yleensä raskaampia pääsyvaatimuksia ja esimerkiksi enemmän laskentatehoa. Käytännössä näillä käyttäjillä rajauksia vähennetään ja oikeuksia lisätään. Riskinä on se, että mikäli nämä käyttäjät eivät osaa käyttää alustaa, niin pahimmassa tapauksessa voidaan tehdä isoakin vahinkoa. Onkin äärimmäisen tärkeää roolittaa käyttäjät oikein ja selvittää yleinen osaamistaso kultakin ryhmältä.

“Tota noin niin tietysti sitten se organisaation tyyppi ja ne käyttäjän roolit eli eli tota noin minkälaisia käyttäjiä siellä on että onko se käyttö sen tyyppistä että niille tulee valmis raportti ilmaantuu johonkin tiettyyn uudelleen tai uudesta saa sen tietyn raportin aina aikavälein tai lähetetään vaikka sähköpostilla tai jotain muuta. Vai onko siellä tota noin niin niinku paljon semmoisia niinku tehokäyttäjiä. Jotka joiden pitää päästä niinku möyhiämaan pöyhimään sitä dataa. Onko siellä niinku oma oma oman omia käyttäjiä jotka on ihan niinku data analytikoita? Tarvitaanko niille jotain jotain niinku tiettyjä tietyn tietyssä muodossa olevaa dataa erikseen ja ja tota noin niin jos jos käyttäjäorganisaatio on sellainen että siellä siellä tehdään paljon niin kun itse itse sitä raportointia ja kehitystä.”(H2)

Tutkija pyrki jatkokysymyksenä kartoittamaan, että mitä asioita käytännössä pitää ottaa huomioon, kun eritellään liiketoimintakäyttäjät ja ammattimaiset raportinkehittäjät. Yksi selkeimmistä näkemyksistä tuli siinä, että on luotava ympäristö, jossa oppiminen on mahdollista ja toteutuksen käyttäminen on selkeää. Tämä tarkoittaa sitä, että raportointiprosessi ja sen osat ovat selkeästi dokumentoitu. Tähän sisältyviä asioita ovat esimerkiksi analytiikka-alustan käyttöohjeet. Ilman selkeää näkemystä ja ohjeita saattaa raportointiprosessin kehittäminen jäädä täysin puolitiehen.

Ohjeistuksen laatiminen ei kuitenkaan ole helppoa sillä se vaatii yhteistyötä tiimin jäsenten kesken ja myös ymmärrystä siitä, minkälaisia asioita todellisuudessa halutaan ja tarvitaan raportille. Samanaikaisesti tulisi kannustaa innovointiin ja luoda ohjeistusta kokonaisuuksien hallintaan ja käyttöön.

“Pitää niinku monet asiat suunnitella aika paljon eri tavoilla kuin se, että kaikki kehittäminen tapahtuisi niinku ammattimaisten kehittäjien toimesta. Eli eli tota noin niin ne ainakin kannattaa kyllä huomioida siinä kun suunnittelee sitä ratkaisua ja niitä välineitä ja ja varsinkin sitten sitä niinku sitä dokumentaatiota mitä se miten sitä tietoa jaetaan siitä ympäristöstä. Jos siellä on paljon semmoista itsepalvelukäyttäjätyyppistä käyttäjää niin niin se vaatii sitten sitä dokumentaatiota ja ja se, että se on oikeat työkalut käytössä ja sitten sitten se vaatii sitä koulutusta että että ne pääsee niinku alkuun ja ja tota noin niin homma ei ei lopahda sitten siihen että ei ne ei ne pääse etenemään asiassa”H2)

9 Johtopäätökset

Tutkimuksen tavoitteena oli löytää vastaukset seuraaviin kysymyksiin:

- Mitä modernit pilvipalvelut ja data-alustat mahdollistavat johdon raportoinnille?
- Mitä pitää ottaa huomioon, kun näitä alustoja suunnitellaan liiketoimintakäyttäjille?
- Millainen prosessi raportoinnissa tulisi olla näillä alustoilla?

Näitä kysymyksiä tarkasteltiin tutkimuksen alussa läpikäymällä teoriaa. Teorian aikana käytiin läpi raportoinnin nykytilaa, menneisyyttä ja raportointiprosessia tiedon keräämisestä aina sen hyödyntämiseen. Teorian lisäksi tutkimuksessa haastateltiin puolistrukturoidusti alan ammattilaisia ja pyrittiin saamaan vastauksia tutkimuskysymyksiin.

Tutkimuksen aikana käytiin läpi sitä, mitä nämä alustat mahdollistavat johdon raportoinnille. Tuloksissa korostui selkeästi se, että tiedon halutaan nykypäivänä olevan entistä reaaliaikaisempaa ja datan tuoreus selkeästi korostuu. Näin ollen raportointi muuttuu kohti ennustavampaa otetta, kuin historian raportointia. Myös uusia raportointitarpeita on ilmennyt, mutta valtavaa raporttitulvaa ei ole välttämättä tullut.

Haastateltavilla oli yhtenevä näkemys siitä, että yritykset Suomessa ja jossain määrin vielä globaalistikin eivät hyödynnä big dataa juurikaan. Datamäärät ovat kasvaneet esimerkiksi sensorien ja IoT-laitteiden myötä, mutta hyödynnettävä datamäärä on pysynyt vielä melko maltillisena. Huomionarvoista on myös se, että isoja datamääriä ei nähdä juurikaan haasteena, sillä ainakin tutkielman tekohetkellä tallennustilaa ja laskentatehoa on tietyissä määrin vielä käytännössä rajattomasti saatavilla suurimmalle osalle yrityksen tarpeista. Haasteena nähdään enemmänkin datan monimuotoisuus, eikä niinkään määrä.

Yritykset voisivat siis hyödyntää dataa vielä enenevässä määrin, mutta selkeää syytä sille, miksi tähän ei olla vielä lähdetty täysin mukaan kaikkien puolesta ei löydy. Yksi selitys, joka kävi haastatteluiden kautta ilmi, on selkeästi se, että yrityksillä ei välttämättä ole itsellään riittävää tietotaitoa, jotta ymmärrettäisiin minkälaisia mahdollisuuksia nämä alustat voisivat tuoda esimerkiksi reaaliaikaisen datan muodossa. Toinen selkeä aspekti

modernien data-alustojen hyödystä on se, että nämä alustat ovat yleensä jonkinlaisessa julkipilvessä, jolloin kustannusten ja toimintojen skaalaaminen onnistuu. Tämä parantaa yrityksen kustanustehokkuutta ja vähentää riskejä.

Vaikka yritykset eivät vielä hyödynnä täysissä määrin modernien analytiikka-alustojen etuja, kuten koneoppimista tai tekoälyä, nämä alustat tarjoavat kuitenkin hyvän lähtökohdan tälle. Yritys voi esimerkiksi aloittaa pienestä alustasta ja maksaa siitä hyvin vähän kuukaudessa, mutta halutessaan kasvattaa infraansa käytännössä rajattomasti ja tarpeen mukaisesti. Tämänkaltaiset alustat voivat toimia myös porttina yrityksen ja sen johdon muodostumiseksi enemmän datavetoiseksi. Datavetoisuutta kuitenkin haittaa se, että vaikka yrityksillä olisi modernit ratkaisut käytössä, niin datan validointiin ja sen oikeellisuuden tarkastamiseen käytetään yhä liian vähän resursseja. Onkin siis selkeää, että liiketoiminta on osallistettava entistä ahkerammin raportointiin, mikäli datasta halutaan kuranttia ja raporteista, sekä analyyseistä mieluisia.

Tutkimuksen aikana käsiteltiin sitä, minkälaisia asioita pitää ottaa huomioon, kun data-alustoja ja raportointiprosesseja suunnitellaan liiketoimintakäyttäjille. Tämä lähtee siitä oletuksesta, että aiemmin raportoinnin tyyli on ollut sellainen, jossa on erikseen raportin rakentaja ja tilaaja tai käyttäjä. Modernit raportointiprosessit, data mesh ja ylipäätään datan tuotteistaminen on luonut tavan, jossa liiketoimintakäyttäjät voivat itse luoda haluamiaan ad-hoc raportteja tai analyyseja hyvin helposti. Tämä ei suoraan poista tarvetta ammattimaisilta raportinkehittäjiltä, vaan pikemminkin rikastuttaa yrityksen raportointia.

Haastateltavilta tuli kattavaa näkemystä siitä, minkälaisia asioita tämänkaltaisessa prosessissa tulisi ottaa huomioon. Yksi tämänkaltaisen aspekti oli rooliperusteinen käyttäjänhallinnan implementointi. Yrityksen siirtyessä enemmän tilanteeseen, jossa kaikki tieto löytyy yhdestä paikasta, on tärkeää pohtia, että kenellä on oikeudet mihinkin dataan. Tiedon anonymisointi ja käyttäjienhallinta roolien, sekä ryhmien avulla nousi erityisen tärkeäksi osuudeksi tätä.

Liiketoimintakäyttäjistä puhuttaessa kävi myös selkeästi ilmi, että käytettävien työkalujen tulee olla helppoja käyttää. Helppokäyttöisyys ulottui paitsi käyttöliittymästä ja ominaisuuksista myös käyttöönottoon. Useat modernit ratkaisut ovat selainpohjaisia, jolloin omalle työasemalle ei välttämättä tarvitse asentaa mitään, vaan yksinkertainen linkki verkkosivulle tuo pääsyn dataan.

Liiketoimintakäyttäjien taidot ovat myös kokeneet selkeän kehityskaaren. Yhtenä syynä tähän nähtiin se, että taulukkolaskentatyökalut ovat olleet monille tuttuja jo vuosikymmeniä, joten hyppäys BI-työkaluihin, kuten PowerBI:hin ei tunnu enää niin pahalta. Työkalujen kehitys on mahdollistanut myös sen, että niiden opettelu ei enää vaadi tutkintoa ja vuosia kokemusta, vaan yksinkertaiset ja päivittävät työt on mahdollista hallita melko vähäisellä teknisellä osaamisella.

Tutkimuksen tarkoituksena oli pohtia myös, että mitä ylipäätään on hyvä raportointi. Monesti prosessia tarkastellaan lähinnä liiketoiminnan näkökulmasta ja teknisempi puoli jää pimentoon. Dynaaminen ja alati globalisoituva yrity maailma kuitenkin vaatii entistä enemmän, sekä teknistä osaamista, että liiketoimintanäkemyksiä, mikäli halutaan toimia datavetoisesti.

Haastateltavilta kyseltiin, minkälaiseksi he kokevat hyvän raportointiprosessin ja minkälaisia asioita siinä tulisi ottaa huomioon. Lähtökohta raportoinnille ja koko prosessille tulisi olla liiketoiminnan tarpeiden kartoitus. Tähän prosessiin sisältyy myös se, että liiketoiminta ja tekninen toteutus keskustelevat keskenään siitä, minkälaisia asioita voidaan raportoida ja mitä ylipäätään on mielekästä raportoida. Visualisointi ja tietojen yhdistely tulee kartoittaa mahdollisimman tarkasti ja selkeästi.

Tarpeiden kartoitus ei välttämättä ole helppoa, joten raportointiprosessin on oltava joustava paitsi prosesseiltaan, mutta myös tekniseltä toteutukseltaan. Prosessit täytyy rakentaa kestäviksi ja sellaisiksi, että ne eivät hajoa, mikäli muutostarpeita, tai uusia tarpeita kehkeytyy. Datan tuotteistaminen ja erilaiset tiimit datan ympärillä ovat tulleet jäädäkseen. Yritysten on pyrittävä itse luomaan kulttuuri, joka kannustaa hyvään raportointiin ja myös innovointiin. Datan parissa työskentely ei pitäisi olla pelkästään kovan teknisen osaamisen työntekijöillä, vaan myös liiketoiminnalla. Hyvä raportointi ei ole pelkästään sitä, että yritykset raportoivat pakolliset taloudelliset ja lainmukaiset raportit, vaan sitä, että yritys pyrkii aktiivisesti luomaan arvoa datasta ja valjastamaan sen voiman käyttöönsä.

Tutkielman rajoitteet liittyivät pitkälti haastateltavien määrään, sekä tutkimuksen substanssiin, jossa pyrittiin pitämään teoria ja empiria mahdollisimman paljon liiketoimintakeskeisenä. Raja on kuitenkin häilyvä, sillä kuten tutkimuksesta käy ilmi, myös liiketoiminnan näkökulma on muuttunut entistä teknisemmäksi.

Haastateltavilla on keskenään kattava kokemus, mutta näkemykset ovat subjektiivisia näkemyksiä, jotka ovat henkilö- ja aikasidonnaisia. Subjektiivinen näkemys muodostuu omien kokemusten, näkemysten, mutta myös työnkuvan kautta. Haastateltavista osa oli selkeästi enemmän keskittyneitä tekniseen toteutukseen ja osa enemmän liiketoimintakeskeisiä. Tutkimus näin ollen sai hyvää poikkileikkausta koko raportointiprosessista, mutta saattaa myös vaikuttaa haastateltavien tietotaitoon eri kysymysten osa-alueilta.

Tutkimuksena tekohehkellä eletään murrosvaihetta siinä, että osa yrityksistä on pitkällä datapolullaan ja osa vasta alussa. Toimiala elää ja muuttuu nopeasti, joten uusia toimintamalleja ja työkaluja tulee nopeaan tahtiin. Usein nämä uudet toimintamallit ja työkalut testataan myös heti käytännössä eli pitkän aikavälin dataa uusista metodeista ei välttämättä ole saatavilla.

Yksi selkeä jatkotutkimusmahdollisuus olisi tarkastella case-tyyppisesti yritystä tai yrityksiä, jotka aloittavat raportointiprosessien luomisen. Case:n avulla voitaisiin syväluotaavasti tarkastella vain yhtä tai muutamaa yritystä. Näin saatava tieto olisi heikommin yleistettävissä, mutta se toisi valideettiä nykyiselle tutkimukselle, tai poikisi uusia jatkotutkimusmahdollisuuksia.

Lähteet

- .Kim, Rosemary et al. “Formal Analysis of Segregation of Duties (SoD) in Accounting: A Computational Approach.” *Abacus (Sydney)* 56.2 (202): 165–212. Web.
 <<https://aws.amazon.com/sports/nhl-opportunity-analysis/>>
 <<https://blogs.microsoft.com/on-the-issues/2023/12/19/trust-privacy-bing-copilot-responsible-ai/>>
 <<https://chartio.com/learn/business-intelligence/a-step-by-step-guide-to-bi-reporting/>>
 <<https://digia.com/en/news/an-ai-developed-by-digia-predicts-ilves-will-come-out-on-top-in-the-finnish-elite-league-ice-hockey-4677183>>
 <<https://learn.microsoft.com/en-us/azure/role-based-access-control/>>
 <<https://towardsdatascience.com/why-data-analytics-is-gaining-hype-in-the-21st-century-b7b1ca289f09> >
 <<https://uutishuone.pwc.fi/vastuullisuuden-raportointi-taloustraportoinnin-tasolle-kuinka-se-tehdaan>>
 <<https://www.dataversity.net/what-is-modern-bi-and-how-is-it-different-from-traditional-bi/> >
 <<https://www.forbes.com/advisor/business/software/ai-in-business/>>
 <<https://www.forbes.com/sites/forbesbusinesscouncil/2023/03/07/financial-reports-vs-management-reports-what-do-they-mean-for-your-service-business/>>
 <<https://www.forbes.com/sites/forbestechcouncil/2021/10/14/flying-blind-how-bad-data-undermines-business/?sh=2efe93df29e8>>, haettu 12.12.2023
- Alexopoulos, C., Zuiderwijk, A., Charapabidis, Y., Loukis, E., Janssen, M. (2014). Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media. In: Janssen, M., Scholl, H.J., Wimmer, M.A., Bannister, F. (eds) *Electronic Government. EGOV 2014. Lecture Notes in Computer Science*, vol 8653. Springer, Berlin, Heidelberg. Web
- Amazon 1.4.2023, NHL Edge IQ, haettu 15.12.2023
- Bandara, Madhushi, Fethi A. Rabhi, and Muneera Bano. “A Knowledge-Driven Approach for Designing Data Analytics Platforms.” *Requirements engineering* 28.2 (2023): 195–212. Web.

- Bendre, Mangesh et al. "DATASPREAD: Unifying Databases and Spreadsheets." Proceedings of the VLDB Endowment 8.12 (2015): 2000–2003. Web.
- Berisha, Blend, Endrit Mëziu, and Isak Shabani. "Big Data Analytics in Cloud Computing: An Overview." Journal of cloud computing : advances, systems and applications 11.1 (2022): 24–24. Web.
- Chartio, A Step-by-Step Guide to BI Reporting, haettu 1.2.2024
- Chen, Chun-houh., Wolfgang. Härdle, and Antony. Unwin. Handbook of Data Visualization. Berlin ; Springer, 2007. Web.
- Dataversity 24.1.2020, What is Modern BI and How is it Different From Traditional BI? , haettu 4.1.2024
- Digia 9.11.2023 An AI developed by Digia predicts: Ilves will come out on top in the Finnish Elite League ice hockey championship, haettu 4.1.2024
- Dong L, Li G, Wang D, et al. Granular data modeling and analysis based on optimized subsets of data. Measurement and Control. 2023;56(1-2):34-48. Web.
- Flesca, Sergio et al. "From Star Schemas to Big Data: 20+ Years of Data Warehouse Research." A Comprehensive Guide Through the Italian Database Research. Vol. 31. Switzerland: Springer International Publishing AG, 2018. Web.
- Forbes 14.10.2021 Flying Blind: How Bad Data Undermines Business
- Forbes 24.4.2023, How Businesses Are Using Artificial Intelligence In 2024, haettu 15.1.2024
- Forbes 7.3.2023, Financial Reports Vs. Management Reports: What Do They Mean For Your Service Business? haettu 6.1.2024,
- Fugini, Mariagrazia, Jacopo Finocchi, and Paolo Locatelli. "A Big Data Analytics Architecture for Smart Cities and Smart Companies." Big data research 24 (2021): 100192-. Web
- Hasanah, F N, F Renaldi, and F R Umbara. "Implementing Cloud Computing Technology on Restaurant's Expenses Monitoring System." IOP conference series. Materials Science and Engineering 1115.1 (2021): 12039-. Web.
- <https://www.montecarlodata.com/blog-what-is-a-data-mesh-and-how-not-to-mesh-it-up/>
- Inês Araújo Machado, Carlos Costa, Maribel Yasmina Santos. Data Mesh: Concepts and Principles of Paradigm Shift in Data Architectures. CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANAgement / HCist - International

- Conference on Health and Social Care Information Systems and Technologies. 2021
- Inmon, W. H., Dan Linstedt, and Mark Rogers. *Data Architecture : A Primer for the Data Scientist : Big Data, Data Warehouse and Data Vault*. Ed. Steven Elliot. 1st edition. Amsterdam, Netherlands: Morgan Kaufmann, 2015. Print.
- Karagiannis, Anastasios, Panos Vassiliadis, and Alkis Simitsis. "Scheduling Strategies for Efficient ETL Execution." *Information systems (Oxford)* 38.6 (2013): 927–945. Web.
- Khine, Pwint & Wang, Zhao. (2018). *Data lake: a new ideology in big data era*. ITM Web of Conferences. 17. Web
- Kimball, Ralph., and Margy. Ross. *The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling*. Third edition. Indianapolis, Ind: Wiley, 2013. Print
- Lubwama, Roni. *The Inside Track to Excelling As a Business Analyst: Soft Skills That Can Accelerate Your Career*. 1st ed. Berkeley, CA: Apress L. P, 2019. Web.
- Manning, Anna. *Databases for Small Business: Essentials of Database Management, Data Analysis, and Staff Training for Entrepreneurs and Professionals*. Berkeley, CA: Apress L. P, 2015. Web.
- Maresova, Petra, Vladimir Sobeslav, and Ondrej Krejcar. "Cost-Benefit Analysis - Evaluation Model of Cloud Computing Deployment for Use in Companies." *Applied economics* 49.6 (2017): 521–533. Web.
- Marilena Mirouniuc, Mihai Carp, Ionela-Corina-Chersan. "The Relevance of Financial Reporting on the Performance of Quoted Romanian Companies in the Context of Adopting the IFRS", *Procedia Economics and Finance*. (2015)
- Maximini, Dominik. *The Scrum Culture: Introducing Agile Methods in Organizations*. 2015. Cham: Springer International Publishing AG, 2015. Web.
- McAfee, Andrew, and Erik Brynjolfsson. *Big Data: The Management Revolution*. Vol. 90. United States: N.p., 2012. Print.
- Microsoft 19.12.2023 *Enhancing trust and protecting privacy in the AI era*, haettu 12.1.2024
- Microsoft 21.8.2022 *What is Azure role-based access control*, haettu 4.1.2024
- MIT Sloan Management Review 7.12.2023 *Leading With Decision-Driven Data Analytics*, haettu 10.1.2024
- Monetcarlodata 31.7.2020 *What is a Data Mesh — and How Not to Mesh it Up*, haettu 5.2.2024

Pałys, Marcin., Andrzej. Pałys. Benefits and Challenges of Self-Service Business

Intelligence Implementation. *Procedia Computer Science* 225 (2023): 795–803.

Web.

Praveen, Shagufta & Chandra, Umesh. (2020). Influence of Structured, Semi- Structured, Unstructured data on various data models. *International Journal of Scientific and Engineering Research*. 8. 67-69.

PwC 9.11.2022 Vastuullisuuden raportointi talousraportoinnin tasolle – kuinka se tehdään? , haettu 11.12.2023

Rinu, Gour. Why Data Analytics is Gaining HYPE in the 21st Century, haettu 10.2.2024

Sajida, S., & Ramakrishna, Dr. S. “A Study of Extract-Transform-Load (ETL) Processes.” *International Journal of Engineering Research & Technology (IJERT)*, Special Issue - 2015, NCACI-2015 Conference Proceedings, ISSN: 2278-0181. Web.

Sandholm, Thomas, and Dongman Lee. “Notes on Cloud Computing Principles.”

Journal of cloud computing : advances, systems and applications 3.1 (2014): 1–10. Web.

Liitteet

9.1 Haastattelujen rakenne

- Minkälainen on hyvä raportti?
- Millä tavoin raportointi on muuttunut oman työurasi aikana?
- Minkälainen on hyvä raportointiprosessi?
- Millä tavoin BI ja analytiikka voivat auttaa liiketoimintaa?
- Vaikuttaako raportointityökalujen ja analytiikka-alustojen valinta vaikuttaa siihen, minkälaisia raportteja voidaan tehdä?
- Mitä asioita yritysten tulisi ottaa huomioon luodessaan yhtenäistä tietovarasto ja raportointialustaa liiketoimintakäyttäjille?