



## **TIEDON REPLIKOINTIMENETELMÄN PÄIVITYS YRITYKSESSÄ : CASE TUK- KUKAUPPA**

Lappeenrannan–Lahden teknillinen yliopisto LUT

Tietotekniikan kandidaatintyö

2024

Janni Timoskainen

Tarkastaja: Tutkijatohtori Jiri Musto

## TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Teknis-luonnontieteellinen

Tietotekniikka

Janni Timoskainen

### **Tiedon replikointimenetelmän päivitys yrityksessä : Case tukkukauppa**

Tietotekniikan kandidaatintyö

2024

51 sivua, 6 kuviota, 2 taulukkoa ja 1 liite

Tarkastaja: Tutkijatohtori Jiri Musto

Avainsanat: Replikointi, Tietokanta, Tietovarasto, Tietojenkäsittely

Tässä kandidaatintyössä tutkitaan kohdeyrityksen tietojen replikointimenetelmän merkitystä. Nykyajan yrityksissä pyörii valtavasti ja yhä enemmän tietoa sähköisessä muodossa. Avainasemassa tietojen saatavuuden kannalta on niiden replikointi. Tietojen replikointiprosessissa tietoja kopioidaan alkuperäisestä lähteestä toiseen, jossa niitä pysytään hyödyntämään mm. erilaisiin tiedon analysointitarpeisiin.

Työn tutkimusmenetelmänä toimii laadullinen tapaustutkimus, jonka tarkoituksena on selvittää nykyisen replikointimenetelmän, että uuden merkitystä kohdeyritykselle. Tutkimuksen aihealueena on kohdeyrityksen laajan projektin pienempi osa-alue, replikointimenetelmän päivittäminen. Tutkimusaineisto kerättiin haastattelun avulla, jossa haastateltiin kohdeyrityksen järjestelmäpäällikköä. Lisäksi tutkimuksen aineistona toimii kohdeyrityksen dokumentaatio nykyisestä replikointimenetelmästä.

Haastattelussa saatiin vastauksia niin nykyiseen, kuin uuteen menetelmään liittyviin kysymyksiin. Näiden avulla voitiin analysoida replikointimenetelmän merkitystä ja asemaa kohdeyrityksessä, luoden samalla yleiskuvaa päätöksestä päivittää nykyinen menetelmä. Tutkimus osoitti, että uusi replikointimenetelmä lisää replikoinnin nopeutta ja tehokkuutta, sekä vähentää henkilösidonnoisuuksia. Tutkimuksen tulokset tuovat uusia näkökulmia käytännön muutoksiin koskien replikointimenetelmän päivittämistä ja tarjoavat laajemman soveltamis-mahdollisuuden.

Työn rajaus tietyn organisaation replikointimenetelmiin luo tutkimuksen tuloksista yksityiskohtaisen, mutta rajoitetusti yleistettävän kokonaisuuden. Haastattelun rajaus yhteen henkilöön rajaa tutkimusaineiston näkökulmaa.

## ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Science

Software Engineering

Janni Timoskainen

### **Updating data replication method in a company : Case wholesale trade**

Bachelor's thesis

2024

51 pages, 6 figures, 2 tables and 1 appendix

Examiner: Postdoctoral Researcher Jiri Musto

Keywords: Replication, Database, Datawarehouse, Data management

This bachelor's thesis examines the importance of the target company's data replication method. In today's companies, a huge amount and more and more information circulates in electronic form. The key to data availability is its replication. In the data replication process, data is copied from the original source to another, where it is kept for use, e.g. for various data analysis needs.

The research method is a qualitative case study and the purpose is to find out the significance of the used replication methods. The topic of the work is a smaller part of the target company's extensive project, updating the replication method. The research material was collected through an interview, where the system manager of the target company was interviewed. In addition, the target company's documentation of its current replication method serves as material.

In the interview, answers were received to questions related to both the current and the new replication method. This made it possible to analyze the importance and status of the method used, while creating an overview of the decision to update the current method. The research showed that the new replication method increases the speed and efficiency of replication and reduces personal ties. The results of the research bring new perspectives to practical changes regarding the updating of the replication method and offer a wider possibility of application.

Limiting the work to the replication methods of a specific organization creates a detailed, but difficult to generalize, research results. In addition, limiting the interview to one person limits the perspective of the research material.

## KIITOKSET/ ACKNOWLEDGEMENTS

Haluan kiittää kaikkia teitä, jotka olette tukeneet minua tämän työn kirjoittamisen aikana. Erityiskiitos ohjaajalleni palautteesta ja opastuksesta työn eri vaiheissa, sekä kohdeyritykselle aiheesta ja mahdollisuudesta.

## LYHENNELUETTELO

AIX	Advanced Interactive Executive
DB	Tietokanta (Database)
DBS	Tietokantajärjestelmä (Database system)
DBMS	Tietokannan hallintajärjestelmät (Database Management Systems)
DW	Tietovarasto (Datawarehouse)
ERP	Toiminnanohjausjärjestelmä (Enterprise Resource Planning)
ETL	Extract, Transform, Load
NoSQL	Ei-relaatiotietokanta (Not Only SQL)
MSSQL	Microsoft SQL
OLAP	Online Analytical Processing
SQL	Structured Query Language

## Sisällysluettelo

Tiivistelmä

Abstract

Kiitokset

Lyhenneluettelo

1	Johdanto.....	8
2	Tutkimuksen tausta.....	11
2.1	Tietokannat.....	11
2.1.1	Relaatiomalli.....	13
2.1.2	Tietokannan hallintajärjestelmät.....	15
2.2	Tietovarastot.....	17
2.3	Replikointi.....	19
3	Tutkimusmenetelmä.....	23
3.1	Kohdeyritys.....	23
3.1.1	Projekti.....	23
3.1.2	Nykyinen replikointimenetelmä.....	24
3.2	Aineistonkeruumenetelmä.....	26
4	Tutkimustulokset.....	29
4.1	Nykyinen replikointimenetelmä.....	29
4.2	Uudistamisprosessi.....	32
4.3	Uusi replikointimenetelmä.....	34
5	Keskustelu.....	37
5.1	Tulosten pohdinta ja merkitys.....	37
5.2	Tutkimuksen luotettavuus.....	42
6	Yhteenveto.....	44
	Lähteet.....	45

## Liitteet

Liite 1. Haastattelukysymykset

## Kuvioluettelo

Kuvio 1. Dokumenttitietokannan dokumentin rakenne

Kuvio 2. Relaation rakenne

Kuvio 3. Tietovaraston rakenne

Kuvio 4. Synkroninen replikointimenetelmä

Kuvio 5. Asynkroninen replikointimenetelmä

Kuvio 6. Kohdeyrityksen replikointiprosessi

## Taulukkuuettelo

Taulukko 1: Nykyisen replikointimenetelmän arviointi

Taulukko 2: Uuden replikointimenetelmän arviointi

# 1 Johdanto

Erilaiset yritykset ja heidän liiketoimintansa luo valtavan määrän tietoa, jota nykypäivänä säilytetään enenevissä määrin sähköisessä muodossa (Gupta & Rani 2019, 325-326). Tämä sähköisessä muodossa säilytettävä tieto, toisin sanoen data (Bhatia & Bansal 2015; Kitchin 2014, 1-2) on myös pidettävä tallessa esimerkiksi erilaisissa tietokannoissa. Tietokannoista voidaan kopioida ja siirtää tietoa esimerkiksi tietovarastoihin, ja tätä prosessia kutsutaan replikoinniksi (Özsu & Valduriez, 2020, 247). Yleisesti replikointiprosessissa tietoja kopioidaan ja siirretään alkuperäisestä lähteestä toiseen sijaintiin.

Nasibullin ja Novikovin (2020) mukaan replikointi pyrkii takaamaan seuraavat ominaisuudet:

- Saatavuus: varmistaa tiedon käytettävyyden, vaikka yhteys pääasialliseen lähteeseen katkeaisi.
- Luotettavuus: perustuu samojen tietojen sijaitsemiseen useissa replikoissa, jolloin voidaan luottaa, että tieto on aina jossain tallessa.
- Suorituskyky: viittaa järjestelmän kykyyn suorittaa toimintoja tietyssä ajassa. Replikat mahdollistavat saman tiedon käsittelyn useille eri ihmisille samanaikaisesti, joka tehostaa tiedon käytettävyyttä. Näin ollen suurempaa määrää tietoa voidaan käsitellä samassa ajassa, mikä nostaa suorituskykyä.

Yrityksien tulisi tavoitella sellaista replikointimenetelmää, joka takaisi aiemmin mainitut ominaisuudet, koska näin saavutetaan suurin hyöty käytetystä replikointimenetelmästä (Mazilu, 2010). Tietojen replikointi yhdessä tietovarastoympäristöjen kanssa on usein osana yritysten liiketoimintatiedon hallintajärjestelmää, jonka tarkoituksena on kerätä ja analysoida esimerkiksi yrityksen liiketoiminnan kannalta tärkeitä tietoja (Van Der Lans, 2012). Myös tämän takia on oleellista, että yrityksillä on ajankohtaista tietoa saatavilla esimerkiksi tietojen raportointia varten.

Tämä kandidaatintyö tutkii kohdeyrityksen eli tukkukaupan laajan projektin yhtä osa-aluetta. Mittavaan projektiin sisältyy yrityksen toiminnanohjausjärjestelmän (ERP) tietokannan version päivittäminen, tietojärjestelmän käyttöjärjestelmän vaihtaminen, sekä



tietovarasto ympäristön ja replikoinnin päivittäminen (Parviainen, 2023). Tämän työn tarkoituksena on tarkastella yrityksen replikointimenetelmän päivittämistä, vertailla nykyisen ja päivitetyn menetelmän eroja ja selvittää replikointimenetelmän merkitystä.

Työssä tehdään laadullinen tapaustutkimus, jonka tavoitteena on ymmärtää tutkimuskohdetta (Laine & Bamberg & Jokinen, 2015, 10) eli kohdeyrityksen replikointimenetelmän merkitystä yksityiskohtaisemmin. Tutkimuksessa aineistonkeruumenetelmänä toimii puolistrukturoitu haastattelu, tarkemmin teemahaastattelu, jossa haastatellaan yrityksen järjestelmäpäällikköä. Tämä avaa ainutlaatuisen tilaisuuden syventyä yrityksen nykyisen, että uuden replikointimenetelmän yksityiskohtiin ja arvioida niitä. Haastattelu on yksi laadullisen tutkimuksen käytetyimpiä aineistonkeruumenetelmiä (Hirsjärvi & Hurme, 2022; Hyvärinen & Nikander & Ruusuvoori, 2017), jonka avulla saadaan kerättyä tutkimukselle aineistoa auttaamaan mm. tutkimuskysymyksiin vastaamisessa. Teemahaastattelun luoma aineisto analysoidaan sisällönanalyyssimenetelmän keinoin. Työssä tutustutaan myös aiheeseen liittyvään kirjallisuuteen, sekä kohdeyrityksen dokumentaatioon, jonka avulla pyritään analysoimaan yrityksen käyttämiä menetelmiä. Kohdeyrityksen dokumentaatio nykyisestä replikointimenetelmästä on osa tutkimuksessa käytettyä aineistoa.

Työn tavoitteena on ymmärtää tarkemmin, miten tietojen replikointi toimii, sekä selvittää miten yrityksen nykyinen replikointimenetelmä eroaa uudesta. Tarkoituksena on luoda syvempi käsitys siitä, miksi yritys on valinnut tietynlaisen uuden menetelmän ja mitä etuja tai haasteita se tuo mukanaan nykyiseen menetelmään verrattuna. Työssä vastataan seuraaviin tutkimuskysymyksiin: *Miten kohdeyrityksen replikointi on toteutettu nykytilanteessa? Entä uuden menetelmän myötä? Mitä vaihtoehtoja replikointimenetelmässä on harkittu ja miksi juuri projektin myötä tuleva menetelmä valittiin?*

Kohdeyritys, sekä lukija voi hyödyntää tämän työn ja tutkimuksen tuloksia yleiskuvana kohdeyrityksen replikointiprosessista sekä nykyisen että uuden järjestelmän kannalta. Työn tuloksia voidaan pitää pohjana sille, miksi replikointimenetelmää kannattaa päivittää, sekä mitä hyötyjä, sekä haittoja päivittäminen voi tuoda mukanaan. Tutkimus tarjoaa yleisen ymmärryksen replikoinnista ja auttaa ymmärtämään, miten replikointimenetelmien päivitys voi parantaa tietojärjestelmien toimivuutta, nopeutta ja luotettavuutta.

Tämä kandidaatin työ sisältää kuusi lukua. Työn toisessa luvussa tarkastellaan tutkimuksen taustatietoja eli tutkimuskontekstiin liittyviä käsitteitä, sekä menetelmiä. Kolmannessa

luvussa selvennetään työn tutkimusmenetelmä, perehdytään kohdeyritykseen ja heidän laajaan projektiinsa, sekä tarkemmin kohdeyrityksen nykyiseen replikointiprosessiin. Lisäksi luvussa määritellään työn aineistonkeruumenetelmä ja kuvaillaan tutkimuksen etenemistä. Neljännessä luvussa käsitellään tutkimuksen tuloksia, ja viidennessä luvussa pohditaan tuloksia, sekä tarkastellaan tutkimuksen luotettavuutta. Viimeisessä eli kuudennessa luvussa käsitellään työn yhteenveto.

## 2 Tutkimuksen tausta

Tässä luvussa käydään läpi tutkimuksen tausta. Tutkimuksen ja tämän työn kannalta on hyvinkin olennaista ymmärtää muutamia erilaisia käsitteitä ja asiakokonaisuuksia. Jo siitä lähtien kun tietoa on voitu kerätä, on yritykset luoneet, käyttäneet ja käsitelleet tietoa heidän liiketoimintansa vahvistamiseksi (Beynon-Davies, 2004, 4). Tehokas tiedon hyödyntäminen, kuten liiketoiminnan kehitystä tukeva raportointi ja analysointi, on usein erittäin tärkeää yrityksille (Sivula & Aho & Laukkanen 2023). Liiketoiminnan kattavan tilannekuvan rakentaminen edellyttää tyypillisesti tietojen yhdistämistä eri tietojärjestelmistä. Tämä tietojen yhdistäminen vaatii usein tietovaraston rakentamisen, jossa tietoja käsitellään ja yhdistellään (Sivula et al. 2023).

Miten yritykset voivat säilyttää suuria määriä tietoa, varmistaa tietojen saatavuutta, sekä hyödyntää tietoa analysoinnissa ja näin ollen kehittää toimintaansa? Tällöin ratkaisuksi käy erilaiset tietokannat, sekä jo aiemmin mainitut, suuremmat ja enemmän tietojen analysointiin tarkoitetut tietovarastot (Velicanu & Matei, 2007). Näiden kahden välillä tietoa voidaan kopioida ja siirtää esimerkiksi replikoinnin avulla, joka mahdollistaa tiedon saatavuuden, sekä sen jatkokäsittelyn mm. raportoinnissa (Mazilu, 2010).

Ensimmäisenä tässä luvussa perehdytään aiheeseen liittyvään käsitteistöön, kuten tietokantoihin, niiden eri malleihin erityisesti relaatiomalliin, sekä siihen miten tietokantoja hallinnoidaan. Lisäksi luvussa käydään läpi mitä tietovarastot ovat. Tämän jälkeen luvussa käsitellään mitä replikoinnilla tarkoitetaan.

### 2.1 Tietokannat

Sähköistä tietoa on pidettävä jossain tallessa ja monia tiedon tallennusrakenteita on tarjolla yhä enenevässä määrin (Laine, 2000). Suurten tietomäärien järjestäminen on kehittynyt aina tiedostoista tietokantoihin ja edelleen tietovarastoihin (Velicanu & Matei, 2007). Edellä mainitut ovat suuruus järjestyksessä, tiedostojen ollessa pienin tietorakenne ja tietovarastojen karkeasti suurin. Yksinkertaisuudessaan tietokanta (DB) voidaan määritellä tiettyä käyttötarkoitusta varten tehdyksi kokonaisuudeksi, joka säilyttää kokoelman toisiinsa liittyviä

tietoja (Laine, 2000). Tietokantojen tiedot ovat jollain tavalla tekemisissä todellisen maailman tapahtumien kanssa (Lahtonen, 2002, 2). Esimerkiksi musiikkitietokannassa voi olla tietoa artisteista ja heidän kappaleistaan, jolloin tietokanta sisältää olennaisen linkin todelliseen maailmaan.

Tietokannat voidaan jakaa karkeasti niiden käyttämän mallin mukaan relaatiotietokantoihin, jotka noudattavat relaatiomallia, sekä ei-relaatio- eli Not Only SQL (NoSQL) -tietokantoihin (Han & E & Le & Du, 2011; Kaufmann & Meier, 2023). NoSQL tietokannat eivät noudata relaatiomallia ja ne voidaan jakaa dokumentti-, graafi-, sekä avain-arvotietokantoihin (Kaufmann & Meier, 2023). Suosituimpia NoSQL tietokantoja on mm. MongoDB (2023), Cassandra (2024) ja HBase (2024) (Ploetz & Kandhare & Kadambi & Wu, 2018, 7).

NoSQL tietokantojen rakenne eroaa relaatiotietokantojen rakenteesta mm. sillä, ettei tietoa säilytetä taulukkomaisissa rakenteissa riippuen NoSQL -tietokannan tyypistä. Lisäksi tietokantojen kieli ei ole Structured Query Language -tietokonekieli (SQL) (Kaufmann & Meier, 2023). Tietojen tallennus ilman kiinteää taulukkorakennetta tarjoaa joustavuutta tietojen mallinnukseen. Käsitellään seuraavaksi ytimekkäästi yhtä NoSQL tietokantatyyppiä, dokumenttitietokantaa.

```
{
  "_id": 1,
  "etunimi": "Etunimi1",
  "sukunimi": "Sukunimi1",
  "osoite": {
    "katuosoite": "Esimerkkikatu 123",
    "postinumero": "00000",
    "kaupunki": "Lappeenranta"
  },
  "puhelinnumero": "Puh1"
}
```

#### Kuvio 1. Dokumenttitietokannan dokumentin rakenne

Dokumenttitietokannat koostuvat nimensä mukaan dokumenteista, jotka yhdessä muodostavat kokoelmia (Nayak & Poriya & Poojary, 2013). Dokumenttitietokannoissa tieto tallennetaan mm. Extensible Markup Language (XML) tai JavaScript Object Notation (JSON) -

muodossa (Kaufmann & Meier, 2023). Kuviossa 1 on esimerkki dokumenttitietokannan yhden dokumentin rakenteesta JSON muodossa. Dokumenttitietokannat perustuvat avain-arvotietokantojen tavoin arvojen tallentamiseen avaimen perusteella (Nayak et al. 2013). Esimerkiksi kuviossa 1 ”etunimi” -kenttä toimii avaimena ja sitä vastaava arvo on merkattu kaksoispisteellä erotettuna avaimen yhteyteen. Dokumenttien sisältämät tiedot voivat olla monimutkaisia ja sisältää sisäkkäisiä rakenteita (Nayak et al. 2013), kuten kuviossa 1 ”osoite” -kentän arvo on olio, joka sisältää kolme avain-arvo paria.

NoSQL:n etuihin kuuluu hyvä joustavuus, skaalautuvuus, suorituskyky, sekä helppokäyttöisyys (Han et al. 2011; Nayak et al. 2013). NoSQL:n heikkouksiin lukeutuu johdonmukaisuuden puute ja joidenkin NoSQL tietokantojen puute toimia Atomically, Consistency, Isolation, Durability (ACID) -periaatteiden mukaisesti, joiden tarkoituksena on turvata tietojen eheys erilaisissa tilanteissa (Nayak et al. 2013; Ploetz et al. 2018, 9).

Tietokantojen sisältämiä tietoja voidaan säilyttää halutun ajanjakson ajan ja erilaisia tietokantoja rakennetaan niiden käyttötarkoituksen mukaan (Beynon-Davies, 2004, 6). Yksi yrityksen suosima tietokantatyyppe on operatiivinen tietokanta. Velicanu ja Matei (2007) kuvailevat operatiivisen tietokannan olevan sellainen tietokanta, jossa tietoja päivitetään useasti, eli lisätään, muokataan, sekä poistetaan. Kun tietokannassa päivitetään tietoja, jokaisen päivityksen yhteydessä aiemmat arvot menetetään. Operatiivisten tietokantojen on usein tarkoitus säilyttää ajankohtaista tietoa enemmän, kuin historiallista tietoa.

Triggerit ovat tietokantaproseduureja, jotka aktivoituvat automaattisesti jonkin tapahtuman myötä (Dalton & Whitehead, 2001; Dewson, 2015) esimerkiksi, kun tietoja lisätään, muokataan tai poistetaan tietokannasta. Dewsonin (2015) mukaan triggerit voivat vaikuttaa tietokantojen suorituskykyyn ja ne tulee optimoida huolellisesti, jotta ne eivät aiheuta tarpeettomia hidasteita tietokannalle. Hidasteita syntyy esimerkiksi, jos triggeri sisältää monimutkaisia toimintoja tai suuri määriä tietokantakyselyitä.

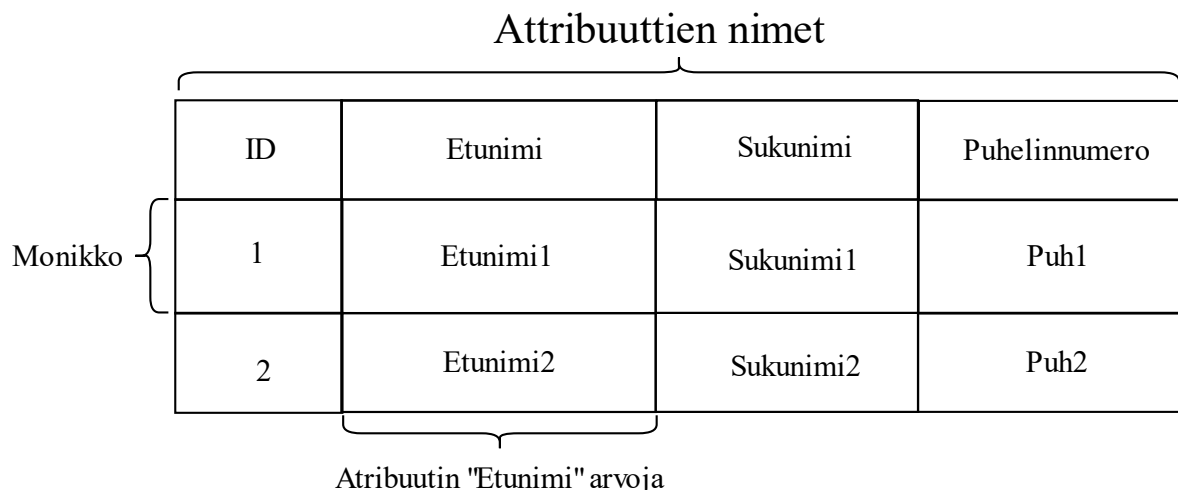
### 2.1.1 Relaatiomalli

Kuten jo aiemmin on tullut ilmi, tietokantoja voidaan rakentaa moniin eri tarkoituksiin ja niitä voidaan jakaa niiden käyttämän tietomallin mukaan (data model). Nykyisin tietokantoja tarkastellaan yleisimmin relaatiomallin kautta ja tämän mallin mukaan toteutettuja

tietokantoja nimitetään relaatiotietokannoiksi (Kaufmann & Meier, 2023; Laine, 2000). Historiassa on käytetty myös verkko- ja hierarkkista mallia (Laine, 2000.), joihin ei tässä työssä keskitytä.

Relaatiotietokantojen tiedot tallennetaan yleisesti tietokantoihin SQL tietokonekieltä käyttäen (Kaufmann & Meier, 2023; Lahtonen, 2002, 45; McQuillan, 2015). Tätä kieltä käytetään niin tietokannan rakenteen määrittelyyn ja muokkaamiseen, tietokantakyselyihin kuin myös tietojen lisäämiseen, poistamiseen ja muokkaamiseen. Tietokantoihin tehdyt tietokantakyselyt (query) perustuvat tietokantojen ominaisuuteen hakea tiedon sisällön perusteella kohteita ja käsitellä niitä. Kuten Harrington (2009, 5) kuvailee, tietokannassa oleva tieto sisältää ymmärryksen tietojen välisistä suhteista. Esimerkiksi yrityksen ostamien tuotteiden tietojen, kuten toimittaja, kappalemäärä ja tilauspäivämäärä, tulee olla yhteydessä toisiinsa. Näin ollen oikeanlaisilla ja tarkoilla tietokantakyselyillä voidaan hakea juuri tarvittavaa ja osaltaan hyvinkin yksityiskohtaista tietoa tietokannasta.

Relaatiotietokantojen tietoa voidaan esitellä selkeästi taulujen avulla. Yleisesti relaatiotietokannat koostuvat tauluista, jotka koostuvat sarakkeista ja nämä puolestaan sisältävät rivejä (Prabhu & Venkatesan, 2006, 65).



Kuvio 2. Relaation rakenne

Relaatiomallissa tietokannan taulua kutsutaan relaatioksi ja kuviossa 2 on esitelty esimerkki relaation rakenteesta. Relaatio sisältää yhden tai useamman attribuutin (sarake) ja monikon (rivi). Yksi attribuutti sisältää aina samankaltaista tietoa, esimerkiksi puhelinnumero attribuutti sisältää puhelinnumeroita. Harringtonin (2009, 86) mukaan relaatio on pari, joka koostuu joukosta attribuutteja ja monikkoja. Jokainen attribuutti koostuu yhdestä tai

useammasta monikosta, ja monikko sisältää attribuutteja, joiden arvo on järjestetty attribuutin nimen mukaan. Relaatiotietokannat sisältävät vastaavanlaisia tauluja kuin kuviossa 2, jotka sisältävät vaihtelevan määrän monikkoja ja attribuutteja ja näin ollen myös vaihtelevan määrän tietoa.

Attribuuttien ja monikkojen järjestyksellä ei ole väliä, mutta monikoiden on oltava erotettavissa toisistaan. Jokaisella monikolla tulee olla vähintään yksi attribuutti, jonka arvo on uniikki muihin monikoihin nähden. Perusavain (primary key) attribuuttina varmistaa, että monikot ovat uniikkeja ja se tulee olla määritelty jokaisessa relaatioissa. Kuviossa 2. perusavaimena toimii *ID* attribuutti ja sen arvo jokaisella monikolla omansa. (Beynon-Davies, 2004, 91–93.)

Toinen olennainen osa relaatioita on viiteavain. Viiteavaimella (foreign key) kytketään saman relaatiotietokannan eri relaatiot toisiinsa (Beynon-Davies, 2004, 93). Tämänlainen kytkentä voidaan muodostaa, kun yhden relaation monikon avainattribuutin arvo tallennetaan myös toisen relaation monikkoon (Laine, 2000).

### 2.1.2 Tietokannan hallintajärjestelmät

Tietokantojen hallintaan käytetään erilaisia tietokantojen hallintajärjestelmiä (DBMS). Tietokannan ideana on, että sen käyttäjän ei tarvitse huolehtia tiedon fyysisestä sijainnista. Sen sijaan käyttäjä voi tehdä tiedonkäsittelypyynnön, eli tietokantakyselyn, jonka jälkeen ohjelmisto vastaa tiedon noutamisesta ja tallentamisesta taustalla ilman, että käyttäjän tarvitsee itse huolehtia näistä toiminnoista (Harrington, 2009, 5). DBMS:sien avulla tietokannoista voidaan helposti organisoida ja hakea tietoja ja rivejä (Bhatia & Bansal, 2015).

Riippumatta DBMS tuotteesta, kuuluisi Harringtonin (2009) mukaan niistä löytyä seuraavat ominaisuudet: 1) DBMS:n on mahdollistettava tietokannan määrittäminen, mukaan lukien tietojen loogisen rakenteen ja näiden välisten suhteiden määrittäminen. 2) DBMS:n on tarjottava toiminnot tietojen muokkaamista varten, eli tietojen lisäämisen, muokkaamisen, järjestämisen, sekä poistamisen. 3) DBMS:n kuuluu mahdollistaa tietojen hakeminen, eli käyttäjän on voitava muotoilla tietokantakyselyjä perustuen tietojen loogisiin suhteisiin. 4) DBMS:n on tarjottava keinoja tietoturvan ylläpitoon. Esimerkiksi sen täytyy mahdollistaa käyttäjätunnusten ja salasanojen luonti tietokannan käyttäjille. Näin ollen voidaan varmistua

siitä, ketkä tietokantaa pääsevät hallinnoimaan ja käyttöoikeuksilla voidaan rajoittaa tietokannan eri osa-alueita eri käyttäjille.

Kuten Laine (2000) listaa, relaatiomallin perusteella toimivista DBMS:istä tunnetuimpia ovat mm. Oracle (2023), Microsoft SQL Server (2023) ja IBM:n DB2 (2023). Huomioiden, että listaus on pari vuosikymmentä vanha, kolme nykyään maailmanlaajuisesti suosituinta DBMS:ää Statistan (2023) mukaan ovat Oracle, MySQL (2023) ja Microsoft SQL Server. Listauksessa ei ole havaittavissa suurta muutosta yhden suosituimman DBMS:n muuttuessa.

Tässä työssä tarkastellaan tarkemmin Microsoft SQL Serveriä, lyhentäen se MSSQL Server, sillä se on myös kohdeyrityksen käyttämä DBMS. Jotta käsite voidaan ymmärtää, on hyvä tietää mitä palvelin (server) tarkoittaa. Nimensä mukaan palvelin eli tietokoneella toimiva ohjelma tarjoaa jonkinlaisen palvelun tai palveluita (Glowniak, 1998). Palvelin voi olla esimerkiksi sähköpostipalvelin, joka käsittelee sähköposteja tai tietokantapalvelin, joka käsittelee DBMS:sää. Kokonaisuudessaan SQL Server on kokoelma palveluita, jotka tukevat tiedon tallennusta ja hakuja jostain tietolähteestä kuten tietokannasta tai tietovarastosta (McQuillan, 2015).

MSSQL Server on Microsoftin kehittämä relaatiomalliin perustuva DBMS (Microsoft, 2023). Fosterin ja Godbolen (2014, 323–327) mukaan MSSQL Server on erityisesti suunniteltu ja kehitetty toimimaan, sekä maksimoimaan hyötynsä Windows-käyttöjärjestelmässä. Lisäksi riippumatta versiosta, MSSQL Serverillä on lukuisia eri ominaisuuksia, kuten tietokantatuki, replikointi-, integrointi-, analysointi- ja raportointipalveluja, sekä hallintatyökaluja (Foster & Godbole, 2014, 323–327). MSSQL Server takaa myös helpon asennuksen ja käytön. Integrointipalveluihin kuuluu mm. ETL, joka käydään läpi seuraavassa luvussa. Analysointipalveluihin kuuluu Online Analytical Processing (OLAP) eli yrityksen päätöksentekojärjestelmää tukeva tekniikka, jonka avulla voidaan järjestää suuria yritystietokantoja (Chaudhuri, & Dayal, 1997). Nämä ominaisuudet muodostavat kattavan kokonaisuuden tietokantojen hallintaan.

Microsoftin verkkosivuilta löytyy dokumentaatiota MSSQL Server versioista aina vuodesta 2005 lähtien tähän päivään asti, version 2005 ollessa vanhin ja 2022 uusin (Microsoft, 2023). Tämän työn kohdeyrityksellä on käytössä 2008 versio MSSQL Serveristä. MSSQL Server 2008 kykenee hallitsemaan valtavia tietokantoja, sekä sisältää monipuolisen joukon OLAP-komponentteja ja työkaluja erilaiseen tietojen käsittelyyn, analysointiin, sekä kuutioiden



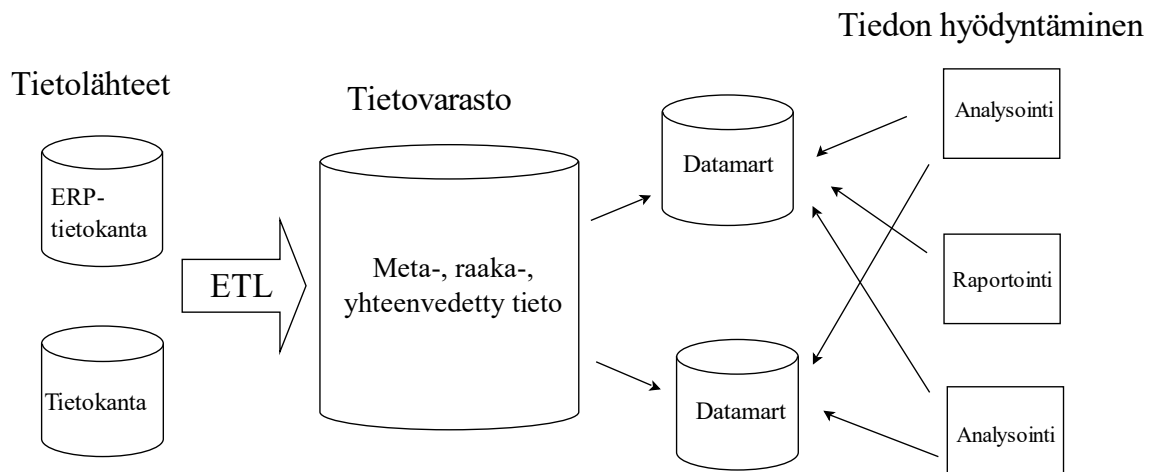
luomiseen (Nielsen & White & Parui, 2009, 3). Lisäksi se sisältää myös yllä mainitut yleiset MSSQL Serverin ominaisuudet.

Vuonna 2018 Microsoft julkaisi päivityksen, jossa kertoivat lopettavansa MSSQL Server 2008 -version tukemisen (Microsoft, 2020). He eivät esimerkiksi tee 2008 versiolle enää säännöllisiä tietoturvapäivityksiä. Suositeltiin, että yritykset, jotka tätä versiota käyttävät siirtyisivät uudempaan versioon, jotta suorituskyky, tehokkuus ja säännölliset tietoturvapäivitykset parantaisivat käytettävyyttä. Tätä tosiasiaa, että kohdeyrityksen käyttämä versio MSSQL Serveristä ei enää ole tuen piirissä, voidaan arvioida vaikuttavan kohdeyrityksen päätökseen päivittää menetelmää.

## 2.2 Tietovarastot

Edellytys aina suurempien tietomäärien varastoinnista, sekä tiedon hyötykäytöstä mm. yrityksen päätöksentekojärjestelmässä on johtanut tietovarastoihin perustuvien analyyttisten järjestelmien suunnitteluun (Vaisman & Zimányi, 2022, 3). Tietovarastojen avulla yritykset voivat saada johdonmukaisen näkemyksen oleellisista asioista ja tiedoista, joita heillä on käsiteltävänä (Velicanu & Matei, 2007). Esimerkiksi he voivat arvioida yrityksen tulosta, sekä tunnistaa tulevaa kehitystä. Mitä enemmän tietoa on sähköisessä muodossa, sitä enemmän yritykset ovat riippuvaisia helposti saavutettavasta-, jaettavasta- ja käsiteltävästä tiedosta, jonka avulla he voivat saavuttaa positiivista kehitystä (Rifaie & Kianmehr & Alhadj & Ridley, 2008).

Tietovarastot toimivat myös kätevästi historiallisen tiedon säilyttämisessä (Inmon & Linstedt, 2014, 42). Tällä tarkoitetaan sitä, ettei uudella tiedolla korvata vanhaa vaan kaikki tieto säilyy tietovarastossa. Lähtökohtaisesti tietovarastoissa on kuitenkin tarkoitus säilyttää tietoa, joka on jollain tavalla validia tiedon käsittelyn ja analysoinnin kannalta (Prabhu & Venkatesan, 2006, 72; Rifaie et al. 2008; Velicanu & Matei, 2007). Voidaankin ajatella esimerkiksi, että yrityksen omistaessa monta eri tietokantaa, siirtävät he vain analysoinnin kannalta merkittävät tiedot tietovarastoonsa, jättäen muut tiedot pois. Toisinaan tietovarastoissa voi olla myös historiallista tietoa, mikäli tämä on olennaista tiedon analysoinnin kannalta. Kuviossa 3. esitellään tietovarasto-ympäristön rakennetta, ja seuraavaksi käydään se läpi.



Kuvio 3. Tietovaraston rakenne

Yrityksen tietovarastoon päätyneet tiedot voivat olla peräisin monesta eri lähteestä, esimerkiksi eri tietokannoista, tai muista yrityksen sisäisistä ja ulkoisista lähteistä. Erilaista tietoa voi olla olemassa monessa eri muodossa. Tietolähteiden sisältäessä eri muodoissa olevaa tietoa tulisi se käsitellä yhtenäiseen muotoon, jotta sitä olisi helpompi hyödyntää jatkossa. Tästä on vastuussa Extract, Transform, Load (ETL) -prosessi, jonka mukaan tietolähteistä kerätään, käsitellään ja saatetaan tietoa tietovarastoon. Tiedon keräämisvaiheessa (extract) kerätään haluttu tieto kasaan tietolähteistä. Tämän jälkeen tieto täytyy käsitellä, jotta ne olisivat yhtenäisessä muodossa. Tässä tiedon käsittelyvaiheessa (transform) suodatetaan myös virheelliset ja ei-toivotut tiedot pois. Viimeisessä ETL-prosessin vaiheessa eli tiedon saattamisessa (load) käsitelty tieto saatetaan tietovarastoon tietyin väliajoin. (Vaisman & Zimányi, 2022, 91-92.)

Tietovarastot rakentuvat yleensä muista tietolähteistä tulevasta tiedosta ja usein niiden sisällä olevia tietoja on järjestetty helpommin analysoitavaan muotoon paikallisvarastoihin (data mart) (Vaisman & Zimányi, 2022, 67). Tietovarastojen voidaan ajatella sisältävän lähtökohteisesti yrityksen määrittelemän määrän paikallisvarastoja, jotka on suunnattu organisaation tiettyyn osa-alueeseen (Prabhu & Venkatesan, 2006, 95; Vaisman & Zimányi, 2022, 89). Esimerkiksi henkilökunnan, tietyn asiakkaan tai liiketoiminnan kannalta olennaisia tietoja säilytetään niille kohdennetuissa paikallisvarastoissa. Tietovaraston tietojen jako paikallisvarastoihin helpottaa suurien tietovarastokokonaisuuksien ymmärtämistä ja osaltaan myös rajoittaa käyttäjien oikeuksia kaikkien tietovaraston sisältämään tietoon. Kuten kuviosta 3. näkee, lähtökohteisesti yrityksen työntekijät pääsevät käsiksi edellä mainittuihin paikallisvarastoihin, joiden sisältämät tiedot ovat helposti luettavissa, sekä analysoitavissa.

Tietovaraston rakenne kannattaa olla suunniteltu niin, että tukee sen käyttäjien tarkoitusta raportoida ja analysoida strategisia tietoja. Näin ollen sen arkkitehtuurin tulisi olla mm. vankka, joustava ja skaalautuva (Rifaie et al. 2008). Tietovarastot tarjoavat käyttäjille mahdollisuuden erilaisiin analyysihin sinne säilytystä tiedoista (Velicanu & Matei, 2007), jolloin yritykset pystyvät kehittämään esimerkiksi heidän liiketoimintaansa.

### 2.3 Replikointi

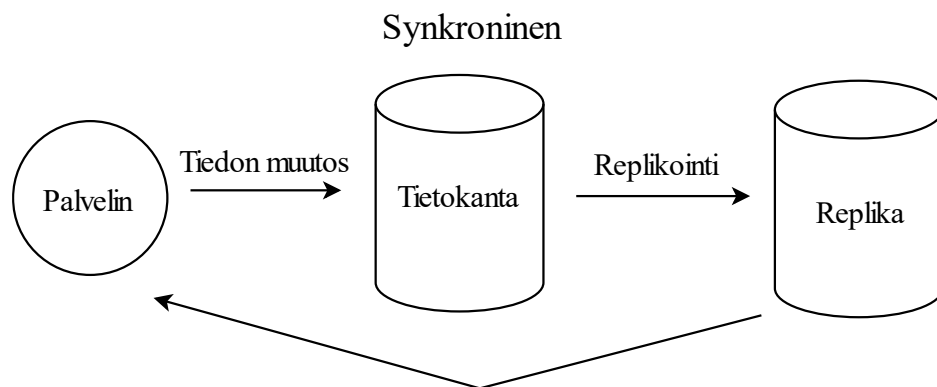
Tiedon tehokkaan käytön, kuten sen analysoinnin kannalta on tärkeää, että sitä on tarjolla niin, että mahdollisesti useat henkilöt pystyvät käsittelemään sitä samanaikaisesti. Tiedon replikointi on prosessi, jossa jostain tietolähteestä tiedot kopioidaan toiseen sijaintiin kuten tietovarastoon. Tällöin sama tieto on tallennettu useammalle tallennusalueelle. Replikoinnissa tiedon kopioita eli replikoita voidaan tehdä useampi, joka jakaa tietoa tällöin moniin samanlaisiin osiin. Nämä replikat voidaan siirtää tietokantaan tai -kantoihin ja siten useampi käyttäjä voi hyödyntää tietoa, joka on saatavilla useassa eri replikassa. Tämä on yksi replikoinnin ominaisuuksista, jonka tavoitteena on parantaa tiedon saatavuutta. (Mazilu, 2010.)

Replikointi mahdollistaa esimerkiksi tietojen säilyttämisen, mikäli alkuperäinen lähde vaurioituu (Mazilu, 2010). Se ei kuitenkaan ole replikoinnin päätarkoitus. Replikoinnin tarkoituksena on tiedon saatavuuden parantamisen lisäksi parantaa suorituskykyä (Munoz, Decker & Armendariz & Gonzales de Mendivil, 2007). Koska replikat ovat riippumattomia toisistaan on esimerkiksi replikan 1 tietoihin helppo päästä käsiksi ilman, että häiritäisi replikan 2 tietojen käyttöä, joka lisää suorituskykyä.

Tietokantojen replikoinnissa on ainakin yksi tietokanta, joka sisältää alkuperäisen tiedon, jonka lisäksi on tietokanta tai -kantoja, jotka sisältävät replikoita. Replikoinnissa on olennaista ylläpitää replikoita ja varmistaa, että tietokannat, jotka koostuvat replikoidusta tiedosta on synkronoitu alkuperäisen tietokannan kanssa. Synkronoinnilla tarkoitetaan prosessia, jonka avulla varmistetaan, että jokainen tietokannan kopio sisältää samat kohteet, sekä itse tiedot. Tietojen synkronoinnissa replikat, joiden tieto on muuttunut, päivitetään uudella ja ajankohtaisella tiedolla. (Mazilu, 2010.)

Replikointi on usein jaettu kahteen eri menetelmään, jotka on nimetty eri lähteiden mukaan hieman eri tavoin. Esimerkiksi muutaman lähteen mukaan nämä menetelmät on jaettu

innokkaaksi (eager) ja laiskaksi (lazy) (Gray & Helland & O’Neil & Shasha, 1996; Özsu, & Valduriez, 2020, 253–254). Toisinaan nämä kaksi menetelmää voidaan jakaa synkroniseen ja asynkroniseen replikointiin (Noor & Zian & Bahri, 2019). Tässä työssä jatkossa käytetään jälkimmäistä nimeämistapaa. Nämä menetelmät on luokiteltu sen mukaan, milloin päivitykset tehdään replikoihin ja missä nämä päivitykset sallitaan (Özsu, & Valduriez, 2020, 252–253).

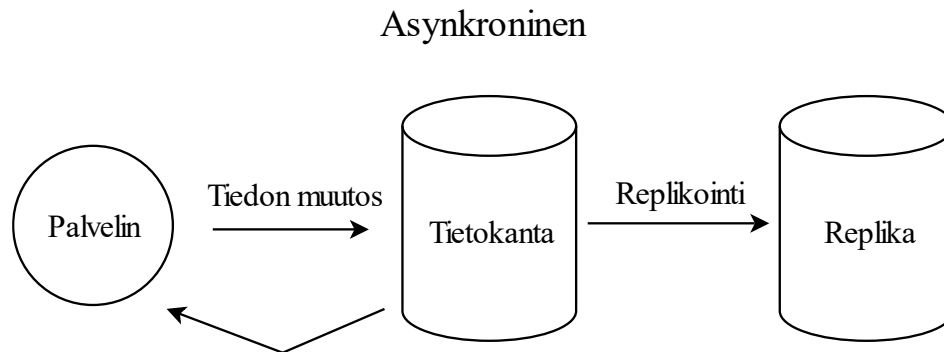


Kuvio 4. Synkroninen replikointimenetelmä

Synkronisessa replikointimenetelmässä päivitetään kaksi tai useampi replika samaan aikaan ja päivitys peruutetaan, mikäli jokin näistä epäonnistuu (Noor et al. 2019). Toisin sanoen synkronisessa menetelmässä muutos alkuperäisessä lähteessä päivitetään kaikkiin replikoihin päivitystapahtuman yhteydessä (Gray et al. 1996). Replikointiprosessi alkaa alusta vasta kun alkuperäiseen tietokantaan tehty muutos on rekisteröity ja replikoitu replikaan (Natanzon & Bachmat, 2013). Tämä tarkoittaa, että jokaisen muutoksen jälkeen replika on ajankohtaisesti synkronoitu alkuperäisen lähteen kanssa. Tämä voidaan myös huomata kuviosta 4. jossa esitellään yksinkertaistettu versio synkronisen replikointimenetelmän prosessista. Siinä palvelimelta tehdään muutos tietokantaan, esimerkiksi poistetaan jokin tieto. Tämän jälkeen se replikoituu suoraan replikaan. Vasta kun replikointi on suoritettu onnistuneesti, voidaan uusia muutoksia replikoida.

Synkronisen menetelmän eduksi muodostuu korkea käytettävyys, automaattinen vikasietokyky ja tietojen menettäminen on minimaalista (Noor et al. 2019). Vikasietokyky perustuu siihen, että tiedot replikoidaan lähes reaaliajassa, jolloin tieto on lähes aina ajankohtaista replikoissa. Toisinaan haitaksi koituu vähäinen joustavuus ja skaalautuvuus, sekä verkkoyhteyden kohdistuva paine muutoksien yhteydessä tapahtuvan replikoinnin ansiosta, jota voi tapahtua pienellä aikavälillä (Noor et al. 2019). Synkroninen menetelmä voi heikentää

suorituskykyä ja lisätä tapahtumavasteaikoja (Gray et al. 1996), esimerkiksi mikäli operatiivisessa tietokannassa, joka replikoidaan toiseen sijaintiin, tehdään jatkuvasti muutoksia, voi replikointiprosessi hidastua valtavasti.



Kuvio 5. Asynkroninen replikointimenetelmä

Asynkronisessa replikointimenetelmässä kaikki muutokset ensisijaisessa tietokannassa kirjataan heti, mutta niiden lähettäminen replikaan on ajastettu (Noor et al. 2019). Tässä replikointiprosessissa replikointia voidaan aktiivisesti jatkaa, vaikka replikasta ei olisi tullutkaan merkkiä siitä, että tiedoissa tapahtunut muutos olisi replikoitu onnistuneesti (Natanzon & Bachmat, 2013). Kuviossa 5. esitellään yksinkertaistettu asynkronisen replikointimenetelmän prosessi. Muutoksista tulleita tapahtumia kerätään ja ne replikoidaan samanaikaisesti tietyin väliajoin (Noor et al. 2019), esimerkiksi kerran minuutissa, tunnissa tai viikossa. Tässä menetelmässä replikat eivät aina ole ajankohtaisesti synkronisoitu lähdetietokannan kanssa toisin kuin synkronisessa replikointimenetelmässä.

Asynkronisen replikointimenetelmän hyviä ominaisuuksia ovat joustavuus ja skaalautuvuus (Noor et al. 2019). Koska asynkronisessa menetelmässä replikointi ei tapahdu reaaliajassa, sietää se enemmän esimerkiksi verkkoyhteyksien heikkenemistä kuin synkroninen menetelmä. Asynkronisen menetelmän suurimmaksi haitaksi muodostuu mahdolliset tietojen menetykset (Natanzon, & Bachmat, 2013; Noor et al. 2019). Tietoja voidaan menettää esimerkiksi, mikäli käyttökatkos tapahtuisi alkuperäisen tiedon tietokannassa ja ajankohtaisimpia tietoja ei olisi tallennettu replikoihin asynkronisen menetelmän aikaviiheen takia.

Lisäksi asynkronisessa replikointimenetelmässä voi ilmetä haitallinen tilanne, jossa tiedot eivät ole kopioituneet kaikkiin replikoihin. Tämän seurauksena osa tietoa hyödyntävistä tahoista voi olla kosketuksissa vanhaan tietoon replikassa tai replikoissa. Toisinaan alkuperäisen tiedonlähteen puolella ollaan kosketuksissa päivitetyn tiedon kanssa, jolloin voi ilmetä

erilaisia ristiriitoja ja epätarkkuuksia. Tämä epäsynkronisuus saattaa aiheuttaa ongelmia erityisesti silloin, kun päätöksentekoon liittyvä tieto kuuluisi olla jatkuvasti ajan tasalla. Tämänlaista tilannetta ei voisi syntyä synkronisen replikointimenetelmän kanssa, sillä siinä kaikki replikat ovat aina samanaikaisesti päivittyneet.

## 3 Tutkimusmenetelmä

Kvalitatiivisen eli laadullisen tutkimuksen tarkoituksena on ymmärtää jotain tapahtumaa tai toimintaa, tai siinä pyritään antamaan tulkinta jostain ilmiöstä (Eskola & Suoranta 1998). Tämän työn laadullisena tutkimuksena toimii tapaustutkimus, jonka avulla on tarkoitus ymmärtää jotain tiettyä tapahtumaa tai ilmiötä yksityiskohtaisemmin (Laine et al. 2015, 10). Tapaustutkimuksen ilmiönä toimii kohdeyrityksen replikointimenetelmä. Työssä tutkitaan niin kohdeyrityksen nykyistä- kuin uutta replikointimenetelmää ja sen merkitystä tarkemmin. Työn aineistonkeruumenetelmänä toimii puolistrukturoitu haastattelu.

Tässä luvussa käydään läpi työn kohdeyritys sekä kohdeyrityksen ajankotaisen projektin sisältö pääpiirteittäin. Tämän jälkeen luvussa 3.1.2 keskitytään kohdeyrityksen replikoinnin nykytilaan, jota pidetään myös osana työssä tehdyn tutkimuksen aineistoa. Luvussa 3.2 määritellään tutkimuksen aineistonkeruumenetelmä, eli työssä tehty haastattelu.

### 3.1 Kohdeyritys

Työn kohdeyritys on vuonna 1919 perustettu tukkukauppa, joka toimii palveluyrityksenä. Kaupan liiton (2023) mukaan tukkukaupat kehittävät tai ostavat erilaisia tuotteita, sekä palveluita ja myy niitä eteenpäin erilaisille toimijoille. Osa tuotteista saattaa tulla ulkomailta, jolloin tukkukaupan toimintaan liittyy myös maahantuonti. Kohdeyritys tuottaa niin logistiikka-, hankinta- kuin digitaalisia palveluita.

#### 3.1.1 Projekti

Kohdeyrityksellä on käynnissä laaja projekti, jonka tavoitteena on kehittää montaa eri osa-aluetta. Tähän projektiin sisältyy toiminnanohjausjärjestelmän eli ERP-järjestelmän tietokannan versiopäivitys, käyttöjärjestelmän vaihtaminen, tietovarasto ympäristön, sekä replikoinnin päivittäminen (Parviainen, 2023). Nämä kaikki projektin osa-alueet ovat hyvinkin olennaisia kohdeyrityksen tietojärjestelmän kannalta. Tämän alaluvun seuraavat kappaleet kuvailevat Parviaisen (2023) mukaan laajan projektin eri osia.

Kuten aikaisemmin on tullut ilmi, kohdeyritys käyttää ERP-järjestelmää, jonka ytimessä on heille tärkeä tietokanta. ERP:n ytimessä olevan Progress (2023) -tietokannan versiopäivityksessä versio päivitetään Progress 10.2B:stä uudempaan Progress 12.X:ään.

Projektin toinen osa-alue on käyttöjärjestelmän vaihtaminen. Käyttöjärjestelmät ovat lähes jokaisen tietokoneen ydin yhdistäen käyttäjän ja laitteiston, tarjoten käyttäjäystävällisen ympäristön (Chauhan, 2014). Kohdeyrityksen käyttöjärjestelmän vaihtaminen tapahtuu Advanced Interactive Executive (AIX) (IBM Corporation, 2022) käyttöjärjestelmästä Linuxiin.

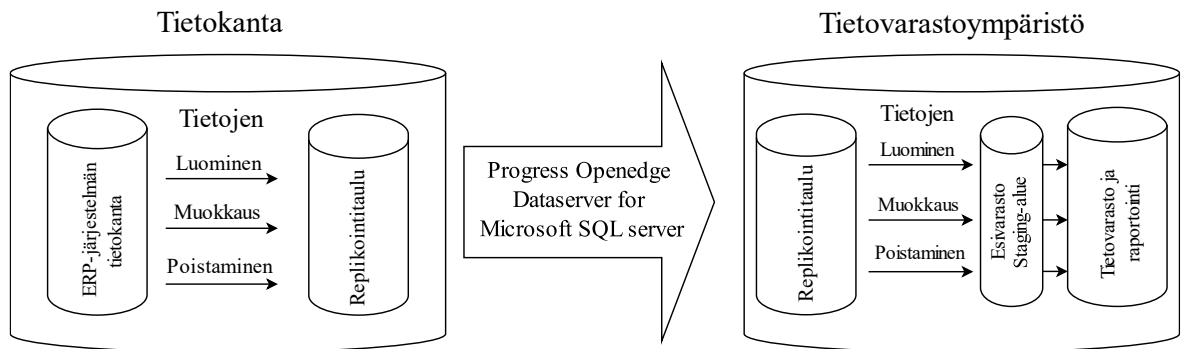
Kohdeyrityksellä on käytössä tietovarasto, johon päätyvä tieto on replikoitu ERP-järjestelmän tietokannasta. Tietovarastoympäristön päivittämisessä on kyse versiopäivityksessä, jossa MSSQL 2008 Server versio päivitetään uudempaan MSSQL 2019 Server versioon (Microsoft, 2023).

### 3.1.2 Nykyinen replikointimenetelmä

Yritykset käyttävät ERP-järjestelmää resurssien suunnitteluun. ERP-järjestelmät ovatkin usein integroitua niin, että sen ytimessä on yksi yhteinen tietokanta, jota eri toiminnot käyttävät. Yrityksien tärkeimpiä toimintoja ovat mm. ostot ja tilaukset, sekä talous ja varastonhallinta. Tällöin tietokannalla on suuri merkitys kohdeyritykselle, sillä tämän kautta nämä monet yrityksen tärkeät toiminnot toimivat. On kriittistä, että tietokanta ja kokonaisvaltainen tietojärjestelmä ovat ajantasalla ja palvelee yritystä niin, että he saavat siitä mahdollisimman suuren hyödyn. (Logistiikan maailma, 2023.)

Työn kohdeyrityksellä on käytössä tietokanta, joka toimii lähinnä yrityksen ERP-järjestelmän ytimenä ja se sisältääkin heidän liiketoimintansa kannalta tärkeimmät osa-alueet. Tämän ERP-järjestelmän tietokannan sisältämät tiedot tietojen replikointimenetelmä replikoi kohdeyrityksen tietovarasto ympäristöön. Yrityksellä on tällä hetkellä replikoinnissa käytössä Progress OpenEdge Dataserver for Microsoft SQL Server -ohjelmisto (Progress, 2023) ja replicate.sh -skripti. Jatkossa yrityksen nykyiseen ohjelmistoon verrataan lyhentäen se Dataserver for MSSQLSR. Yrityksen on tarkoitus päivittää nykyinen ohjelmisto Progress Open Edge pro2 -versioon. (Parviainen, 2023.)





Kuvio 6. Kohdeyityksen replikointiprosessi

Tarkastellaan kohdeyityksen replikointiprosessia, joka on myös havainnollistettu kuviossa 6. Kohdeyityksen replikointiprosessissa uudet ja muuttuneet rivit replikoidaan ensin tietokannan replikointitauluun triggerien avulla. Jokaisella kohdeyityksen replikoitavalla taululla on kolme triggeriä; luomis-, muokkaamis-, sekä poistamistriggerit (Parviainen, 2023). Parviaisen (2023) mukaan kohdeyityksen replikoinnissa koko rivi replikoidaan aina, mikäli sen sisällä edes yhden sarakkeen sisältämä tieto muuttuu.

Tiedot sittemmin siirretään tietokannan replikointitaulusta tietovarastoympäristön replikointitauluun. Tietovaraston replikointitaulusta tieto siirretään esivarastoon ja itse tietovarastoon jälkikäsitteilyyn. Samalla siirtyy tieto siitä, mistä taulusta tieto on peräisin, jotta voidaan varmistua siitä, että tieto siirretään sitä vastaavaan sijaintiin kohdereplikointitaulussa tietovarastoympäristössä. (Parviainen, 2023.)

Kohdeyityksellä on replikointitaulut niin ERP-järjestelmän tietokannassa kuin tietovarastoympäristön tietokannassa. Kaikkien replikoitavien taulujen sarakkeiden yhtenevissä nimissä niin tietokannan kuin tietovarastoympäristön tietokantapalvelimella tulee olla huomioitu taulun rakenteen nimi (Schemaholder). Tämän avulla ohjelma osaa viedä uuden sarakkeen tiedot oikeaan paikkaan tietovaraston replikointitaulussa. Esimerkiksi mikäli tietokannasta jo replikoituun tauluun lisätään sarake, tulee saman niminen sarake olla luotuna myös tietovaraston replikointitaulussa. (Parviainen, 2023.)

Replikoidessa uutta taulua ensimmäistä kertaa, tulee tietovarastoympäristön tietokantapalvelimella olla valmiiksi alustettu tietokantataulu täysin samalla nimellä kuin tietokannan replikointitaulussa. Alustuksella tarkoitetaan sitä, että tietokannan replikointitaulun rakenne (Schema) validoidaan Dataserver for MSSQLSR replikointiohjelmalle. Tämänlainen alustus voidaan tehdä myös tauluun, jossa on jo replikoita, mutta se lähtökohtaisesti kestää

pidempään riippuen sarakkeiden ja rivien määrästä. Jotta replikointiskripti toimii, täytyy tietovaraston kohdetauluun myös luoda kaksi saraketta, eli replikointiaika, sekä rivin yksilöivä arvo -sarakeet. Uuden kohdesarakkeen tietotyyppi määräytyy tietokannan lähdesarakkeen mukaan, mutta merkkimäärä on tietovarastoympäristössä asetettu suuremmaksi. Tämän avulla voidaan välttää tiedonsiirto-ongelmia tietokannan ylipitkistä kentistä verrattuna Open Edgen tietokannan taulun rakenteeseen. (Parviainen, 2023.)

Kuvion 6. mukaan tietovarastoympäristössä on replikointitaulun lisäksi tiedon jalostukseen tarkoitettu esivarasto, sekä itse tietovarasto ja raportointiin tarvittavat taulut. Kohdeyrityksen replikointi prosessissa Dataserver for MSSQLSR-ohjelmistoa ja replicate.sh -skriptiä käytetään tiedon siirtämiseen tietokannan replikointitaulusta tietovarastoympäristössä sijaitsevaan vastaavaan replikointitauluun. Replicate.sh käynnistyy minuutin välein ja se käsittelee tietokannan replikointitauluun kertyneet rivit. Replikointiskripti sisältää replikointiluokkia, joiden perusteella replikointitaulun sisältämät tiedot siirretään tietovaraston kohdereplikointitauluun. Luokat ovat 0 eli jatkuva siirtäminen, 1 eli kerran tunnissa, 5 eli puolentunnin välein, sekä 9 eli kerran vuorokaudessa kello 18:00. Näiden luokkien mukaan tiedon voidaan ajatella olevan tärkeysjärjestyksessä niin, että tärkeät tiedot siirtyvät tietovaraston replikointitauluun heti eli käyttäen luokkaa 0 ja loput ovat ajastettu tapahtumaan muiden luokkien mukaan. (Parviainen, 2023.)

Esitietovarastossa on Staging-alue, jossa tietoa jatkojalostetaan käytettäväksi erilaisiin tarpeisiin. Tämän jälkeen tiedot siirretään tietovarastoon raportointia ja analysointia varten. Käyttäjät ovat kosketuksissa tietokannassa sijaitsevaan ERP-järjestelmän tietokantaan, sekä tietovarastoympäristössä olevaan tietovarastoon ja raportointitauluihin. Kohdeyrityksen tarkoituksena on, ettei tietovarastoympäristön sisältämiin replikointitauluihin ole käyttäjäkosketusta. (Parviainen, 2023.)

### 3.2 Aineistonkeruumenetelmä

Tämän työn aineistonkeruumenetelmänä käytetään haastattelua, mikä mahdollistaa syventävän vuorovaikutukseen tutkijan ja haastateltavan välillä. Haastattelu auttaa tutkijaa hahmottamaan yrityksen replikointimenetelmää sekä siihen liittyviä päätöksiä. Lisäksi haastattelu tukee tapaustutkimuksen luonnetta mahdollistaessaan ilmiön eli replikointimenetelmän merkityksen tarkemman tutkimuksen. Haastattelun käsitellessä tiettyä ilmiötä voidaan

hankittavan aineiston määrää, sekä näkökulmaa rajata vaivattomasti (Alasuutari, 2011). Rajaus auttaa keskittymään tarkemmin tiettyyn aihealueeseen mikä edistää kokonaisvaltaiseman käsityksen muodostumista (Laine et al. 2015, 50–51).

Puolistrukturoitujen haastattelumenetelmien käyttö on lisääntynyt viime aikoina (Hirsjärvi & Hurme, 2022). Työn haastattelu toimii puolistrukturoituna haastatteluna, tarkemmin teemahaastatteluna, joka mahdollistaa esimerkiksi sen, että haastattelijalla voi muuttaa kysymyksiä, sekä niiden kysymisjärjestystä haastattelun edetessä (Hirsjärvi & Hurme, 2022; Hyvärinen et al. 2017). Lisäksi teemahaastattelussa kysymyksiä voidaan myös lisätä tarvittaessa (Hyvärinen et al. 2017), mikä helpottaa tarkentamaan vastauksia ja näin ollen voidaan varmistua mahdollisimman kattavasta, mutta tiivistä haastattelukokonaisuudesta.

Haastattelun avulla kerätystä aineistosta tulisi pystyä tekemään luotettavasti tutkittavasta ilmiöstä päätelmiä (Hirsjärvi & Hurme, 2022), joka lisää painetta niin haastattelun rajauksen kuin kysymyksien määrittelyille. Kysyttävien kysymysten ollessa suuressa roolissa tulevan aineiston sisällön kannalta (Hyvärinen et al. 2017), tulisi niiden suunnitteluun käyttää harkintaa. Järkevästi suunniteltu haastattelurunko, sekä kysymyksien laatu voi saada aikaan hyvin yksityiskohtaisen, tai toisinaan pelkistetyn ja epäselvän aineiston. Tämän työn haastattelurunko (Liite 1.) ja näin ollen kysymykset on suunniteltu etukäteen, vaikka niistä voidaankin tarpeen tullen poiketa. Teemahaastattelun luonne puolistrukturoituna haastatteluna luo mahdollisuuden avoimempaan keskusteluun haastateltavan kanssa, joka kehittää myös haastatteluympäristöstä luotettavamman, sekä myötäilevämmän haastateltavaa kohtaan.

Haastattelussa haastatellaan kohdeyrityksen järjestelmäpäällikköä. Haastateltavan valinta perustuu siihen, että hänellä on laaja tietämys yrityksessä tapahtuvasta laajasta projektista, sekä tietojen replikoinnista. Lisäksi hän on työskennellyt kohdeyrityksessä parin vuosikymmenen ajan, joten hänellä on näkemystä käytetyistä järjestelmistä, sekä menetelmistä pitkältä aikaväliltä. Tällöin tutkimuksen kannalta suurin hyöty on haastatella häntä.

Haastattelu pidetään etähaastatteluna Teams-sovelluksen avulla ja siihen on varattu 60 minuuttia aikaa. Haastattelun ohjaa haastattelijalla ja haastattelussa kysyttävistä kysymyksistä ei julkaista etukäteen. Haastattelukysymykset jakautuvat kolmeen osa-alueeseen eli teemaan, jotka ovat nykyinen replikointimenetelmä, uudistamisprosessi sekä uusi replikointimenetelmä. Haastattelun kysymykset ovat liitteessä 1. Haastattelun tarkoituksena on keskittyä lähinnä replikointimenetelmän uudistamisprosessiin, sekä uuteen replikointimenetelmään.

Haastattelussa käsitellään myös kysymyksiä, jotka liittyvät kohdeyrityksen nykyiseen replikointimenetelmään. Nämä kysymykset toimivat pääosin tarkentavia kysymyksinä, vahvistukseen kohdeyrityksen dokumentaatiosta luotua aineistoa.

Hirsjärven ja Hurmeen (2022) mukaan teemahaastattelussa hypoteesien eli olettamuksien asema on ongelmallinen ja ohjeistuksia niiden muodostamiseen ei voida antaa. Tämä johtuu lähinnä siitä, että teemahaastattelussa ollaan kiinnostuneita jostain ilmiöstä, sekä sen ominaisuuksista kuin ennalta asetettujen hypoteesien todentamisesta. Näin ollen ennalta asetettuja oletuksia teemahaastattelun tuloksista ei tehdä, vaan niitä pikemminkin löydetään teemahaastattelun keinoin.

Teemahaastatteluun kuuluen haastattelu nauhoitetaan (Hirsjärvi & Hurme, 2022) ja tämän jälkeen se litteroidaan eli puretaan kirjoittamalla haastattelu kokonaisuudessaan tekstiksi. Aineiston litteroinnin jälkeen aineistoa analysoidaan sisällönanalyysin mukaan, joka sopii erityisen hyvin teemahaastattelulla kerättyyn aineistoon (Tuomi & Sarajärvi, 2018). Sisällönanalyysin mukaan ensimmäisenä aineisto tiivistetään (Puusa & Juuti, 2020; Tuomi & Sarajärvi, 2018), eli aineistosta poistetaan epäolennaisia asioita ja vastauksia jäsenellään, sekä yhdistellään niiden yhteneväisyyksien mukaan. Tämän jälkeen aineisto ryhmitellään. Aineistossa pidetään jako kolmeen samaan teemaan kuin haastattelukysymyksissä, joita ovat nykyinen replikointimenetelmä, uudistamisprosessi, sekä uusi replikointimenetelmä. Aineiston analysoinnin jälkeen suoritetaan sen tulkinta.

## 4 Tutkimustulokset

Tässä luvussa käydään läpi tämän työn tutkimuksessa tehdyn aineistonkeruumenetelmän eli haastattelun tulokset ja kootaan aineisto kasaan. Luku on jaettu alalukuihin aineiston teemojen mukaan. Kokonaisuudessaan haastattelu kesti noin 40 minuuttia eli hieman alle sille asetun ajan. Haastattelussa haastattelija sai kysytyä suurimman osan haastattelurungon kysymyksistä (liite 1). Kuten aiemmin on tullut ilmi, puolistrukturoitu haastattelu sallii lisäkysymykset, joita myös tämän teemahaastattelun aikana kysyttiin muutama kappale. Tämän luvun alaluvuissa esiintyvät kursiiiset kohdat ovat suoria sitaatteja haastatteluaineistosta, jotka ovat peräisin haastateltavalta, eli kohdeyrityksen järjestelmäpäälliköltä.

### 4.1 Nykyinen replikointimenetelmä

Vaikka kohdeyrityksen nykyisestä replikointimenetelmästä on dokumentaatiota, eli aineistoa valmiiksi haastattelun ulkopuolelta, tuli haastattelussa kysytyä muutamia tarkentavia kysymyksiä nykyisestä replikointimenetelmästä. Haastattelu aloitettiin keskittymällä muutamisiin kysymyksiin kohdentuen kohdeyrityksen nykyiseen replikointimenetelmään.

Tutustuttaessa kohdeyrityksen nykyiseen replikointimenetelmään, ensimmäinen olennainen kysymys koskee sen alkuperää eli sitä, miten siihen on alun perin päädytty. Haastateltava kertoi, että ennen nykyistä menetelmäänsä heillä oli replikoinnin sijaan olemassa tiedon massa-ajo siirto kerran vuorokaudessa. Tämä on hänen mukaansa edelleen hyvin yleistä organisaatioissa, sisältäen kuitenkin omat huonot ominaisuutensa. Kuitenkin kohdeyrityksellä replikoituva tieto on avainasemassa esimerkiksi myös tuotannon ohjauksessa, joten tiedon reaaliaikaisuus on tärkeää. Myös reaaliaikaisempi raportointi oli kohdeyrityksen tavoitteena, joten alkoivat he käymään eri replikoinnin mahdollisuuksia läpi huomioiden ERP-järjestelmässä sijaitsevan Progress tietokantansa.

Nykyinen menetelmä valittiin ja sitä aloitettiin työstämään, sen ollen alussa hiukan haasteellista. Haasteellisuutta nykyisen menetelmän käyttöönotossa loi uuteen ohjelmistoympäristöön ja malliin totuttelu. Tämän myötä kohdeyrityksen tuli tehdä muutoksia niin Progressin

tietokannassa kuin tietovaraston puolelle, jotta replikointimenetelmä toimisi hyvin. Haastateltava kuvailee nykyisen replikointimenetelmän toimivan tällä hetkellä moitteettomasti:

*”Nythän se toimii erittäin hyvin, se on hyvin ratkassu meidän ton haasteen, meillä on hyvin reaaliaikainen replikointi plus raportointi. Eli sen [päätyamisen nykyiseen replikointimenetelmään] ohjasi nämä meidän nykyset järjestelmät tuon valinnan kanssa.”*

Haastateltavan mukaan replikointitietokannan sijainti niin ERP-järjestelmän tietokannassa kuin tietovarastoympäristön puolella on välttämättömyys. Replikointityökalu mikä heillä on käytössä nykyisen menetelmän kanssa, vaatii tämänlaisen replikointikannan olemassaolon ennen kuin tieto siirretään tietovaraston puolelle. Kuten aiemmin käydyn nykyisen replikointimenetelmän mukaan tiedot replikoituessaan siirtyvät triggerien avulla ensin ERP-järjestelmän tietokannasta replikointikantaan, jonka jälkeen ne siirretään tietovaraston vastaavaan replikointikantaan (Parviainen, 2023). Näin ollen määrittämällä triggereitä kohdeyritys voi valita mitä tietoja tietokannasta replikoituu tietovaraston puolelle, sekä estää joidenkin tietojen pääsyn replikointikantaan. Esimerkiksi suuret massapäivitykset ERP-järjestelmän tietokantaan sisältäessään lähinnä ylläpidollisia asioita ei vaikuta yrityksen tapahtumiin kuten saldo- tai myyntitapahtumiin, jolloin näiden tietojen ei tarvitse siirtyä tietokannasta tietovaraston puolelle.

*”-- Plus meille se ainakin tuo helpotusta siltä osin, että me pystytään rajaamaan et me ei kaikkia tietoja edes haluta sinne [tietovaraston puolelle], koska ne ei ole millään tavalla niinkun merkitseviä. Niin ne aiheuttaisi vaan ihan ylimääräistä replikoitavaa”*

Kohdeyrityksen lähdetietokanta eli kaikkia ERP-järjestelmän tietokannan sisältämiä tietoja ei replikoida kokonaan. ERP-järjestelmän tietokanta sisältää noin 130 taulua sisältäen myös tauluja, jotka eivät liity millään tavalla replikointiin. Tällaisia tauluja ovat taulut, jotka liittyvät ohjaukseen tai ovat aputaulua, joiden kautta tieto siirtyy niille tarkoitettuihin tauluihin, eli niitä käytetään vain hetkellisesti.

Kysyttäessä soveltuuko replikat varmuuskopiointiin, mikäli päätietokanta kaatuisi tuli haastateltavalta kieltävä vastaus. Tämä johtuu siitä, ettei replikat sisällä kaikkea ERP-tietokannan tietoa, johtuen aikaisemmin mainituista triggereistä, jotka osaltaan rajaavat replikoitavia tietoja. Lisäksi kohdeyritykseltä löytyy jo valmiiksi reaaliaikainen varmuuskopiointi, jota haastateltava kuvailee seuraavalla tavalla:

*”-- Ja meillä on ERP:issä reaaliaikainen backup:i [varmuuskopiointi], elikkä siis me oteetaan ihan niinkun päivänpäätteeksi toki backupit, mut meillä on semmoinen reaaliaikainen backuppi, et jos meillä tulee crashi [tilanne jossa tietokanta mm. kaatuu] niin meillä on sekunti siitä se data tallessa.”*

Haastateltavan mukaan tietojen määrä ERP-järjestelmän tietokannassa on ”paisunut kuin pullataikina” vuosien varrella, jolla viitataan tietokannan sisältämän tiedon määrän kasvuun. Kohdeyrityksen nykyinen ERP-järjestelmä on ollut ollut heillä vakituksessa käytössä vuodesta 1997 ja näin ollen kyseiseen tietokantaan on kertynyt paljon tietoa vuosien varrelta. Tällä hetkellä tietojen määrä kasvaa eksponentiaalisesti. Tähän tietokantaan on kehitetty ”automaattisuus” joka haastateltavan mukaan poistaa automaattisesti nykyhetkestä yli 10-vuotta vanhat tiedot. Toisinaan kohdeyrityksellä on tietovarastoympäristön puolella historiatietokanta, josta sinne tulleita tietoja ei poisteta ollenkaan. Kohdeyrityksen kaikki replikoitu tieto päätyy myös tähän tietokantaan ja sieltä voi hakea jopa 10 vuoden takaisia tietoja edelleen, tämän ollessa kuitenkin nykyään hyvin harvinaista.

Haastateltava painottaa, että nykyinen replikointimenetelmä on kehitetty nykyiseen muotoonsa pitkän ajan kuluessa, jonka myötä siitä on saatu toimiva. Heidän luokittelunsa mukaan ylimääräistä tietoa ei replikoidu vaan ainoastaan kaikki tarpeellinen tieto. Replikointi toimii itsekseen, on luotettava, sekä toiminut pitkään hyvin, vaikka menetelmän käyttöönoton jälkeen sitä paranneltiinkin muutaman vuoden ajan.

Nykyisellä replikointimenetelmällä on omat rajoituksensa, joka lähinnä kohdistuu menetelmän suorituskykyyn. Nykyinen replikointimenetelmä toimii moitteettomasti, mikäli siirrettävän tiedon määrä on maltillinen. Mikäli replikoitavaa tietoa on paljon tukkeutuvat ne replikointiluokkiin, niin ettei tietoa siirry tavoitteisiin nähden oikealla aikataululla tai tehokkuudella. Huomion arvoista on se, ettei niin suuria muutoksellisia rivimääriä pitäisi replikointiin tulla, että nykyinen replikointiprosessi tukkeutuisi täysin, ellei kyseessä olisi virhetilanne.

*”--nyt jos mä haluisin ton mejän nykyisen replikoinnin saada tukehtumaan niin mä saisin aika helposti. Eli jos mä yhteen yksittäiseen tauluun ampuisin semmosen viisisataatuhatta – miljoona päivitysriviä niin se ois sen jälkeen niinko tosi takkunen.”*

Mikäli replikointi täyttyy suurella määrällä tietoa, joutunee kohdeyrityksen järjestelmäpäällikkö korjata tilanteen manuaalisesti. Tällöin hän jakaa replikoitavia tietoja eri replikointiluokkiin, jotka suorittavat replikoinnin luokkaan määritellyn ajan mukaan. Hän voi myös käynnistää lisäluokkia auttamaan tiedon replikoituvuuden tehokkuutta. Lisäksi versioiden vanhuus on myös osoittautunut nykyisen replikointimenetelmän heikkoudeksi, jolloin uuteen versioon päivittäminen ei tapahdu enää niin kätevästi. Nykyjärjestelmät ovat yksinkertaisesti tehokkaampia.

## 4.2 Uudistamisprosessi

Kohdeyrityksellä on haastateltavan mukaan muutama selkeä tavoite uuden replikointimenetelmän kannalta. Tavoitteista ensimmäinen on menetelmän nopeuden lisääminen. Nykyisen replikointimenetelmän heikkoutena on suurien tietomäärien käsittely, sillä ne tukkeuttavat replikointiluokat helposti. Tavoitteena on, ettei uuden menetelmän kanssa tulisi olemaan tämänlaista ongelmaa, joka mahdollistaa menetelmälle lisää nopeutta ja näin ollen myös tehokkuutta.

*”Niin nyt kun me ollaan tota uuttapuolta testattu niin siin ei tuu tämmöst [tukkeutumista] ollenkaa. Et me ollaan niinko ihan keinotekosesti aktivoitu 10 000 riviä tai 100 000 riviä, josta sit tippuu eri, eri tauluihin niit päivitettävii [tietoja], niin se menee sit niiko minuuteissa, kun nyt [nykyisen replikointimenetelmän kanssa] se saattaa kestää monii, monii tunteja.”*

Toinen tavoite uuden replikointimenetelmän kannalta on varmuuden lisääminen. Nykyinen replikointimenetelmä perustuu ERP-tietokannassa oleviin triggereihin, joka ei menetelmänä ole niin varma kuin uusi replikointimenetelmä sen perustuessa itse tietokantamuutoksiin. Näin ollen menetelmän perustuessa itse tietokantamuutoksiin on se haastateltavan mukaan tällöin aukoton. Haastateltava kuitenkin mainitsee jälleen, että nykyinen replikointimenetelmä, perustuen triggereihin, toimii aikaisemmin kerrotun ”hiomisen” tuloksena, eli kohdeyritys on jalostanut menetelmän toimimaan niin, että ERP-järjestelmästä replikoituu ainoastaan valitut tiedot. Lisäksi uuden replikointimenetelmän tavoitteena on riisua siitä henkilösidonaisuutta, haastateltavan näin viitaten siihen, että menetelmän seuranta ja valvonta mielellään ulkoistetaan.

*”-- tavoite on mun mielestä se että se [uusi replikointimenetelmä] toimii yhtä aukottomasti [kuin nykyinen menetelmä], et se ei vaadi sit niiku mitään manuaalista työtä --”*

Edellä mainitut tavoitteet toimivat osaltaan rohkaisuna kohdeyritykselle päivittää nykyistä menetelmää muiden syiden ohessa. Replikointimenetelmän päivittämiselle on olennaisia syitä. Esimerkiksi uusi replikointimenetelmä tulee olemaan tuen piirissä, eli uudemmat järjestelmät saavat tukea niiden valmistajilta mm. käyttöönottoon, ongelmatilanteisiin, sekä päivitystarpeisiin. Lisäksi päivittäminen on välttämätöntä, sillä nykyinen menetelmä ei sovi laajan projektin myötä päivitettävien muiden järjestelmien kanssa yhteen.

Nykyinen replikointimenetelmä ei sovi yrityksen laajan projektin myötä käyttöön otettavien järjestelmien versioiden kanssa yhteen, joka osaltaan pakottaa myös replikoinnin uudistamisen. Kysyttäessä eri vaihtoehtoja uudelle replikointimenetelmälle kertoi haastattelija, että



heillä oli hyvä sauma siirtyä astetta korkeamman teknologian versioon, joka edesauttoi uuden menetelmän, eli Progress Open Edge Pro2 (Progress, 2023), hankinnan.

Kohdeyrityksellä on nykyisenä toimittajana suuri informaatioteknologia-alan (IT) yritys, joka tuottaa konsultointi- ja ulkoistuspalveluita niin Suomessa kuin ulkomailla. Toimittajalla tarkoitetaan tässä kontekstissa yritystä, jolta ostetaan palveluita. Tällä hetkellä kohdeyrityksen tietovarastoympäristö on tämän aiemmin mainitun toimittajan hallinnassa. Replikointimenetelmän uudistamisprosessiin vaadittuihin resursseihin kuuluu kohdeyrityksen toimittajan puolelta kaksi henkilöä ja kohdeyrityksen puolelta itse haastateltava, järjestelmäpäällikkö. Tarpeiden mukaan uudistamisprosessiin saadaan lisää henkilöitä ja näin ollen lisää resursseja mukaan.

Replikoinnin uudistamisprosessin aikana olennaista on, että tietojen eheys säilyy ennallaan, eli tietoja ei menetetä. Kohdeyrityksellä on muutama eri keino varmistaa tietojen eheyttä. Niistä ensimmäisenä on kohdeyrityksen tarkoitus suorittaa raakatestejä eli vertailla tietokantaa tehtyjen kyselyjen avulla, että taulujen rivimäärät täsmäävät molemmissa järjestelmissä. Kohdeyritys valitsee jonkin tietyn ajankohdan, kuten kuukauden tai vuoden taaksepäin, jonka myötä raakatesti suoritetaan.

Toisinaan kohdeyritys varmistaa tietojen eheyttä tekemällä taulukohtaisen vertailun ”replikoinnin päässä”. Tällä haastateltava tarkoittaa, että vertailu tehdään tiettyyn tauluun ERP-järjestelmän replikointitietokannassa, että tietovarastoympäristön puolella. Mikäli eroja löytyy näiden välillä samassa taulussa, voidaan ne päivittää. Tässä ei kuitenkaan ole järkeä vertailla suurta määrää tietoja vaan vain viimeisimmät, sillä jossain tauluissa saattaa olla jopa kymmeniä miljoonia rivejä, haastateltava huomioi.

Kolmas keino, jonka haastateltava mainitsi tietojen eheyden varmistamiseen replikointimenetelmän uudistamisprosessin aikana, on mahdolliset automatisoinnit tietojen tarkistamiseen. Tällä hetkellä kohdeyrityksen tarkoituksena on selvittää erilaisia mahdollisuuksia toteuttaa haluttu automaatiikka. Tämänlainen automatisointi loisi varmuutta siitä, että tietoja ei ole poistunut uudistamisprosessin aikana. Automaatio perustuisi jonkinlaiseen häiriöilmoitukseen, johon esimerkiksi järjestelmäpäällikkö tai muu kohdeyrityksen työntekijä voi reagoida. Tämä häiriöilmoitus ilmoittaa, mikäli jokin tieto ei kulje tai ei ole siirtynyt järjestelmien läpi. Haastateltavan mukaan tämänlainen ilmoitus on normaalistikin häiriötilanteessa varsin kätevä tapa saada selville, mikäli ongelmia on havaittu.

*”-- meillä on nyt tossa selvityksessä vielä se, että saako siihen jonkun automatiikan siltä osin, että nousee joku lippu pystyyn, jos joku tieto ei meekkään niinkon läpi.”*

Haastateltavan mukaan riskit, joita kohdeyritys tunnistaa replikointimenetelmän uudistamisprosessissa ovat hyvin ihmisvaikutteista. Käytännössä tietojen varmistus, sekä riskien välttäminen on paljolti kohdeyrityksen työstä kiinni. Itse riskejä ei ole tarkemmin tunnistettu. Koska replikointimenetelmän uudistaminen on osa laajaa projektia, tulee haastateltavan mukaan kaikki järjestelmät olla testattu niin, että jokainen toiminto toimii yhtä hyvin kuin tällä hetkellä nykyisien menetelmien kanssa. Näin ollen kohdeyritys rakentaa uuden ympäristön kokonaan nykyisen ympäristön rinnalle, jonka jälkeen tälle suoritetaan tuotantotesti. Ympäristöllä tarkoitetaan koko laajan projektin muodostamaa suurta järjestelmä kokonaisuutta. Tämän tarkoituksena on ajaa kaikki muutokset läpi niin kuin nykyisessäkin siirrosta ja jos ja kun ongelmia tulee vastaan, niitä korjataan.

Kohdeyrityksen laajan projektin sisältäessä paljon muitakin osa-alueita kuin tiedon replikointimenetelmän päivittämisen on hyvinkin oleellista ymmärtää, kuinka monta osaa laajasta projektista tulee olla toteutettuna ennen kuin uusi replikointimenetelmä voidaan ottaa käyttöön. Haastateltava painottaa, että jokainen laajan projektin osa-alue pitää olla testattu, sekä otettu käyttöön, jotta uutta replikointimenetelmää voidaan käyttää.

*”No siis kaikki pitää olla valmiina, siinä ei oo mitään, että ei voi yksikään osuus olla niinkon testaamatta tai todettu et se ei ois valmis et se pitää kaikki olla [toteutettuna].”*

#### 4.3 Uusi replikointimenetelmä

Kohdeyritys ei ole vielä päässyt ottamaan käyttöön uutta replikointimenetelmää. Haastateltava kertoi, että laajan projektin toteutuksen aikataulu tarkentuu, kun suurin osa koko projektin ja siten myös replikointimenetelmän keskeisistä testeistä on suoritettu. Haastattelun aikaan kohdeyrityksen ainoa vielä testattava kokonaisuus oli tuotannonverkkopohjaisten sovellusten testaus. On kuitenkin selvää, että replikointimenetelmän ja projektin käyttöönotto menee vuoden 2024 puolelle.

Haastateltavan mukaan, kun kohdeyrityksen uusi replikointimenetelmä on otettu käyttöön, siitä ei juurikaan seuraa helpotusta, mutta samalla ei myöskään monimutkaista hänen työnsä. Uuden replikointimenetelmän ylläpitoa tullaan siirtämään aiemmin mainitun toimittajan piiriin, jolloin toimittaja huolehtii kohdeyrityksen replikoinnin valvonnasta. Mikäli

joitain toimenpiteitä joudutaan tekemään katsovat kohdeyritys ja toimittaja kumman suunnalta toimenpide toteutetaan.

*”--tottakai me ollaan varauduttu, että se alku niin sitä pitää myös valvoa, siis että tulee jälkiseurantaa siihen, et tehdään niinku niitä vertailuja siinä, vaikka mitään ei olisi virheessä, eikä ketään ois ilmottanu mistään ongelmist, ni tottakai me niinku halutaan todentaa että ne kaikki toimii niinkun pitääkin.”*

Kohdeyrityksen mittareihin, joilla he arvioivat replikointimenetelmän päivittämisen onnistumisen, kuuluu menetelmälle asetettujen tavoitteiden onnistumien arviointi. Tavoitteiden arviointi tapahtuu mm. virheiden määrän, siihen menneiden työtuntien ja selvitysten tarkasteluna. Nämä asiat tulevat näkymään helposti uuden menetelmän käyttöönoton jälkeen. Kun uusi replikointimenetelmä on saatu käyttöön, on kohdeyrityksen tavoitteena se, ettei siihen tarvitsisi koskea. Näin ollen menetelmä toimii mutkitta itsekseen, kuten haastateltavan mukaan perus integraatiopalvelu tosin hieman kompleksisempänä versiona.

Haastateltavalta myös löytyy luottamusta siihen, että uusi järjestelmä tulisi saamaan hyvällä tahdilla käyttöön, sekä ennen kaikkea toimivaksi. Tähän hän luottaa siksi, että he tietävät valmiiksi erilaisia tilanteita ja kohtia, joissa nykyisen järjestelmän kanssa on tullut ongelmia. Tällöin heidän on helpompi reagoida niihin. Kuitenkin, kun koko laaja projekti otetaan käyttöön voi suuressa kuvassa niin replikoinnissa kuin sen ulkopuolella olla ongelmia, joita joudutaan kohtaamaan.

*”-- nyt me tiedetään niinku noi tietyt sudenkuopat ni mä luulen et se, jos nyt toivoa sais nii sehän menis niinku ykkösellä että ei tarvii tehä mitään, mutta aina pitää olla vähän pessimisti et jotainhan varmasti tulee vastaan. -- Mut sitte ku testataan sitä koko ketjuu niin se on sit toinen juttu, mutta vahva usko on että saadaan se ongelmitta käyttöön hyvin nopeessa ajassa.”*

Tiedon määrää, jota siirtyy replikoinnissa, on mahdotonta sanoa. Haastateltavan mukaan rivejä voi siirtyä päivätasolla ja päivän mukaan sadoistatuhansista miljooniin. Kohdeyrityksen keräysprosessissa yhteen riviin voidaan tehdä jopa kahdeksan muutosta. Tietoihin voi tulla muutoksia tuotteen kulkiessa mm. keräyksen, kuittauksen, sekä rahdituksen läpi. Vaikka tämä tietokannan rivi voi saada monta päivitystä, ei rivi kuitenkaan monistu moneksi riviksi. Haastateltavan mukaan tällä tavoin ajatellessa, vaikka tapahtumia on paljon niin rivejä ei kuitenkaan ole tolkkuttomasti.

Rakenteeltaan uusi replikointimenetelmä tulee sisältämään samanlaista logiikkaa kuin nykyisessä menetelmässä mutta tietovaraston puolella tulee olemaan joitain muutoksia. Esimerkiksi tietovaraston puolella tietokantataulut tullaan tekemään hieman eri tyylillä kuin

nykyisessä menetelmässä. Muuten tietovaraston logiikka pyritään pitämään samanlaisena kuin nykyhetkellä. Myös replikointikannat tulevat jatkossakin olemaan niin ERP-järjestelmän kuin tietovarastoympäristön puolella ja tietovaraston puolella tullaan jatkossakin säilyttämään historiallista tietoa sille tarkoitettussa historiatietokannassa. Tälle historiatietokannan sisältämälle tiedolle ei kohdeyrityksellä ole mitään valmista suunnitelmaa jatkon kannalta, eli tilanteen mukaan sieltä poistetaan tietoa vanhimmasta päästä.

Uusi replikointimenetelmä on lähtökohtaisesti nopeampi kuin nykyinen, ottaen huomioon, että jokaisella järjestelmällä on omat suorituskykynsä rajat. Haastateltava mainitsi, että yhdessä testissä he syöttivät satatuhatta päivitettyä riviä replikointiin, ja uusi menetelmä käsiteli ne minuutissa. Koska uusi menetelmä tulee olemaan reaaliaikaisempaa sen perustuessa ERP-järjestelmän tietokannassa tapahtuviin muutoksiin, tulee myös virheet menemään läpi tietovaraston puolelle heti. Esimerkiksi näppäilyvirheet voivat aiheuttaa suuria muutoksia tilattavien tuotteiden tilausmääriin. Toisinaan haastateltava kertoo, että tällaisia virheitä voidaan välttää erilaisilla rajoituksilla kuten kappalemäärien tupla varmistuksella. Tämänlaisen tarkoitukseen tehtäviä tarkistuksia voidaan tehdä moneen paikkaan, mutta niitä tulisi satoja.

Säilytettävän tiedon määrän kasvulla ei ole merkittävää vaikutusta replikoinnille. Replikoinnissa siirtyy tietty määrä valittua tietoa ja ns. vanha tieto tilastoidaan, jolloin siihen ei pysty koskea. Näin ollen tilastotilassa oleva tieto ei liiku enää replikoinnissa ja tiedon määrän kasvu ei siltä osin kuormita replikointia.

Kysyttäessä kohdeyrityksen replikointimenetelmän tulevaisuudesta on vastaus ymmärrettävä. Kohdeyrityksellä ei ole tällä hetkellä suunnitelmia päivittää uutta replikointimenetelmää. Koska replikointimenetelmä on osa laajempaa järjestelmien kokonaisuutta, on se näin ollen myös alttiina uudistuksille. Haastateltavan mukaan kohdeyritys harkitsee uutta vaihtoehtoa ERP-järjestelmälle ja tämä muutos voisi vaikuttaa myös replikointimenetelmään. Koska nykyinen replikointimenetelmä on osaltaan riippuvainen kohdeyrityksen nykyisestä ERP-järjestelmästä, tulisi sen vaihtuessa myös replikointi tehdä uudella tavalla. Haastateltava kuitenkin vakuuttaa, että nykyinen, että uudistettu replikointi toimisi mallina myös mahdollisille uusille replikointimenetelmille, eikä esimerkiksi massa-ajot ole enää vaihtoehto.

*”-- jos tää ERP:i vaihtuu taustalta, niin ei me tätä [nykyistä tai uutta replikointimenetelmää] voida käyttää enää. Et sit me toki kokemuksen pohjalta ni haetaan vastaavan tyyppistä ratkaisua.”*

## 5 Keskustelu

Tässä luvussa tuodaan esille tutkimuksen tärkeimpiä tuloksia mm. tutkimuskysymyksiin vastaamalla. Luvun 5.1 tarkoituksena on myös pohtia tutkimustulosten merkitystä aineistoa tulkiten ja hyödyntäen. Tulkittaessa aineistoa siitä tehdään johtopäätöksiä, joiden myötä selvitetään, saavuttiko tutkimus sille asetetut tavoitteet. Luvussa 5.2 käsitellään tutkimuksen luotettavuutta ja rajoitusten merkitsevyyttä.

### 5.1 Tulosten pohdinta ja merkitys

Tämä työ kartoittaa kohdeyrityksen käyttämiä replikointimenetelmiä, sekä luo ymmärrystä nykyisen menetelmän uudistamiselle. Tarkoituksena oli tarkastella yrityksen replikointimenetelmän päivittämistä, vertailla nykyisen ja uuden menetelmän eroja, sekä selvittää replikointimenetelmän merkitystä. Kokonaisuudessaan tutkimuksessa käytetty aineisto sisältää tietoa kohdeyrityksen replikointimenetelmän nykytilasta, replikoinnin uudistamisprosessista ja uudesta replikointimenetelmästä. Haastattelun tulokset heijastavat kohdeyrityksen järjestelmäpäällikön näkemystä edellä mainituista asioista.

Ominaisuudet, joita replikointi pyrkii takaamaan, on esitelty työn johdannossa. Näitä ovat luotettavuus, saatavuus, sekä suorituskyky (Nasibullin & Novikov, 2020). Luotettavuus yhtenä replikoinnin ominaisuutena on kohdeyrityksen tilanteessa hieman merkityksellinen. Luotettavuus takaa tiedon olemassaolon aina jossain replikassa. Kohdeyritys ei käytä replikoita tietojen varmuuskopioina, joten tästä ominaisuudesta ei ole suoranaista hyötyä kohdeyritykselle. Muut replikoinnin ominaisuudet eli saatavuus, sekä suorituskyky ovat avain asemassa kohdeyrityksen kannalta.

Kuten aiemmin työn taustasta on tullut ilmi, replikointi voidaan jakaa synkroniseen- ja asynkroniseen replikointiin. Kohdeyrityksen nykyinen replikointimenetelmä on malliltaan asynkronista replikointia, sillä siinä replikointia jatketaan aktiivisesti replikointikantaan, vaikka tästä tietokannasta muutokset eivät olisivat vielä siirtyneet eteenpäin tietovaraston puolelle. Tämä perustuu kohdeyrityksen asettamiin ajastuksiin, eli replikointiluokkiin, joiden mukaan muutoksia siirretään tietyin ajoin eteenpäin.

Käsiteltäessä kohdeyrityksen replikointimenetelmän päivittämistä on hyvä ymmärtää replikoinnin nykytila. Työn ensimmäiset tutkimuskysymykset ovat ”Miten kohdeyrityksen replikointi on toteutettu nykytilanteessa? Entä uuden menetelmän myötä?”. Kohdeyrityksen dokumentaatio sisältää tietoa heidän nykyisestä replikointimenetelmästä ja siitä luotu aineisto (luku 3.1.2) vastaa tutkimuskysymyksistä ensimmäiseen. Yhteenvedona nykyinen replikointimenetelmä replikoi tietoa kohdeyrityksen ERP-järjestelmässä olevasta tietokannasta tietovaraston puolelle. Nykyinen replikointimenetelmä perustuu triggereihin, jotka ovat osoittautuneet tehokkaiksi erityisesti kohdeyrityksen tietojen replikointiin, johtuen myös niiden, sekä itse menetelmän pitkästä kehittämisestä sen käyttöönoton jälkeen.

Kohdeyrityksen dokumentaation myötä luodun aineiston avulla nykyisen replikointimenetelmän rakenne kuin -prosessi oli suoraviivaista selvittää. Muutama asia kuitenkin herätti mielenkiintoa nykyisen replikointimenetelmän toteutuksessa, jonka takia haastattelussa käytiin läpi myös nykyiseen menetelmään liittyviä asioita. Esimerkiksi nykyisen replikointimenetelmän rakenne herätti pientä kummastusta replikointitietokannan sijaitessa niin ERP-järjestelmän kuin tietovaraston puolella (Kuvio 6.). Kuten haastattelun tuloksista käy ilmi, on replikointitietokanta olemassa molemmissa olennaisesta syystä, eli replikointityökalun sitä vaatiessa.

Taulukko 1: Nykyisen replikointimenetelmän arviointi

Vahvuudet	Heikkoudet
<ul style="list-style-type: none"> <li>- Pitkän testauksen tuotos</li> <li>- Yhteensopivuus</li> <li>- Tiedon saatavuus</li> <li>- Enemmän hallintaa triggerien avulla</li> </ul>	<ul style="list-style-type: none"> <li>- Hidas</li> <li>- Vanhat järjestelmät, ei tukea</li> <li>- Tukkeutumisvaara</li> <li>- Triggerit hidastavat</li> <li>- Henkilösidonnainen</li> </ul>

Lisäksi mielenkiintoa herätti replikointimenetelmän vahvuudet ja heikkoudet, joita on lisätty taulukkoon 1. Vahvuuksista ensimmäinen viittaa nykyisen replikointimenetelmän vakauteen ja toimivuuteen. Pitkän kehityksen ja testauksen jälkeen on oletettavaa, että menetelmä on hiottu niin hyvään malliin, että erilaisia virhetilanteita esiintyy harvoin. Nykyinen replikointimenetelmä voidaan katsoa olevan hyvin optimoitu. Lisäksi nykyisen replikointimenetelmän vahvuutena korostuu järjestelmien yhteensopivuus sekä replikoinnissa että laajemmin muiden järjestelmien kanssa. Nykyinen replikointimenetelmä takaa tiedon

saatavuutta raportointiin, mikä onkin yksi replikoinnin keskeisistä tehtävistä. Vahvuuksiin kuuluu myös ERP-tietokannassa sijaitsevat triggerit, sillä niiden avulla kohdeyritys voi määrittää replikoitavaa tietoa kätevästi kohdentaessaan triggerit vain raportoinnin kannalta merkitykselliseen tietoon.

Haastattelusta luodusta aineistosta kuitenkin nousi esille replikointimenetelmän haasteet sen nopeudessa, mikäli replikoitavien tietojen määrä kasvaa. Vaikka nykyinen replikointimenetelmä onkin hyvin vakaa, sekä saavuttaa ominaisuuksia, joita replikointimenetelmältä odotetaan, on se kuitenkin suhteellisen hidaskäyttöinen, sekä alttiina tukkeutua, mikäli replikoitavaa tietoa on paljon. Nykyisen replikointimenetelmän hitaus voi osittain johtua myös järjestelmäversioiden vanhentumisesta tai replikoinnin apuna käytettävistä triggereistä. Kuten työn taustasta käy ilmi, triggerit voivat hidastaa tietokantoja esimerkiksi sisältäessään monimutkaisia toimintoja tai suuria määriä tietokantakyselyjä. Tämän voidaan nähdä vaikuttavan myös kohdeyrityksen nykyisen replikointimenetelmän suorituskykyyn ja nopeuteen, sillä kaikki replikointiin menevä tieto on määritelty nimenomaan näiden triggerien avulla. Viimeinen listattu kohdeyrityksen nykyisen replikointimenetelmän haaste on sen henkilösidonaisuus, mikä tarkoittaa, että replikointia valvoo ja mahdollisesti korjaa työntekijä.

Toisen tutkimuskysymyksen vastaus, joka käsittelee kohdeyrityksen tietojen replikointia uuden menetelmän myötä, on johdettu haastattelun tuloksista muodostuneesta aineistosta. Uuden replikointimenetelmän rakenne on hyvin samanlainen kuin nykyisen. Erona tulee olemaan joitain muutoksia tietovarasto ympäristön puolella, tarkemmin siellä sijaitsevien tietokantataulujen tyylien muuttumisessa. Suurin muutos nykyiseen menetelmään on tietojen reaaliaikaisempi, tietokantamuutokseen perustuva replikointi. Tällöin triggereitä ei ole ERP-järjestelmän puolella replikoinnin tarkoituksiin. Koska kohdeyritys ei halua replikoida kaikkea tietoa ERP-järjestelmän tietokannasta, on heillä käytössä jatkossakin replikointikannat, sekä replikointiluokat.

Replikointitietokantojen pysyminen niin ERP-järjestelmän, kuin tietovarastojen puolella tarkoittaa myös sitä, että replikoinnista tehdään osittain ajastettua. Kohdeyritys on katsonut replikointiluokkien olevan kätevä vaihtoehto jakaa replikointiin menevän tiedon määrää ja näin ollen replikoinnin kuormitusta. Koska replikointi perustuu suoraan tietokantamuutoksiin, on se malliltaan synkronista replikointia. Kohdeyrityksen muodostamat replikointiluokat kuitenkin estävät menetelmän toimintaa synkronisena menetelmänä, sillä tällöin replikointi ei siirry suoraan tietovarasto ympäristön puolelle kumoten synkronisen menetelmän

pääidean. Uuden replikointimenetelmän voidaan näin ollen katsoa olevan hybridi mallinen menetelmä, sisältäen ominaisuuksia niin asynkronisesta- kuin synkronisesta menetelmästä.

Koska kohdeyrityksen replikointimenetelmä on toiminut moitteettomasti, herättää se lisää mielenkiintoa siihen, miksi menetelmää on päätetty lähteä päivittämään ja miksi juuri uusi menetelmä on valittu. Työn viimeinen tutkimuskysymys on ”Mitä vaihtoehtoja replikointimenetelmässä on harkittu ja miksi juuri projektin myötä tuleva menetelmä valittiin?”. Uuden replikointimenetelmän vaihtoehtoja mietittäessä on olennaista tiedostaa, että kohdeyrityksellä ei varsinaisesti ollut minkäänlaista ongelmaa heidän nykyisen replikointimenetelmänsä kanssa. Raskaimpana syynä menetelmän päivittämiselle toimi muiden järjestelmäpäivitysten ohessa tullut tarve päivittää myös replikointimenetelmää. Lisäksi kuten työn taustasta tuli ilmi, Microsoft lopetti MSSQL Server 2008 -version tukemisen vuonna 2018, jota voidaan myös pitää yhtenä kohdeyrityksen motiivina päivittää replikointimenetelmänsä nykyisestä uuteen. Kohdeyritys katsoiärkevimmäksi vaihtoehdoksi päivittää nykyisen replikointimenetelmän versiota uudempaan, sillä se on yhteensopiva muiden laajan projektin myötä tulleiden järjestelmäpäivityksien kanssa. Lisäksi uusi versio sisältää merkityksellisiä parannuksia nykyiseen replikointimenetelmän versioon nähden.

Taulukko 2: Uuden replikointimenetelmän arviointi

Vahvuudet	Heikkoudet
<ul style="list-style-type: none"> <li>- Tehokas</li> <li>- Uudet järjestelmät ja tuki</li> <li>- Ei triggereitä hidastamassa</li> <li>- Vapauttaa henkilöstöä (pitkällä aikavälillä)</li> <li>- Tiedon saatavuus ja suorituskyky</li> </ul>	<ul style="list-style-type: none"> <li>- Virheiden mahdollisuus lisääntyä tietokantamuutoksia siirrettäessä</li> <li>- Ei triggereitä avustamassa tiedon siirrossa</li> <li>- Kehitysaika, testausaika</li> <li>- Uuden järjestelmän käytön opettelu</li> </ul>

Uuden replikointimenetelmän valintaan vaikuttaneet asiat tulivat ilmi tutkimusaineistosta. Menetelmän suunnittelussa ja näin ollen tavoitteissa keskityttiin erityisesti korvaamaan nykyisen menetelmän heikkouksia (ks. taulukko 1). Taulukossa 2 on listattu uuden replikointimenetelmän vahvuuksia ja heikkouksia, joita tarkastellessa voidaan selvittää uuden replikointimenetelmän ominaisuuksia. Tarkastellaan ensin uuden replikointimenetelmän vahvuuksia. Uudella replikointimenetelmällä pyritään mm. nopeuden ja luotettavuuden



lisäämiseen. Nopeutta lisää uuden replikointimenetelmän uusi teknologia. Luotettavuutta lisää replikointimenetelmän perustuminen triggereiden sijasta tietokantamuutoksiin. Nämä itessään nostavat uuden replikointimenetelmän tehokkuutta ja toisinaan tehokkuutta parantaa se, ettei lukuisia triggereitä ole hidastamassa prosessia.

Yksi merkittävimmistä vahvuuksista on myös se, että uusi replikointimenetelmä kuuluu tuen piiriin. Näin ollen kohdeyritys voi varmistaa esimerkiksi yrityksen tietoturvaa ja päivittää ohjelmistonsa ajantasaisiksi erilaisten järjestelmäpäivitysten myötä. Toisinaan uudet järjestelmät myös takaavat parempaa suorituskykyä nykyiseen menetelmään verrattuna. Lisäksi mm. työkalut ovat uudempia, joiden avulla menetelmän ylläpito on nykyaikaisempaa.

Oleennaista on myös uuden menetelmän henkilösidonnaisuuden riisuminen ja replikointimenetelmän hallinnan ulkoistaminen, joka on merkittävä uuden replikointimenetelmän tuoma etu kohdeyritykselle. Uuden replikointimenetelmän hallinnan ulkoistaminen mahdollistaa vakaan pohjan replikoinnin suorituskyvylle, sekä ylläpidolle.

Käydään seuraavaksi läpi heikkouksia, mitä uudessa replikointimenetelmässä voi ilmetä. Yksi suurimpia haasteita mitä uusi replikointimenetelmä aiheuttaa, on virheiden replikoitumisen riskin nousu. Koska uuden menetelmän avulla saadaan suuria määriä tietoa nopeasti replikoitumaan tietovaraston puolelle, kulkeutuu myös virheet samassa tahdissa jopa asiakkaiden nähtäville. Virheiden lisäksi myös paljon mm. raportoinnin kannalta turhaa tietoa voi siirtyä tietovaraston puolelle. Nykyisessä replikointimenetelmässä virheitä on torjuttu mm. triggerien avulla. Koska näitä triggereitä ei enää ole uudessa menetelmässä, tulee virheet estää muilla keinoilla.

Nykyisessä replikointimenetelmässä triggereillä oli erittäin olennainen rooli niiden suodattaessa replikoitavaa tietoa. Uudessa menetelmässä triggereitä ei ole, vaan replikointi perustuu suoraan tietokantamuutokseen. Näin ollen triggerien puute voidaan nähdä uuden menetelmän heikkoutena, sillä uudessa menetelmässä ne eivät ole rajoittamassa replikoitavaa tietoa.

Lisäksi uuden replikointimenetelmän heikkouksina voidaan pitää myös järjestelmien opetelu-, kehitys-, sekä testaus vaiheita. Nämä eri vaiheet tulevat viemään kohdeyritykseltä aikaa, sekä resursseja. Lisäksi on tärkeää huomata, että nämä vaiheet eivät ainoastaan vaadi aikaa ja resursseja, vaan voivat myös aiheuttaa mahdollisia haasteita uuden replikointimenetelmän sovittamisessa ja integroinnissa kohdeyrityksen laajan projektin myötä luotuun

uuteen toimintaympäristöön. Näiden seikkojen huomioiminen on olennaista varmistettaessa sujuvaa siirtymistä uuteen replikointimenetelmään.

Tehdyn tutkimuksen perusteella näyttää siltä, että replikointimenetelmän uudistaminen on perusteltua ja replikointi on kohdeyritykselle merkityksellistä. Merkityksen luo kohdeyrityksen tarve saada reaaliaikaisempaa tietoa heidän raportointiinsa ERP-järjestelmän tietokannasta. Replikointi luo yhteyden ERP-järjestelmän toiminnallisen tietokannan, sekä yrityksen tietovarasto ympäristön välille. Uusi replikointimenetelmä tulee olemaan yhteensopiva kohdeyrityksen laajan projektin myötä tulevien muiden järjestelmien kanssa. Lisäksi se tulee nostamaan replikoinnin nopeutta, sekä suorituskykyä suhteessa nykyiseen menetelmään. Yhteenvedona voidaan todeta, että kohdeyrityksen uuden replikointimenetelmän käyttöönotto vaikuttaa harkitulta, sekä hyvin suunnitellulta. On mielenkiintoista nähdä, miten muutos tulee vaikuttamaan kohdeyrityksen toimintaan pitkällä aikavälillä. Tulevaisuuden kannalta tutkimusta voisi laajentaa kohdistamaan useampaan yritykseen, tarkastellen laajempaa kuvaa replikoinnin toteuttamisesta.

## 5.2 Tutkimuksen luotettavuus

Laadullisen tutkimuksen arviointi pelkistyy tutkimusprosessin uskottavuuden, sekä luotettavuuden tarkastelemiseksi (Eskola & Suoranta 1998; Tuomi & Sarajärvi, 2018). Tutkimuksen luotettavuutta voidaan tarkastella esimerkiksi valitsemalla tarkoituksenmukaiset lähestymistavat ja menetelmät, sekä erityisesti perustelemalla niiden valinta (Puusa & Juuti, 2020). Työn tutkija tavoittelee päättelypoluista mahdollisimman läpinäkyviä niin tutkimuksen eri vaiheiden havainnollistamisessa kuin aineiston analysoinnissa ja tulkinnassa. Näin myös mahdollistetaan se, että tutkija kykenee ratkaista tutkimusongelman ja toteuttaa tutkimuksen luotettavasti.

Työn tuloksissa tulee ottaa huomioon myös tapaustutkimus työn tutkimusmenetelmänä. Itse tutkimusmenetelmä ei ole lähtökohtaisesti luotettava tai epäluotettava (Puusa & Juuti, 2020), vaan luotettavuutta tulee tarkastella erilaisten näkökulmien kautta ja suhteuttaa tutkittavaan ilmiöön. Tapaustutkimuksessa tarkastellaan ilmiötä, joten on hyvä huomioida, että työssä keskitytään nimenomaan kohdeyrityksen replikointimenetelmään, sekä sen uudistamiseen. Tietyn organisaation ilmiön tarkastelu voi luoda haasteita tulosten yleistämisen kannalta.

Eskolan ja Suorannan (1998) mukaan laadullisen tutkimuksen lähtökohta on tutkijan subjektiivisuus, jolla viitataan viittaa tutkijan henkilökohtaiseen tulkintaan, näkemyksiin ja kokemuksiin, jotka vaikuttavat osana tutkimusprosessia. Näin ollen tutkijan subjektiivisuus vaikuttaa laajasti työssä tehtyihin päätöksiin niin käytettyjen lähteiden valintaan kuin aineiston läpikäyntiin ja sen analysointiin. Työssä on kuitenkin pyritty valitsemaan luotettavia lähteitä, sekä toimintamalleja.

Haastattelun rajaus vain yhteen henkilöön voi aiheuttaa aineistoon yksipuoleisen näkökulman, joka tulee ottaa huomioon aineistoa lukiessa ja arvioidessa. Tämänlainen rajaus saattaa vaikuttaa tutkimuksen luotettavuuteen negatiivisesti mm. mikäli haastateltavalle kehittyy painetta vastata tietyllä tavalla esimerkiksi yrityksen maineen kannalta. Huolimatta tästä rajauksesta, haastattelusta saatiin kuitenkin muodostettua hyvä määrä hyödyllistä aineistoa.

Rajaus kyseiseen haastateltava toisinaan myös luo luotettavuutta aineistolle ja siten koko tutkimukselle, sillä hänellä on laaja ja pitkäaikavälin käsitys työn aiheesta, eli kohdeyrityksen replikointimenetelmästä. Luotettavuutta vahvisti myös haastattelussa kysytyt kysymykset kohdeyrityksen nykyisestä replikointimenetelmästä, sillä nämä käsitelivät osaltaan tutkimuksen taustaan, sekä kohdeyrityksen dokumentaatioon liittyviä asioita.

Haastattelussa saadusta aineistosta voi tulla epäluotettava, mikäli haastattelija ohjaa tai johdattelee haastateltavaa tietynlaisiin vastauksiin. On kuitenkin huomionarvoista, että haastattelija on tietoinen tästä mahdollisesta haasteesta ja pyrki aktiivisesti välttämään tällaista tilannetta. Aineiston käsittelyssä pyrittiin käyttämään mahdollisimman paljon suoria lainauksia aineistosta, jotta lukija pystyy seurata tutkijan tekemää päättelyketjua (Puusa & Juuti, 2020). Tämänlainen käytäntö luo luotettavuutta aineistoon ja siitä tehtyihin päätelmiin.

## 6 Yhteenveto

Tässä työssä tutkittiin kohdeyrityksen replikointimenetelmän merkitystä. Työn tavoitteena oli selvittää, minkälainen kohdeyrityksen nykyinen replikointimenetelmä on, sekä miten uusi menetelmä tulee eroamaan siitä. Tavoitteena oli luoda käsitys menetelmien eduista ja haasteista, sekä selvittää miksi kohdeyritys on päätenyt juuri näihin menetelmiin. Tutkimus toteutettiin tapaustutkimuksena ja aineistonkeruumenetelmänä käytettiin teemahaastattelua.

Replikointimenetelmän uudistaminen oli hyvin ajankohtaista kohdeyritykselle, jonka myötä tutkimus sai alkunsa. Tiedon määrän kasvaessa sähköisessä muodossa on liiketoiminnan kannalta tärkeää, että yritykset pystyvät käyttämään, sekä analysoimaan tietoja reaaliaikaisesti. Replikointi on avainasemassa esimerkiksi tietojen saatavuuden kannalta ja se toimii olennaisena yhteytenä kohdeyrityksen ERP-järjestelmän, sekä tietovarasto-ympäristön välillä, mikä lisää aiheen tärkeyttä etenkin kohdeyritykselle.

Tutkimuksen tuloksena saatiin koottua käsitys niin kohdeyrityksen nykyisestä kuin uudesta replikointimenetelmästä. Rakenteeltaan uusi menetelmä tulee pysymään samanlaisena, huomioiden jotkin muutokset tietovarasto-ympäristön tietokantojen taulujen rakenteessa. Pääidea replikoinnissa pysyy ennallaan. Uusi menetelmä on nykyistä menetelmää nopeampi, sekä luotettavampi ja se tulee jatkossakin takaamaan replikoinnin ominaisuuksia, kuten tiedon saatavuuden, mahdollistaen mm. lisää suorituskyykyä.

Tutkimuksessa onnistuttiin saavuttamaan työlle asetetut tavoitteet. Uusi menetelmä korvaa nykyisen menetelmän heikkouksia hyvin, tuoden vain muutamia uusia heikkouksia mukanaan. Nykyisten järjestelmien vanha ikä, järjestelmätuen puute, sekä tavoite reaaliaikaisemmasta replikoinnista ohjasivat kohdeyrityksen uuden replikointimenetelmän valintaa ja loi merkitystä uudistamisprosessille ja uudelle replikointimenetelmälle.

Jatkotutkimuksessa olisi harkittava monitapaustutkimusta tai tapaustutkimuksen yhdistämistä muihin tutkimusmenetelmiin, mikä voisi tukea laajempaa yleistettävyyttä ja antaa monipuolisemman näkökulman tutkittavaan ilmiöön. Tämä voisi tarkoittaa myös mahdollisten vertailukohteiden, eli toisten organisaatioiden lisäämistä tutkimukseen, mahdollistaen laajemman aineiston. Tällöin tutkimustuloksia voidaan yleistää enemmän ja niistä voisi hyötyä myös muut tahot, jotka suunnittelevat replikointimenetelmänsä uudistamista.

## Lähteet

- Alasuutari, P. (2011) Laadullinen tutkimus 2.0. Tampere: Vastapaino.
- Beynon-Davies, P. (2004) Database systems. Third edition. London: Palgrave Macmillan.
- Bhatia, A. B. & Bansal, V. (2015) Database management system. Oxford, Englanti: Alpha Science International Ltd.
- Cassandra. (2024) Open Source NoSQL Database. Viitattu 23.1.2024. Saatavissa: [https://cassandra.apache.org/\\_/index.html](https://cassandra.apache.org/_/index.html)
- Chaudhuri, S. & Dayal, U. (1997) An overview of data warehousing and OLAP technology. SIGMOD record. 26 (1), 65–74.
- Chauhan, N. (2014) Principles of Operating Systems. Oxford University Press.
- Dalton, P. & Whitehead, P. (2001) SQL Server 2000 Black Book. Phoenix: Paraglyph Press.
- Dewson, R. (2015) Beginning SQL Server for Developers Fourth Edition. Fourth Edition. Berkeley, CA: Apress.
- Eskola, J. & Suoranta, J. (1998) Johdatus laadulliseen tutkimukseen. Tampere: Vastapaino.
- Foster, E. C. & Godbole, S. V. (2014) Database Systems A Pragmatic Approach. Berkeley, CA: Apress.
- Glowniak, J. (1998) History, structure, and function of the internet. Seminars in nuclear medicine. 28 (2), 135–144.
- Gray, J., Helland, P., O’Neil, P. & Shasha, D. (1996) The dangers of replication and a solution. SIGMOD record. 25 (2), 173–182.
- Gupta, D. & Rani, R. (2019) A study of big data evolution and research challenges. Journal of information science. 45 (3), 322–340.
- Han, J., E, H., Le, G. & Du, J. (2011) Survey on NoSQL database, 2011 6th International Conference on Pervasive Computing and Applications, Port Elizabeth, 2011, 363-366.
- Harrington, J. L. (2009) Relational database design and implementation clearly explained. 3rd ed. Amsterdam : Morgan Kaufmann/Elsevier.

HBase. (2024) Viitattu 23.1.2024. Saatavissa: <https://hbase.apache.org/>

Hirsjärvi, S. & Hurme, H. (2022) Tutkimushaastattelu : teemahaastattelun teoria ja käytäntö. [2. painos]. Helsinki: Gaudeamus.

Hyvärinen, M., Nikander, P. & Ruusuvaori, J. (2017) Tutkimushaastattelun käsikirja. Tampere: Vastapaino.

IBM (2023) IBM Db2. Viitattu 23.10.2023. Saatavissa: <https://www.ibm.com/products/db2>

IBM Corporation. (2022) IBM AIX operating system for IBM Power. New Orchard Road, Armonk, NY 10504: IBM Corporation.

Inmon, W. H. & Linstedt, D. (2014) Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault. Waltham, MA: Elsevier Science.

Kaufmann, M. & Meier, A. (2023) SQL and NoSQL Databases: Modeling, Languages, Security and Architectures for Big Data Management. Second edition. Cham: Springer.

Kauppan liitto. (2023) Toimialat. Viitattu 11.10.2023. Saatavissa: <https://kauppa.fi/kauppan-ala/toimialat/>

Kitchin, R. (2014) The data revolution: big data, open data, data infrastructures & their consequences. Los Angeles, California: SAGE Publications Ltd.

Lahtonen, T. (2002) SQL ToolKit. Jyväskylä: Docendo.

Laine, H. (2000) Tietokantojen perusteet. Helsingin yliopisto.

Laine, M., Bamberg, J. & Jokinen, P. (2015) Tapaustutkimuksen taito. 3. painos. Helsinki: Gaudeamus Helsinki University Press.

Logistiikan maailma. Toiminnanohjausjärjestelmä. Viitattu 9.10.2023. Saatavissa: <https://www.logistiikanmaailma.fi/logistiikka/ohjausjarjestelmat/toiminnanohjausjarjestelma/>

Mazilu, M. C. (2010) Database replication. Database Systems Journal, 1(2), 33-38

McQuillan, M. (2015) Introducing SQL Server. Berkeley, CA: Apress.

Microsoft. (2020) Announcing new options for SQL Server and Windows Server 2008 End of Support. Viitattu 29.10.2023. Saatavissa: <https://learn.microsoft.com/en-us/lifecycle/announcements/sql-server-windows-server-2008-end-of-support>

Microsoft (2023) SQL Server technical documentation. Viitattu 26.10.2023. Saatavissa: <https://learn.microsoft.com/en-us/sql/sql-server/?view=sql-server-ver16>

MongoDB (2023) Viitattu 15.1.2024. Saatavissa: <https://www.mongodb.com/>

Munoz, F. D., Decker, H., Armendariz, J. E. & Gonzales de Mendivil, J. R. (2007) Database replication. Institute Technology of Information, Technical Report TR-ITI-ITE-07/19.

MySQL (2023) Viitattu 23.10.2023. Saatavissa: <https://www.mysql.com/>

Nasibullin, A. R. & Novikov, B. A. (2020) Replication in Distributed Systems: Models, Methods, and Protocols. Programming and computer software. 46 (5), 341–350.

Natanzon, A. & Bachmat, E. (2013) Dynamic Synchronous/Asynchronous Replication. ACM transactions on storage. 9 (3), 1–19.

Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL databases and its comparison with relational databases. International Journal of Applied Information Systems, 5(4), 16-19.

Nielsen, P., White, M. & Parui, U. (2009) Microsoft SQL server 2008 bible. 1st edition. Indianapolis, IN: Wiley.

Noor, A. S. M., Zian, N. F. M & Bahri, F. N. M. S (2019) Survey on replication techniques for distributed system. International journal of electrical and computer engineering (Malacca, Malacca). 9 (2), 1298–.

Oracle. (2023) Database Management. Viitattu 23.10.2023. Saatavissa: <https://docs.oracle.com/en-us/iaas/database-management/index.html>

Parviainen, T. (2023) Tiedonsiirto eli replikointi kirjapakista (Progress) Datacenteriin (Ms Sql).

Rifaie, M., Kianmehr, K., Alhajj, R. & Ridley, M. J. (2008) Data warehouse architecture and design, in 2008 IEEE International Conference on Information Reuse and Integration. 2008 IEEE. 58–63.

Ploetz, A. & Kandhare, D. & Kadambi, S. & Wu, X. (2018) Seven NoSQL Databases in a Week: get up and running with the fundamentals and functionalities of seven of the most popular NoSQL databases. 1st edition. Birmingham: Packt Publishing

Prabhu, S. & Vēnkaṭēcan, N. (2007) Data mining and warehousing. New Delhi: New Age International P Ltd., Publishers.

Progress. (2023) Viitattu 23.10.2023. Saatavissa: <https://www.progress.com/>

Puusa, A. & Juuti, P. (2020) Laadullisen tutkimuksen näkökulmat ja menetelmät. Helsinki: Gaudeamus.

Sivula, A., Aho, M. & Laukkanen, M. (2023) Datasta liiketoimintaan : 10 tehokasta työkalua. Helsinki: Alma Talent.

Statista. Ranking the most popular database management systems worldwide, as of September 2023. Viitattu 11.10.2023. Saatavissa: <https://www.statista.com/statistics/809750/world-wide-popularity-ranking-database-management-systems/>

Tuomi, J. & Sarajärvi, A. (2018) Laadullinen tutkimus ja sisällönanalyysi. Uudistettu laitos. Helsinki: Kustannusosakeyhtiö Tammi.

Vaisman, A. & Zimányi, E. (2022) Data warehouse systems : design and implementation. Second edition. Berlin, Germany: Springer.

Van Der Lans, R. F. (2012). Data replication for enabling operational BI. White Paper on Business Value and Architecture, 1-26.

Velicanu, M. & Matei, G. (2007) Database versus data warehouse. Editura Economică.

Özsu, M. T. & Valduriez, P. (2020) Principles of Distributed Database Systems. 4th ed. 2020. Cham: Springer International Publishing.



## Liite 1. Haastattelukysymykset

Nämä haastattelukysymykset ovat luotu työn tutkimusosuutta varten. Tämän teemahaastattelun tavoitteena on ymmärtää tarkemmin yrityksen nykyistä replikointimenetelmää, sekä projektin myötä käyttöön otettavaa menetelmää. Lisäksi tämä liite toimii haastattelurunkona haastattelijalle.

### Nykyinen replikointimenetelmä

1. Miten päädyitte juuri tähän menetelmään aikoinaan?
  - 1.1. Miten päädyitte kahteen identtiseen replikointitauluun toisen sijaitessa tietokannassa ja toisen tietovarastoympäristössä?
2. Mitä haittoja koette olevan tässä menetelmässä?
  - 2.1. Minkälaisia konkreettisia ongelmia olette kohdanneet?
3. Mitä hyötyjä koette olevan tässä menetelmässä?
4. Replikoitteko lähdetietokannan kokonaan?
5. Kuinka paljon historiallista tietoa säilytätte tietovarastossa?
6. Käytättekö replikoita tarvittaessa varatietokantana eli varmuuskopiona, jos pää tietokanta jostain syystä kaatuu?

### Uudistamisprosessi

7. Mitä tavoitteita teillä on uudistetun replikointimenetelmän osalta?
8. Minkälaisia resursseja ja tukea tarvitsette uudistamisprosessiin?
9. Miten varmistatte, että tietojen eheys säilyy siirtymän aikana? Millä tarkistatte tämän?
10. Mitä riskejä ja ongelmia tunnistatte päivityksessä? Miten aiotte käsitellä niitä?
  - 10.1. Miten varaudutte mahdollisiin teknisiin ongelmiin tai katkoksiin replikoinnissa päivityksen aikana?

11. Kuinka suuri osa laajasta projektista tulee olla tehty ennen replikointimenetelmän päivittämistä?

11.1. Mitkä projektin osa-alueet ovat kriittisessä asemassa replikoinnin kannalta?

11.2. Kuinka varmistetaan, että vanhojen ja uusien järjestelmien välinen siirtymävaihe on mahdollisimman sujuva?

#### Uusi replikointimenetelmä

12. Oletteko päässeet jo ottamaan uuden replikointimenetelmän käyttöön?

12.1. Minkälaisia vaikeuksia kohtasitte uuden replikointimenetelmän käyttöönotossa? Mikä sujui hyvin?

12.2. Kuinka aiotte seurata ja ylläpitää replikointijärjestelmää käyttöönoton jälkeen?

13. Miten päädyitte juuri tähän menetelmään?

13.1. Mitä mahdollisuuksia tämä menetelmä avaa nykyiseen verrattuna?

13.2. Mitkä ovat uuden replikointimenetelmän tärkeimmät hyödyt?

13.3. Mitkä ovat keskeiset mittarit, joilla arvioitte replikointimenetelmän päivityksen onnistumisen?

14. Mitä muita vaihtoehtoisia ratkaisuja harkittiin ennen kuin päätettiin siirtyä juuri Progress Open Edge Pro2:een?

15. Miten uusi menetelmä vaikuttaa päivittäiseen toimintaan, sekä päivitetty replikointiprosessi työnkulkuun?

15.1. Miten replikointiprosessin rakenne on muuttunut vanhaan menetelmään verrattuna?

16. Kuinka useasti replikointi tapahtuu?

16.1. Kuinka paljon tietoa replikoituu kerralla?

16.2. Kuinka nopeasti replikointi tapahtuu (millisekunneissa)?

17. Muuttuuko replikoitava tieto vanhaan replikointimenetelmään verrattuna?

- 17.1. Kuinka paljon historiallista tietoa tulette säilyttämään tietovarastoympäristössä jatkossa?
- 17.2. Miten uusi replikointimenetelmä sopii säilöttävän tiedon määrän kasvaessa?
18. Uskotteko, että menetelmää jouduttaisiin päivittämään tulevaisuudessa?