

# DATA-ANALYTIKKAJÄRJESTELMÄN VAIHTOPROSESSIN SUUNNITTELU

Googlega Snowplow'iin

Marttala Malla

Opinnäytetyö

Tietojenkäsittelyn koulutus  
Tradenomi (AMK)

2023

Tietojenkäsittelyn koulutus  
Tradenomi (AMK)

---

<b>Tekijä</b>	Malla Marttala	<b>Vuosi</b>	2023
<b>Ohjaaja(t)</b>	Pekka Reijonen		
<b>Toimeksiantaja</b>	Visma Software Oy		
<b>Työn nimi</b>	Data-analytiikkajärjestelmän vaihtoprosessin suunnittelu: Googlesta Snowplow'iin		
<b>Sivumäärä</b>	42		

---

Opinnäytetyön toimeksiantaja halusi eroon nykyisestä data-analytiikkajärjestelmästä ja implementoida toiminnanohjausjärjestelmäänsä Snowplow Analyticsin avoimen lähdekoodin data-analytiikka-alustan. Opinnäytetyön tavoitteena oli suunnitella ja toteuttaa data-analytiikkajärjestelmän vaihto niin, että alkuperäisiin mittauspisteisiin ei tarvitsisi koskea lähdekoodin osalta. Tutkimuskysymyksiä aiheita olivat vanhasta analytiikkajärjestelmästä poistuminen GDPR:n vuoksi, uuden analytiikkajärjestelmän implementoinnin suunnittelu ja toteutus koskematta lähdekoodiin, GDPR:n noudattamisen varmistaminen ja mahdolliset uudet toiminnallisuudet, kuten NPS-mittaus.

Opinnäytetyön tietoperustassa käytiin läpi data-analytiikkaa sekä Snowplow'n avoimen lähdekoodin toimintaa ja dokumentaatiota .NET Tracker -kirjaston osalta. Koska aihe johtuu osittain GDPR-asetuksesta, myös yleistä tietosuojasetusta ja sen vaatimia asioita tutkittiin. Lisäksi tutustuttiin ERP-järjestelmiin ja työn kohteeseen. Tässä tutkimuksellisessa kehittämistyössä käytettiin metodologisena lähestymistapana konstruktivistista tutkimusotetta. Työn aikana tehtiin jatkuvaa dokumentaatiota muistiinpanojen ja päiväkirjan muodossa sekä haastatteluita, keskusteluista että varsinaisen työn toteuttamisesta, jotka toimivat työn ai-neistona.

Snowplow'n implementointi onnistui tavoitteen mukaisesti, eli analytiikka-alusta saatiin vaihdettua koskematta lähdekoodiin. Opinnäytetyönä tehdyn suunnitelman mukaan ei täysin edetty, koska työn edetessä jouduttiin ratkaisemaan muutamia yllättäväkin ongelma. GDPR:n mukaan toimiminen varmistettiin, mutta uusien toiminnallisuuksien lisääminen jäi tulevaisuuteen aikatauluongelmien vuoksi. Vanha analytiikkajärjestelmä tullaan poistamaan käytöstä, kun uudesta on saatu tarpeeksi dataa. Aiemman datan siirtäminen uuteen tietovarastoon ei olisi ollut ajankäytöllisesti järkevää.

Avainsanat

Data, analytiikka, Snowplow, toiminnanohjausjärjestelmät, GDPR

Business Information Technology  
Bachelor of Business Administration

---

<b>Author</b>	Malla Marttala	<b>Year</b>	2023
<b>Supervisor(s)</b>	Pekka Reijonen		
<b>Commissioned by</b>	Visma Software Oy		
<b>Title</b>	Planning the process of replacing data analytics system: from Google to Snowplow		
<b>Number of pages</b>	42		

---

The thesis commissioner wanted to discard their existing data analysis system and implement an open-source platform of Snowplow Analytics to their Enterprise Resource Planning (ERP) system. The goal of the thesis was to plan and implement the switch between data analysis systems so that the measuring points of the source code wouldn't have to be touched. The topics of the research questions concerned the decommissioning the old data analysis systems due to the GDPR, the planning and implementation of the new systems without touching the source code, ensuring compliance with the GDPR and possible new functionalities such as NPS-calculation.

The knowledge base of the thesis covered data-analytics as well as the functionality and documentation of Snowplow's open-source code especially for the .NET Tracker library. As the subject is tied to the GDPR, additional study was done on the data protection regulation and its requirements. Additionally, the ERP systems and the goal of the project were explored. A constructive research approach was used as the methodology in this research development project. During the course of work, continuous documentation was made in the form of notes, work diaries, as well as interviews and discussions on the execution of the work itself, which served as data for the project.

The implementation of Snowplow was successful, and the analytics systems was replaced without touching the source code. The plan laid out in the thesis was not fully realized, though, as a few unexpected issues had to be solved during the course of work. Compliance with the GDPR was achieved, but the addition of new features was left out due to time constraints. The old data analysis systems will be decommissioned once enough data is gathered from the new one. Transferring prior data to a new database would not have been reasonable in terms of time performance.

**Keywords** Data, analysis, enterprise resource planning systems, GDPR

## SISÄLLYS

1	JOHDANTO .....	5
1.1	Toimeksianto .....	5
1.2	Opinnäytetyön aihe .....	7
2	TIETOPERUSTA .....	9
2.1	Visma Nova ja ERP .....	9
2.2	GDPR .....	10
2.3	Data-analytiikka .....	10
2.4	Snowplow .....	12
2.5	Snowplow'n datan hallinta .....	14
2.6	BigQuery ja Tableau .....	16
3	MENETELMÄLLINEN TOTEUTUS.....	18
3.1	Tarkoitus, tavoitteet ja tutkimuskysymykset .....	18
3.2	Toimintaympäristö ja työn tarkoitus .....	18
3.3	Tutkimuksellinen kehittämistyö .....	19
3.4	Konstruktiivinen tutkimus .....	21
3.5	Kehittämisprosessin vaiheet .....	22
4	TYÖN SUUNNITELMA JA TOTEUTUS.....	26
4.1	Visma Novan asiantuntijoiden haastattelut .....	26
4.2	Data-analytiikkajärjestelmän vaihtamisen suunnitelma.....	28
4.2.1	Tiimi.....	28
4.2.2	Kirjastot .....	28
4.2.3	Snowplow'n tutkinta.....	29
4.2.4	Trackerin rakennus.....	30
4.2.5	Snowplow'n toiminta.....	31
4.2.6	Tietovarasto .....	32
4.2.7	NPS.....	32
4.2.8	Käyttöliittymä .....	33
4.3	Toteutus, testit ja ongelmat.....	33
5	POHDINTA .....	36
	LÄHTEET.....	40

# 1 JOHDANTO

## 1.1 Toimeksianto

Data-analytiikka on välttämätön osa nykyajan tietojenkäsittelyalaa. Erityisesti liiketoimintaan liittyvät päätökset ovat riippuvaisia ostajien ja asiakkaiden käyttäytymisen tiedostamisesta. Data-analytiikan avulla saatavaa tietoa käytetään, kun halutaan tehdä liiketoiminnallisesti viisaita päätöksiä organisaatiossa. Data-analytiikan yksi tärkeimmistä tehtävistä onkin auttaa liiketoimintaa menestymään. Data-analytiikka on parhaimmillaan kerätessään arvokasta tietoa käyttäjistä ja asiakkaiden käyttäytymisestä tuotteen tai palvelun parissa. (Stedman 2020.) Kumar (2020) mainitsee blogissaan viisi tärkeää syytä käyttää data-analytiikkaa liiketoiminnassa: tieto auttaa organisaatiota tekemään optimaalisia ratkaisuja ja sen avulla voidaan maksimoida työn tehokkuus. Data-analyysi antaa ajantasaista tietoa käyttäjien ja asiakkaiden toiminnasta sekä muutoksista heidän käyttäytymisestään, ja näin ollen sen avulla on mahdollisuus personoida ratkaisuja ja palveluja. Lisäksi tiedon avulla voidaan lisätä tuotteen tai palvelun laatua. Kaikki tämä auttaa saavuttamaan parempia tuloksia liiketoiminnassa; tiedon avulla voidaan analysoida ostajien ja käyttäjien mennyttä käyttäytymistä, huomata siinä muutoksia ja näiden avulla optimoida tulevia prosesseja.

Opinnäytetyön aiheena on data-analytiikkajärjestelmän vaihtoprosessin suunnittelu. Toimeksiantajana on Visma Software Oy, ja opinnäytteen kohde on ERP-järjestelmänä (Enterprise Resource Planning) eli toiminnanohjausjärjestelmänä toimiva Visma Nova. Visma Novan data-analytiikkajärjestelmänä tällä hetkellä toimivan Google Analyticsin (GA) vaihtaminen toiseen johtuu useasta syystä, mutta pääsyyinä on GA:n datan säilyttäminen EU:n ulkopuolella. Euroopan yleisen tietosuoja-asetuksen eli GDPR:n vastaisesti toimiva Google Analytics voisi olla mahdollinen haitta kansainvälisen yhtiön maineelle kyseisen ongelman vuoksi (Pyyny 2022). Danish Data Protection Agency vanhempi juridinen neuvonantaja, Senior Legal Advisor, on antanut lausunnon koskien GDPR:ää ja Google Analyticsia: ”GDPR on valmisteltu Euroopan kansalaisten suojelemiseksi, ja tämänhetkisillä Google Analyticsin mahdollisuuksilla data-analytiikkajärjestelmää ei voi muokata sellaiseksi, ettei tietoja voisi päästä väärin käsiin”. Näin ollen GA:ta ei lainmukaisesti voi käyttää EU:n alueella (Datatilsynet 2022.)

Usea taho on tutkinut Google Analyticsin käyttöä GDPR:n kannalta katsottuna, ja se on verkkosivujen analytiikkajärjestelmänä todettu tietosuojalainsäädännön vastaiseksi (Tietosuojavaltuutetun toimisto 2022). Vaikka Visma Nova ei kerää sensitiivistä tietoa, on parempi siirtyä pois kyseisestä analytiikkajärjestelmästä korvaamalla se toisella. Muita syitä vaihtoprosessille on, että Visma Novan käyttämä Google Analyticsin versio lopettaa tukensa 1.7.2023, joten vaihto toiseen olisi joka tapauksessa edessä.

Opinnäytteen aihe on tällä hetkellä ajankohtainen Google Analyticsin GDPR-ongelmien vuoksi. Aiheesta löytyy huomattava määrä tietoa aina uutisista blogikirjoituksiin ja lakipykäliin saakka. Hokkanen (2022) on tehnyt aiheesta sivuten opinnäytetyön, jossa hän vaihtaa Snowplow Analyticsin analytiikka-alustaksi SaaS-palveluun. Vaikka Hokkasen työssään tekemä analytiikka-alustan vaihtoprosessi ei ollut lähtöisin GDPR-ongelmista, totesi hän työssään, että yleinen tietosuojasetus ja sen tuomat rajoitteet on otettava huomioon analytiikka-alustaa valitessa. Data-analytiikasta ja GDPR:stä löytyy runsaasti sähköistä materiaalia tutkittavaksi. Tietojenkäsittelyn saralla maailma ja validi tieto muuttuu nopeasti, joten kovin vanhat lähteet eivät kyseiseen aiheeseen enää päde. Avainsanoja, jolla teoreettista pohjaa on lähdetty hakemaan, ovat ERP, data-analytiikka, GDPR sekä Snowplow Analytics.

Toimeksiantaja Visma Software Oy tarvitsee tuotteelleen Visma Novalle uuden data-analytiikkajärjestelmän. Visma Nova on noin 30 vuotta vanha ERP-järjestelmä. Kyseinen toiminnanohjausjärjestelmä on hyvin laaja ja monipuolinen pitää sisällään useita sovelluksia. Visma Novaa käytetään edelleen laajasti ja se on ylläpidon lisäksi uusiutumassa; viime vuosina ERP-järjestelmää on ryhdytty kääntämään modernimmalle koodikielelle sovellus kerrallaan. Opinnäytteessä Visma Novaan implementoitu Google Analytics 360 vaihdetaan toiseen analytiikkajärjestelmään, Snowplow Analyticsiin. Snowplow on avoimen lähdekoodin analytiikka-alusta, ja se on valikoitunut korvaamaan GA:n Visma-konsernin toimesta. (Snowplow 2022.) Visma Software haluaa varmistaa, että Snowplow pystyy toimimaan Visma Novan kanssa vähintään samalla tasolla kuin tämänhetkinen järjestelmä. Harkinnassa on myös tuotteen NPS-arvon (Net Promoter Score eli suosittelemisindeksi) mittauksen lisäämistä Visma Novaan Snowplow'n avulla. Snowplow'lla ja Google Analyticsilla on huomattavia eroja keskenään. (Key2Market

2016.) Työssä tutkitaan Snowplow'n laajoja mahdollisuuksia ja toimintoja sekä etsitään niistä sopivimmat ratkaisut laajaan ERP-järjestelmään.

Visma on kansainvälinen konserni, ja suurin osa Visman yhtiöistä on siirtynyt käyttämään Snowplow Analyticsia konsernin vahvasta toiveesta. Snowplow-analytiikka-alustalla käyttäjä omistaa oman datansa. Näin ollen on käyttäjän päätös, missä kerättyä tietoa säilytetään. Jo pintapuolisena vilkaisuna Snowplow'n dokumentaatioon ja verkkosivuun on nähtävissä, että sillä on hyvin laajat mahdollisuudet ja se on monipuolisemmin muokattavissa kuin Google Analytics. Snowplow Analytics on huomattavasti työläämpi ottaa käyttöön kuin muutamalla klikkauksella toimiva Google Analytics, mutta se tarjoaa muita etuja, joista tärkeimpiä ovat kerätyn datan omistusoikeus sekä pääsy raakadataan. (Igloo Analytics 2019a; Key2Market 2016.)

## 1.2 Opinnäytetyön aihe

Tämä opinnäyte on tutkimuksellinen kehittämistyö, jossa suunnitellaan data-analytiikkajärjestelmän vaihtoprosessi Visma Novaan. Työssä tutkitaan Snowplow Analyticsin toimintaa ja mahdollisuuksia peilaten Visma Novan tarpeisiin. Visma Novan asiantuntijoita kuullaan koskien kyseisen toiminnanohjausjärjestelmän tarpeita data-analytiikan osalta. Työn aikana tutkitaan myös Visma Novan lähdekoodia Google Analyticsin osalta ja otetaan selvää mitä tietoa nykyisellä järjestelmällä tallennetaan ja millaista dataa ylipäänsä Visma Novassa seurataan tällä hetkellä. Tämän jälkeen rakennetaan tietojen pohjalta suunnitelma analytiikkajärjestelmän vaihtamiseen; pyritään mahdollistamaan optimaalinen Snowplow'n käyttö data-analytiikka-alustana Visma Nova toiminnanohjausjärjestelmässä huomioiden järjestelmän tarpeet. Lisäksi toteutetaan varsinainen Google Analyticsin poistaminen ja korvaaminen Snowplow'lla, kun tutkimus ja suunnitelma on tehty. Tämän opinnäytteen tavoiteltava tulos on siis suunnitelma Google Analyticsin vaihtamisesta Snowplow Analyticsiin niin, että Snowplow'n avulla saadaan vähintään sama data kuin GA:lla on saatu. Suunnitelman lisäksi myös toteutus raportoidaan.

Opinnäytetyö aloitetaan Visma Novan tarpeiden kartoittamisesta data-analytiikan osalta, Google Analyticsin nykyisen käytön tutkimisesta ja Snowplow'n mahdollisuuksien ja toiminnan kartoittamisesta. Työssä tutkitaan Snowplow'n dataputkea (data pipeline) ja hahmotellaan, miten Snowplow ylipäätään toimii. Tämän jälkeen tehdään Visma Novan tarpeiden pohjalta sekä Snowplow'n tutkimisen perusteella suunnitelma vaihtoprosessista. Suunnitelman perusteella tehdään keikiluja ja testejä kehitysympäristössä mahdollisten ongelmien löytämiseksi ja korjaamiseksi. Lopuksi toteutetaan data-analytiikkajärjestelmän vaihtaminen sekä arvioidaan kehitysprosessin ja implementoinnin onnistuminen.



## 2 TIETOPERUSTA

### 2.1 Visma Nova ja ERP

Visma Nova on toiminnanohjausjärjestelmä eli ERP (Enterprise Resource Planning). Toiminnanohjausjärjestelmä voi mahdollisesti huolehtia isonkin organisaation tarpeista erilaisten sovellusten osalta. ERP ohjaa liiketoimintaprosesseja aina tilauksesta toimitukseen saakka. Toiminnanohjausjärjestelmällä voi tehdä niin ostot kuin myynnin, maksaa palkat, hoitaa kirjanpidon ja hallinnoida varastoa. Visma Novassa on monta erillistä sovellusta, joita voi ostaa joko kokonaisuena pakettina, osan sovelluksista tai vaikka vain yhdenkin sovelluksen riippuen täysin yrityksen tarpeista. Näin ollen Visma Nova voi sopia aina tilitoimistoista isoihin tukkuihin. Yrityksen pääasiallinen kohderyhmä on kuitenkin keskisuuret tukkukaupat. Visma Nova sisältää seuraavat sovellukset: myyntitilaus, ostotilaus, laskutus, myyntireskontra, ostoreskontra, materiaalihallinto, varastonhallinta, EDI-yhteydet, raportointi, tuotannonohjaus sekä kirjanpito. Eri asiakkaat ostavat järjestelmään erilaisia lisenssejä, jotka määrittävät, mitkä sovellukset ja toiminnot ovat yrityksellä käytössä. (Visma 2022.)

ERP-järjestelmä integroi yhteen monia liiketoiminnan prosesseja ja jakaa tarpeellista dataa eri osioiden välillä. Useilla suurilla yrityksillä, esimerkiksi tukkukaupat, onkin käytössä koko Visma Nova, sillä he tarvitsevat kaikkia aiemmin mainittuja sovelluksia toimiakseen optimaalisesti. Visma Novalla on kuitenkin asiakkaina pienempiäkin yrityksiä; esimerkiksi tilitoimistoja, jotka tarvitsevat vain kirjanpito-sovelluksen. ERP-järjestelmän tarkoitus on vähentää datan määrää. Kun kaikki data tulee yhdestä tietokannasta joka tarpeeseen, ei tarvitse päivittää samaa tietoa useaan lähteeseen. Näin myös jokaisella osastolla ja henkilöllä yrityksen sisällä on sama tieto reaaliajassa. ERP jakaa tietoa yrityksen sisällä, joten eri osastot jopa eri puolilta maailmaa pystyvät kommunikoimaan keskenään paremmin. Toimivan ja laajan ERP-järjestelmän kanssa organisaatio ei välttämättä tarvitse muuta järjestelmää tai sovellusta. Yhden järjestelmän kanssa on aina helpompi tallentaa ja kerätä dataa, kun tietoa ei tarvitse tallentaa duplikaattina minnekään. Tämä vähentää myös mahdollisten virheiden määrää. (The Investopedia Team 2022; Oracle 2022.)

## 2.2 GDPR

General Data Protection Regulation eli yleinen tietosuojasetus on Euroopan unionin asettama henkilötietojen käsittelyä säätelevä laki, joka tuli voimaan toukokuussa 2018. Laki määrittelee rajat henkilökohtaisen datan keräämiseen ja prosessointiin yksilöiltä. GDPR:n tarkoitus on antaa henkilölle itselleen enemmän kontrollia päättää omien henkilökohtaisten tietojensa käsittelystä ja levittämisestä. (Frankenfield 2020.) Vaikka GDPR on Euroopan unionin laki, se koskee myös maanosasta riippumatta kaikkia organisaatioita, jotka keräävät dataa EU:n alueella asuvista ihmisistä. Se koskee myös kaikkia organisaatioita, jotka käsittelevät tietoaan muualla kuin EU:n alueella. (Wolford 2022.)

GDPR:n alaisiin suojeltaviin henkilökohtaisiin asioihin kuuluu muun muassa henkilön nimi, sosiaaliturvatunnus tai muu henkilötunnus, sijaintitiedot, etniset tiedot, kuvat, internet-käyttäytyminen (evästetiedot), terveys ja terveystiedot, poliittiset kannat, uskonnolliset asiat sekä liittojen jäsenyydet. Myös muut tunnistettavuustiedot, kuten sormenjäljet tai kasvopiirteiden analysointi ovat GDPR:n alla suojeltavaa tietoa, eli kaikki tieto ”luonnollisen henkilön fyysisestä, fysiologisesta, geneettisestä, henkisestä, taloudellisesta, kulttuurisosiaalisesta identiteetistä”. (Castagna 2021.) Jos kyseessä on oikeushenkilö tai henkilö on kuollut, tietosuojasetusta ei tarvitse soveltaa. Tietosuojasetus ei myöskään koske lainvalvontaa tai kansallista turvallisuutta. (Your Europe 2022; Official Journal of the European Union 2016.)

## 2.3 Data-analytiikka

Hyvin yksinkertaistettuna data-analytiikka on olemassa olevan tiedon analysoimista niin, että siitä voidaan tehdä erilaisia päätelmiä. Yleensä datan analysointi on automatisoitu, ja se suoritetaan erilaisilla prosesseilla ja algoritmeilla. Data-analytiikkaa on monenlaista. Kuvaileva analytiikka yrittää hahmottaa, mitä tapahtui. Onko esimerkiksi joku tietty numero, vaikkapa myynti, noussut ylös. Diagnostinen analytiikka miettii, miksi jotain tapahtui, eli vaikuttiko esimerkiksi kampanjatarjous myynnin nousuun. Ennustava analytiikka taas pohtii mitä tulee tapahtumaan, eli katsotaan ja analysoidaan menneitä dataa; mitä tapahtui myynnille viimeksi, kun oli kampanjatarjous. Ohjeellinen analytiikka yhdistää nämä kaikki ja

päättelee mitä pitäisi tehdä seuraavaksi. Esimerkissä edellisen kampanjatarjouksen aikaan myynti nousi, joten seuraavalla kertaa kampanjatarjouksen aikaan lisätään myyjien määrää, lisätään kampanjatarjouksessa olevan tuotteen valmistusta sekä jatketaan aukioloaikaa iltaan saakka. (Frankenfield 2022; Santos & Costa 2020.)

Liiketoiminnassa syntyy suuri määrä uutta dataa joka päivä. Tietoa on monenlaista ja merkitykseltään erilaista; asiakkaan palautetta, tilauksia, klikkauksia, sovelluksen aukaisukertoja, raportin tulostamisia, ynnä muita tapahtumia, jotka ovat relevantteja liiketoiminnalle. (Frankenfield 2022; Santos & Costa 2020.) Ahuja, Asthana, Ahuja, ja Agarwal (2020) mainitsevat data-analytiikan tärkeimpiä tavoitteita olevan liiketoiminnan strategian löytäminen, uusien tuotteiden suosion ennustaminen markkina-analyysillä, mainoskampanjoiden suunnittelu, asiakkaiden uskollisuuden ennustaminen, mahdollisten tulevien ongelmien syiden löytäminen, asiakkaiden yhdessä käyttämien tuotteiden kategoriointi ja luokittelu sekä sijoitetun pääoman tuoton ja riskien ennustaminen. Kaikki nämä tähtäävät riskien vähentämiseen ja liiketoiminnan tuoton kannattavuuteen. (Santos & Costa 2020.)

Organisaatiossa eri rooleissa työskentelevät ihmiset tarvitsevat erilaista tietoa työnsä optimoimiseen. Kehittäjät tarvitsevat tietoa kehittääkseen uusia käytänteitä ja toiminnallisuuksia. Analyttikot ja suunnittelijat tarvitsevat dataa kokeillakseen erilaisia ratkaisuja ja luodakseen analyysoivia sovelluksia. Liiketoiminnan päättäjät tarvitsevat dataa saadakseen näkemystä päätöksien kannattavuudesta. (Ye 2022.)

Tietojen analysointi perustuu useisiin päätöksiin. On tärkeää yrittää kerätä massan seasta kyseiselle liiketoiminnalle validia tietoa. Mikäli data-analytiikalla kerätään kaikki mahdollinen tieto, datan määrä voi olla valtava. Tällöin datamassan siivousvaiheessa on hankalaa yrittää löytää tiedon seasta se optimaalinen tieto, jolla on merkitystä tuotteen tai palvelun parantamiseen ja päätöksentekoon. Tästä syystä jo ennen analytiikkajärjestelmän integrointia ohjelmaan täytyy analysoida mitä tietoa halutaan ja tarvitaan sekä milloin sitä tietoa halutaan ja tarvitaan. Data-analytiikkaa käytetään siis prosessien optimointiin yleisen tehokkuuden lisäämiseksi. (Frankenfield 2022.)

Frankenfield (2022) on jakanut data-analyysin neljään askeleeseen. Ensimmäinen askel on määrittellä kriteerit kerättävälle datalle ja kuinka data halutaan luokitella. Toinen askel on datan kerääminen. Kolmas askel on datan järjestely analyysia varten; järjestellään data käyttötarkoituksen mukaan riippuen siitä, mitä tietoa datasta halutaan lopullisessa muodossa. Viimeinen eli neljäs askel on datan siivoaminen ennen analysointia. On varmistettava, ettei datassa ole virheitä tai duplikaatteja eikä prosessi ole jäänyt kesken. Tämän jälkeen data on valmis analysoitavaksi.

Data-analytiikka on kehittynyt paljon viimeisten vuosikymmenien aikana. 1990-luvun puolivälissä data-analytiikka tarvitsi vielä fyysistä laitteistoa, ja sen käyttö oli hankalaa ja kallista. 2000-luvun taitteessa syntyivät ensimmäiset SaaS-analytiikkayritykset, jotka tarjosivat data-analyysia palveluna. Se oli edelleen hintavaa, ja haittapuolena oli oman raakadatan menetys; palvelun tuloksena saatiin vain valmiiksi analysoitua tietoa. Pilvilaskennasta tuli 2000-luvun edetessä edullisempää ja sen myötä saatavuuskin parani. Jopa pienemmillä yrityksillä oli varaa data-analytiikkaan. Dataa tarvitaan jatkuvasti tuotteiden ja palveluiden kehittämiseen koska maailma, ihmiset ja ihmisten käyttäytyminen muuttuu jatkuvasti. Ajan kuluessa myös tarpeet muuttuvat; se mikä houkutteli viisi vuotta sitten, ei houkuttele enää. Snowplow Analytics syntyi vuonna 2012, ja oli ensimmäinen yritys, joka tarjosi uutta teknologiaa ja pilvilaskentaa antaen pääsyn ja omistuksen omaan raakadataansa. (Igloo Analytics 2019b.)

## 2.4 Snowplow

Snowplow Analytics on avoimen lähdekoodin analytiikka-alusta. Sillä voi kerätä tietoa, jopa useista eri lähteistä. Kerättyä tietoa voi itse hallita ja analysoida haluamallaan tavalla. (Poddar 2021.) Koska Snowplow ei ole sidottu Googlen tai Adoben kaltaisiin datamalleihin, Snowplow’lla voi kerätä tietoa kaikkialta: nettisivuilta, mobiilisovelluksista, peleistä, ohjelmistoista ja jopa laitteistoista. (Snowplow 2022.)

Snowplow’ssa ei ole käytännössä minkäänlaisia ohjelman sisäisiä rajoituksia tallennettavan datan suhteen, joten sitä voi käyttää täysin oman tarpeensa mukaan. Snowplow’ta käytettäessä tieto kerätään ja tallennetaan reaaliajassa ja kaikki

saatava raakadata omistetaan itse. Snowplow onkin loistava vaihtoehto data-analytiikka-alustaksi, mikäli halutaan säilyttää kontrolli omaan raakadataansa. (Poddar 2021.)

Snowplow'ta käyttäessä hyötyjä on useita; yksi tärkeimmistä on täysi läpinäkyvyys missä data varastoidaan ja palvelut prosessoidaan. Snowplow on rakennettu avoimen standardin mukaan, jonka avulla voidaan välttää yhden toimittajan kanssa jumiin jääminen. Alusta myös mahdollistaa analytiikan liikkumisen lähemmäs ohjelmistokehitystä, koska sinne voidaan luoda omia tapahtumia ja skeemoja tiedon laadun varmistamiseksi. Tiedon keräämistä varten luotu oma verkotunnus varmistaa kerättyjen tietojen omistuksen. Snowplow'n avoimen lähdekoodin versiota käytettäessä myös kustannukset ovat olemattomat. Käytännössä ainoita "haittapuolia" ovat Snowplow'n implementoinnin haastavuus, sekä itse analysoinnissa graafisen käyttöliittymän puute. Nämä vaativat omia resursseja toimiakseen. (Ye 2022.)

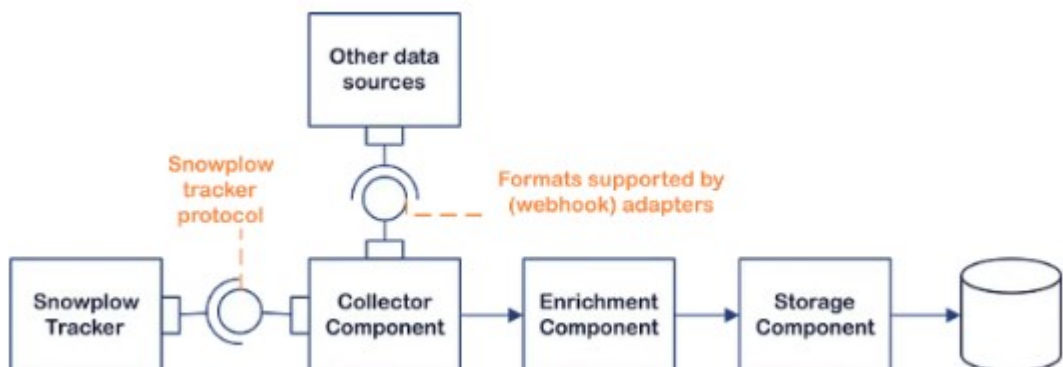
Snowplow voitti vuonna 2022 "Data Tech Deployment of the Year" -palkinnon maailman johtavana käyttäytymisen data-alustana (behavioral data platform). Snowplow'ta kiitetään erityisesti korkealaatuisesta käyttäytymisen datan luomisesta, hallitsemisesta ja mallintamisesta. Data sopii hyvin käytettäväksi muun muassa keinoälyratkaisuissa, koneoppimisessa ja sovelluksissa, jossa tarvitaan edistyksellisistä analytiikkatietoa. Integroituna muihin nykyaikaisiin työkaluihin saatava tieto on merkityksellistä edistyneemmällekin liiketoiminnalle, oikein käytettynä. Snowplow on uniikki alusta tiedonkeruuseen mahdollistaen täyden kontrollin ja omistuksen raakadatalle ilman, että ulkopuolinen taho määrittelee kuinka dataa tulisi kerätä, mallintaa tai käyttää. Snowplow'n edistyksellinen joustavuus ja vapaus erottaa sen muista data-analytiikkaratkaisuista. (Johnson 2022.)

Snowplow'ssa käytetään skeemoja ja tapahtumaskeemat voi määritellä itse. Snowplow aloitti verkkosivujen analyysialustana, mutta nykyään se tähtää jokaisen alustan Event-tapahtumien analytiikkajärjestelmäksi. Riippumatta siitä haluaako tietää sovelluksen avauksen on-premise-ohjelmistossa, verkkosivuilla sivulta toiselle liikkumisen tai oman sovelluksen kustomoidun tapahtuman tiedon, Snowplow pystyy sen keräämään. Snowplow hyväksyy tiedot kaikkialta ja se tarjoaa pääsyn kaikkeen kerättyyn raakadataan. (Merlevede 2020.)

Snowplow'ssa ei ole laisinkaan graafista käyttöliittymää. Mikäli käyttöliittymän haluaa, se täytyy itse rakentaa tai yhdistää se jo olemassa olevaan käyttöliittymään. Monet yhdistävät esimerkiksi Google Analyticsin ja Snowplow'n niin että he ohjaavat Snowplow'lla kerätyt tiedot Google Analyticsiin, joka näyttää ne GA:n tarjoamalla käyttöliittymällä. Snowplow kerää ja tallentaa ainoastaan raakadatan kuten vaikkapa tietyn sivun katselukerrat, tai klikkaukset tietystä linkistä. Snowplow ei itsessään osaa analysoida tietoja käyttäjävirroiksi tai tilastoiksi. Toisin sanoen Snowplow ei osaa näyttää montako kertaa tietyn ajan sisällä on käyty tietyllä sivulla, vaan Snowplow'n tietokantaan tallentamia Event-tapahtumia voidaan tarkastella yksittäisinä tietoina. Tietokannasta löytyy kyllä kaikki kerrat, kun sivulla on käyty, mutta irrallisina tapahtumina. Oikean haun avulla voidaan laskea, kuinka monta kyseistä tapahtumaa on tapahtunut tietyn ajan sisällä. Tietokannasta ei kuitenkaan voi suoraan nähdä, että tänä päivänä sivulla on käyty 1256 kertaa, vaan nähdään 1256 tapahtumaa, joissa kaikissa on käyty kyseisellä sivulla. Snowplow'lla on tarjota erilaisia SDK:ita (Software Development Kit), jotka auttavat analysoimaan saatua raakadataa. Snowplow'lla on kehitteillä myös niin kutsuttu "Web Data Model Project", joka osaa ryhmitellä klikkauksia selausistunnoittain. (Merlevede 2020.)

## 2.5 Snowplow'n datan hallinta

Snowplow'ta voi käyttää monella eri tavalla ja monesta eri kohteesta; se on open-source alusta eli avoimen lähdekoodin alusta, ja sen mahdollisuudet ovat valtaavan suuret. Snowplow'ssa data kuitenkin kulkee samankaltaista dataputkea pitkin jokaisessa versiossa (kuvio 1). (Snowplow 2022.)



Kuvio 1. Snowplow'n yksinkertaistettu dataputki (Merlevede 2020)

Snowplow Analyticsin toiminta lähtee liikkeelle niin kutsutuista Trackereista. Tracker on SDK (Software Development Kit), joita Snowplow tarjoaa usealle alustalle. Tracker-protokollassa Tracker muodostaa Event-tapahtumia, jotka ovat joko GET- tai POST-kutsuja Snowplow'n Collectorille. Trackerit ovat siis vastuussa raakadatan lähettämisestä Collectorille. Trackerit rakennetaan sen mukaan, mitä dataa halutaan kerätä: se on lista tarvittavista parametreista. Trackerit siis luovat datan ja lähettävät sen eteenpäin. (Igloo Analytics 2019B; Merlevede 2020; Poddar 2021; Snowplow 2022.)

Collector-komponentti on avoin perusverkkopalvelin, joka vastaanottaa http-pyyntöjä. Kun Event-tapahtuma tulee Collectoriin se tallentaa raakadatan, enkryptaa eli salaa sen ja lähettää tapahtuman edelleen dataputkessa eteenpäin Enrichiin. Mikäli http-pyyntö ei sisällä käyttäjän tunnistettavaa evästettä, Collector upottaa satunnaisesti luodun evästeen dataan mukaan. Tämä päättää onko saatu data hyvää vai huonoa ja pistää sen oikeaan linjaan joko hylättyjen tai hyväksytyjen tapahtumien putkeen. (Igloo Analytics 2019b; Merlevede 2020; Poddar 2021; Snowplow 2022.)

Enrichment-komponentti eli datan rikastuttaja, on varsinainen datan "tilaaja". Se puhdistaa datan, ja varmistaa, että se vastaa määritellyn Event-tapahtumaskeeman kriteerejä. Mikäli data ei täytä kriteerejä se siirtyy "epäonnistuneiden skeemojen" lokitietoihin. Näin myös virheellistä dataa pystyy tarkastamaan ja käsittelemään. Virheellinen data ei kuitenkaan näy tietokantahauissa kyseisen tapahtumaskeeman alla, vaan "epäonnistuneet" tapahtumaskeemat haetaan eri haulla "failed"-osastolta. Kriteerit täyttäviä tapahtumia Enrichment parantaa määrättyllä tavalla ja tallentaa ne valitulle alustalle. Enrichmentissä voi tehdä erilaisia itsemäärättyjä asioita, muun muassa anonymisoida eli salata käyttäjän ID tai muita tunnistetietoja, voi lukea ja lisätä tallennukseen evästetietoja, muuttaa ip-osoitteita tunnistamattomaksi ja niin edelleen. Myös tämä komponentti on itse määriteltävissä. (Igloo Analytics 2019B; Merlevede 2020; Poddar 2021; Snowplow 2022.)

Storagen eli tallennuspaikan voi Snowplow'ssa määritellä itse useista mahdollisista vaihtoehtoista. Suositeltuja vaihtoehtoja ovat esimerkiksi Redshift, Snowflake tai BigQuery. Storageen tallennuksen jälkeen loppuu Snowplow'n perustointi ja Snowplow'n tehtävä on valmis. Raakadata on tallennettu ja sitä voi hyödyntää haluamallaan tavalla: sitä voi esimerkiksi pilkkoa pienemmiksi osiksi, tallentaa ryhmissä, tarkastella niitä paloina, yhdistellä toiseen dataan tai striimata reaaliaikaisena. (Igloo Analytics 2019B; Merlevede 2020; Poddar 2021; Snowplow 2022.)

## 2.6 BigQuery ja Tableau

BigQuery on suuri enterprise-tason datan pilvivarastointiratkaisu. BigQueryn omistaa Google, ja se on PaaS-palvelu (Platform as a Service). BigQuerya voi käyttää osana Googlen omaa data-analyysiratkaisua, mutta se soveltuu erinomaisesti käytettäväksi Googlen analysointiratkaisusta irrallisena datavarastona. Ilman analysointiosuutta, datavarastona, BigQueryyn voidaan tehdä SQL-kyselyitä, jolla saadaan haluamaansa informaatiota ulos. (Google Cloud 2023; Lindström 2019.)

BigQuery esittää haetun datan taulukoina, riveinä ja kolumneina. BigQuerya kehitetään jatkuvasti, ja se on jo nyt todella tehokas SQL-kyselyjen vastaaja Googlen laskentatehon vuoksi. BigQuery voi vastata nopeasti massiivisiinkin kyselyihin; jopa teratavuja sekunneissa tai petatavuja minuuteissa. Datavarasto on kustannustehokas, sillä hinnoittelu on aidosti käyttöön perustuva; hinta koostuu laskennasta ja tallennustilasta. Varastoon voi tallentaa millaista dataa tahansa; jäsenneilyä, jäsenneilyä sekä hiukan jäsenneilyä dataa. (Google Cloud 2023; Lindström 2019.) Vaikka Google Cloudin muiden tuotteiden tuella BigQuery voisi toimia kokonaisuutena data-analyysijärjestelmänä, tässä työssä sitä käsitellään vain tietovarastona.

Tableau on visuaalisen analytiikan ohjelmisto. Tableau voisi myös olla kokonainen analytiikka-alusta BigQueryn tavoin, mutta sitä voi pilkkoa myös osiin. Tässä työssä käsitellään Tableautta ainoastaan työpöytäsovelluksena eli käyttöliittymänä, jonka on tarkoitus näyttää visuaalisesti Snowplow'lla kerätty data. Table-



auun liitetään datalähde, joka tässä tapauksessa tulee olemaan BigQuery, ja datan saa rakennettua näkyviin tahtomallaan visuaalisella tavalla. Tableaulla on jo valmiita vaihtoehtoja datan näyttämiseen, mutta sillä pystyy rakentamaan myös itselleen sopivan käyttöliittymän juuri yrityksen tarpeisiin räätälöitynä. (Tableau, 2023.)

### 3 MENETELMÄLLINEN TOTEUTUS

#### 3.1 Tarkoitus, tavoitteet ja tutkimuskysymykset

Opinnäytteen tarkoitus on suunnitella data-analytiikkajärjestelmän vaihdon toteutus Visma Novaan niin, että tietokantaan saatava data pysyy samana. Lisäksi tutkitaan mahdollisuutta mitata tuotteen NPS-arvoa Snowplow analytiikka-alustan avulla. Tämän työn päätavoitteena on päästä eroon Google Analyticsista GDPR:n vuoksi sekä siirtää Visma Novan tarvitsema data onnistuneesti Snowplow Analyticsin avulla BigQueryyn, mittauspisteiden ja datan pysyen samankaltaisena kuin GA:n kanssa. Sivutavoitteena on uuden analytiikkajärjestelmän implementointi Visma Novaan keskitetysti eli koskematta sen laajaan lähdekoodiin sovelluksen alla. Pyrkimyksenä on mahdollisimman hienovarainen siirtyminen uuteen analytiikkajärjestelmään, koska Visma Novan jo olemassa olevaan lähdekoodiin ei haluta enää ”turhaan” koskea, sillä järjestelmää uudistetaan koko ajan. Pyritään siis käyttämään jo olemassa olevia Google Analyticsin kanssa käytettyjä funktioita Snowplow’n kanssa, mikäli se todetaan työn edetessä mahdolliseksi.

Opinnäytteessä pääkysymyksenä on, miten data-analytiikkajärjestelmän vaihtoprosessi tulisi toteuttaa. Alatutkimuskysymyksiä ovat:

- Onnistuuko data-analytiikkajärjestelmän vaihto niin, että rajapinta muualle Visma Novan lähdekoodissa ei muutu?
- Miten jo olemassa olevaa dataa tulisi käsitellä?
- Miten varmistetaan uuden järjestelmän GDPR yhteensopivuus?
- Saisiko Snowplow’n avulla mitattua tuotteen NPS-arvoa eli suositteluindeksiä Visma Novaa hyödyttävällä tavalla?

#### 3.2 Toimintaympäristö ja työn tarkoitus

Opinnäytteen toimintaympäristö on ERP-järjestelmä Visma Nova. Työssä suunnitellaan data-analytiikkajärjestelmän vaihtoprosessi Visma Novaan, sekä toteutetaan ja raportoidaan vaihtotapahtuma. Visma Nova on laaja ohjelmisto, se sisältää noin 1,2 miljoonaa riviä koodia (Viita, 2022). Toimintaympäristö on siis

laaja, ja vaatii paljon lähdekoodin lukemista ja tutkimista; on etsittävä missä ja mitä tietoa Google Analytics tällä hetkellä kerää. Nämä tiedot ovat tärkeä paikallistaa, jotta päästään eroon GA:sta täysin, sekä saadaan varmistettua, että uudella järjestelmällä kerätään vähintään samat tiedot kuin aiemmin. Visma Novasta on olemassa pilvipalvelusovellus ja on-premise-ohjelmisto (asiakkaan omassa ympäristössä toimiva), ja dataa on kerättävä analytiikkaa varten molemmista. Kyseessä on kehittämistehtävä, joka toteutetaan tutkimuksellisena kehittämistyönä. Opinnäytteessä käytetään laadullista tutkimusmenetelmää, ja metodologisena lähestymistapana käytetään konstruktivistista tutkimusotetta.

### 3.3 Tutkimuksellinen kehittäminen

Koska tässä opinnäytteessä tehtävä suunnitelma toteutetaan myös käytännössä, on kyseessä tutkimuksellinen kehittäminen eli kehittämistutkimus. Kananen (2012, 43) mainitsee kehittämistutkimuksen vaarana olevan ”tutkimuksen” jäämisen kehittämistyön asteelle. Kehittämistutkimuksen tulokset ovat yksittäistapauksia, joten niistä ei voida tehdä laajoja yleistyksiä. Tuloksia voi olla mahdollista kuitenkin hyödyntää ottamalla huomioon, että juuri tämä tutkimus on pätevä vain täsmälleen samoissa olosuhteissa missä tutkimus on tehty, ja tutkimuksen hyödyntäminen vaatii muutoksia toimenpiteisiin (Kananen 2012, 41–43). Tutkimuksen kohteena on tässä yksi tietty järjestelmä, mutta opinnäyte aiheena on yleispätevä, sillä tällä hetkellä useat yritykset ohjelmistoihin ja verkkosivuihin ovat samassa tilanteessa Google Analyticsin kanssa.

Usein tutkimuksellisessa kehittämistyössä on tapaustutkimuksen piirteitä, niin myös tässä työssä. Tapaustutkimuksessa kerätään yleensä ensin tutkimusaineistoa, ja vasta sitten valitaan tutkimuksen kohde, joten täysin puhdas tapaustutkimus tämä ei ole. Tapaustutkimuksessa ei myöskään yleensä mennä toteutukseen saakka. Tapaustutkimuksessa ei pyritä yleistettävyyteen, vaan pyritään ymmärtämään ja tulkitsemaan yhtä tapausta. (Koppa 2015; Humak 2022; Yin 2018, 3–6.) Vaikka tässä työssä saatava tieto on osaltaan yleispätevää ja sitä voi hyödyntää muihin samankaltaisiin tapauksiin, sitä käsitellään tässä vain Visma Novan ja Snowplow Analyticsin yhdistämisenä optimaalisella tavalla.

Tapaustutkimuksessa kiinnitetään tutkittava tapaus aina kiinteästi omaan paikkaansa eli se on sidonnainen aikaan ja paikkaan. Myös siis tutkimuksen tulos on pätevä ainoastaan paikkaan ja aikaan sidottuna. Tämä vaikuttaa myös työskentelytapoihin. Kun tavanomaisessa haastattelututkimuksessa kysellään tietoja työntekijöiltä, tarvitaan pohjatietoa yrityksestä ja muusta taustasta, mutta tätä taustatietoa ei käytetä niinkään aineistona vaan yleistietona. Tapaustutkimuksessa sen sijaan näitä pohjatietoja voi käyttää myös aineistona, koska itse tutkimuskohde on sidottu kyseiseen paikkaan ja aikaan. (Vuori 2021.)

Tutkimus alkaa aina tutkittavaan ilmiöön perehtymisellä, niin myös kehittämistutkimuksessa. Perehdytään siihen mitä ilmiöstä jo tiedetään. Rakennetaan teoreettista viitekehystä kirjallisuuskatsauksen avulla: yhä etenevässä määrin myös sähköisistä tiedonlähteistä. Aiheeseen perehtyminen voi antaa kirjoittajalle osaan ongelmista vastauksia tai ainakin viitteitä niistä. Tutkimuksella pitäisi aina olla uutuusarvoa, hyötyä tai merkitystä alalla. Tutkija joutuu suunnitteluvaiheessa tekemään tärkeitä päätöksiä ongelmien ja kehityskohteen rajauksesta sekä käytettävistä työkaluista. Suunnitteluvaiheessa on tärkeää itse kehittämistehtävän lisäksi suunnitella myös opinnäytteen kirjoittamista, koska kun kehittämistyö on saatettu loppuun, myös ilmiö on hävinnyt. Tästä syystä dokumentointi on tärkeää. Dokumentointi on suunniteltava tarkkaan, jotta tarvittavia tietoja saadaan kehitystyön jälkeenkin tarkasteltavaksi tarpeen mukaan. (Kananen 2012, 45–48.)

Kenttävaiheessa aloitetaan kehitystyö ja dokumentaation tekeminen. Kananen (2012, 48) mainitsee helpoimman dokumentaatiokeinoon olevan tutkimuspäiväkirjan, johon kirjataan tehdyt asiat. Koska työtä tehdessä ei välttämättä osaa arvioida mikä tieto on relevanttia, on tärkeää kirjoittaa mieluummin liian paljon kuin liian vähän muistiinpanoja. Kenttätyön tulos on opinnäytteen aineisto, jonka perusteella työ kirjoitetaan. Perinteisesti kehittämistyön kenttävaihe ja opinnäytteen kirjoitus erotellaan toisistaan ja tehdään peräkkäin, mutta mahdollisuuksien mukaan myös limittäinen tekeminen on mahdollista, kuten tässä opinnäytteessä on tehty. Limittäinen tekeminen mahdollistaa syvemmän analyysin. Laadullisessa tutkimuksessa aineiston kerääminen ja analysointi tapahtuu samaan aikaan, jolloin tutkijan ymmärrys aiheeseen kasvaa koko ajan. Näin tutkija voi myös huomata uusia analysoitavia asioita aineiston kasvaessa, tai tarkentaa ongelman myötä analysoitavaa osaa. (Kananen 2012, 48, 98.)

Kehittämistutkimuksessa tutkijalta vaaditaan osallistumisen ja prosessinhallinnan lisäksi tutkimusosaamista. Jotta kehittämistutkimus ei jää pelkäksi kehittämistoinnaksi on tulos voitava jotenkin mitata. Hankalemmassa tapauksessa on käytettävä erilaisia mittareita mittaamaan lopputulosta verrattuna entiseen. Helpommassa tuloksenasettelussa kuten tässä työssä, riittää se, että ratkaisu toimii. Kyseessä on siis tekninen parannus, jossa ongelmat eivät liity kohteen suoritukseen. Tavoiteltava lopputulos on siis tuottaa kaikkien hyväksymä, toimiva ratkaisu. (Kananen 2012, 48–50.)

### 3.4 Konstruktiivinen tutkimus

Tässä opinnäytteessä käytetään metodologisena lähestymistapana konstruktiivista tutkimusta. Yksinkertaistettuna konstruktiivisessa tutkimuksessa rakennetaan tuotos, joka ratkaisee ongelman: Visma Novassa on GDPR ongelma, johon opinnäyte tuo ratkaisun. Vaikka Snowplow Analytics on käytössä ympäri maailmaa, puhdasta kopiota ratkaisusta ei voi tehdä vaan ratkaisu määräytyy aina kohteen, tässä tapauksessa Visma Novan, mukaan. (Koppa 2015.) Lukka (2001) kertoo, että konstruktiivisen tutkimusotteen ydin on, että tutkimus keskittyy tosielämän ongelmiin, tuottaa konstruktion sekä yrityksen toteuttaa sen ja sisältää kokemuksellista oppimista. Lisäksi konstruktiivinen tutkimus sisältää tutkijan ja toimeksiantajan tiimimäistä yhteistyötä, on kytketty olemassa olevaan tietoon, sekä reflektoi löydöksiä teoriaan. Konstruktiivisessa tutkimuksessa luodaan uudenlainen ratkaisu tutkimustiedon sekä käytännöstä kerättävän tiedon pohjalta (Ojasalo, Moilanen & Ritalahti 2009, 65).

Konstruktiivisessa tutkimuksessa on järkevintä tutkijan liittyä jäseneksi työryhmään, tai jopa vetää sitä. Näin kaikki osapuolet saadaan osallistettua tavoitteiden saavuttamiseen. Lisäksi työryhmä voi antaa arvokasta osaamista aiheesta tutkijalle. (Lukka 2001.) Opinnäytetyössä koostettiin tiimi Snowplow-implemентаation toteuttamiseen, vaikka vetovastuu olikin opinnäytetyön tekijän vastuulla. Työryhmän tarkoitus oli nimenomaan jakaa omaa osaamistaan ja tietoaan. Tutkimussopimuksien tekemistä suositellaan tehtäväksi aina, sillä mikäli tuloksia ei päästä julkaisemaan todenmukaisesti, voi akateeminen tutkimuksen anti jäädä vaillinaiseksi (Lukka 2001).

Konstruktiiivisen tutkimuksen toteutusvaiheessa suunniteltua ratkaisua ei testata ainoastaan teoreettisesti toimivaksi, vaan konstruktiiivisessa tutkimuksessa saatu tulos otetaan organisaatiossa käyttöön. Tämä vaatii ratkaisun olevan testattu toteuttamiskelpoiseksi. Ratkaisun toimivuutta arvioidaan tiimissä sekä organisaatiossa, johon se on tehty. (Lukka 2001.) Tässä tutkimusprojektissa julkaistaan työn raportti, joka sisältää sekä suunnitelman että toteutuksen raportoinnin. Ainoastaan talon sisäinen artefakti, eli varsinainen koodi, on ei-julkaistavaa materiaalia. Analysointivaiheessa on tärkeää pystyä etääntymään työstä ja siihen liittyvistä tunteista, ja tarkastella tuloksia objektiivisesti (Lukka 2001).

### 3.5 Kehittämisen prosessin vaiheet

Kehittämistutkimukseen ja -prosessiin on olemassa useita koulukuntia, jotka määrittelevät kehittämissuorituksen vaiheet hieman eri malleihin perustuen (Kananen 2012, 53). Yhteistä kaikille on kuitenkin prosessin lineaarinen eteneminen, ja perusajatuksena kaikissa on ongelman määrittely, ratkaisun esitys, kokeilu ja arviointi. Käytännössä kuitenkin vain harvoin prosessi etenee aivan täysin lineaarisesti, vaan osia vaiheista voidaan työstää yhtä aikaa, tai tehdä osittain limittäin. (Kuvio 2) (Salonen, Eloranta, Hautala & Kinos, 2017, 51–53; Kananen 2012, 53.)



Kuvio 2. Kehittämissuorituksen lineaarinen eteneminen (Salonen ym. 2017)

Ojasalo, Moilanen ja Ritanen (2014) jakavat tutkimuksessaan kehittämissuorituksen kuuteen eri osa-alueeseen, joista ensimmäinen on kehittämissuorituksen tunnistaminen ja alustavien tavoitteiden määrittäminen. Opinnäytteessä kehittämissuorituksen kohde oli toiminnanohjausjärjestelmä Visma Novan on-premise- ja pilvipalvelusovellukset. Alustava tavoite tuli toimeksiantona; Google Analytics tulisi vaihtaa toiseen data-analytiikkajärjestelmään, tietosuojasetusten vuoksi mahdollisimman nopeasti.

Kehittämisen prosessin toinen osa-alue on kehittämiskohteeseen perehtyminen teoriassa ja käytännössä (Ojasalo ym. 2014). Kehittämiskohteeseen päästiin tutustumaan sekä teoriassa että käytännössä työskentelemällä Visma Novan parissa jo ennen opinnäytetyön aloitusta. Kehittämiskohde oli saatavilla koko ajan työtä tehdessä, joten jatkoperehtyminen onnistui myös opinnäytteen teon lomassa.

Ojasalo ym. (2014) kertoo kolmannen osa-alueen kehittämisen prosessissa olevan kehittämistehtävän määrittäminen ja kehittämiskohteen rajaaminen. Kehittämistehtävä oli vaihtaa Google Analytics Snowplow Analyticsiin Visma Nova toiminnanohjausjärjestelmään, sekä on-premise-ohjelmistoon että pilvipalveluun. Visma Novaa ollaan kääntämässä uudelle koodikielelle, ja vielä kääntämättömiin sovelluksiin ei haluta kuin pakolliset päivitykset. Tästä syystä toivottiin, että muutostyö voitaisiin tehdä koskematta ohjelman lähdekoodiin. Snowplow'lla haluttiin talteen kaikki sama data, mitä Google Analyticsilla saatiin. Opinnäytettä tehdessä harkittiin, ryhdytäänkö mittaamaan lisäksi esimerkiksi NPS-arvoa, mikä oli toimeksiantajan intresseissä mukana. Kehittämistehtävää määriteltiin tarkemmin tietoperustan avulla, toimeksiantajan ehdoilla ja hyväksynnällä. Tutkimuksellinen kehittämistyö tuottaa aina jotain, ja tässä tapauksessa se tuotti uuden järjestelmän data-analyysiin tarvittavan datan keräämiseen (Taulukko 1).

Taulukko 1. Tutkimuksen ja kehittämisen erot (Humak 2022)

	Tieteellinen tutkimus	Tutkimuksellinen kehittäminen	Arkinen kehittäminen
<b>Tavoite</b>	Uuden teoreettisen tiedon tuottaminen	Uusi tieto työelämän käytännöistä & todellisten ongelmien ratkaisu	Ratkaista käytännön ongelmia
<b>Tietoperusta</b>	Tutkimuskirjallisuus	Tutkimuskirjallisuus ja hiljainen tieto	Hiljainen tieto
<b>Kohde</b>	Tiedeyhteisö	Koko ammattiala ja tilaajaorganisaatio	Oma työ tai oma organisaatio
<b>Menetelmät</b>	Yleensä yksi tutkimusmenetelmä	Monien kehittämistyön menetelmien yhdistelmä	Sattumanvaraiset menetelmät tai "mutu"
<b>Vuorovaikutus</b>	Vähäistä, voi tehdä "kammiossa"	Kiinteä, edellyttää yleensä kenttätöitä	Riippuu tavoitteista
<b>Kriittisyys</b>	Systemaattisuus, kriittisyys	Systemaattisuus, kriittisyys	Sattumanvaraisuus
<b>Työn päättymisen</b>	Tutkimustulokset julkaistaan	Työn tulokset on jalkautettu käytäntöön ja uusi tieto ammattialalle	Ongelma on ratkaistu

Neljäs kehittämisen prosessin osa-alue on tietoperustan laatiminen sekä lähestymistavan ja menetelmien suunnittelu (Ojasalo ym. 2014). Tietoperusta koostuu teoriapohjasta sekä tutkimusaineistosta. Teoriapohjaa kerättiin muun muassa

Snowplow'n mahdollisuuksista, rajoitteista ja toiminnasta, dokumentaatiosta, data-analytiikasta ja GDPR:stä. Teoriapohjan tutkimus ei ensisijaisesti tähtää käytännön sovellukseen (Tilastokeskus 2015)

Tutkimusaineisto oli itsekerättävää aineistoa; haastattelut (mm. Visma Novan arkkitehti, Visman Snowplow asiantuntijat) sekä muistiinpanot havainnoinnista, joka tässä tapauksessa oli Visma Novan lähdekoodin tutkimista muun muassa Google Analyticsin osalta. Dokumentaatiota tehtiin muistiinpanojen muodossa työn etenemisestä, sekä esimerkiksi datan mittauspisteistä.

Vaikka haastattelut ja keskustelut eivät ole kovin kunnianhimoista kehittämistä, siitä lähdettiin liikkeelle; niillä kartutettiin välttämätöntä, hiljaista tietoa, esimerkiksi millaista dataa Visma Novasta pitäisi saada esille data-analytiikan avulla, sekä lisätietoa koskien Visma Novan tiimin toivetta NPS-arvon mittaamisesta. Lisäksi Visman Snowplow-asiantuntijalta saatiin tarkempaa tietoa Visma-konsernin käytänteistä koskien Snowplow-implemентаatiota. Tutkimusaineistoa kerättiin myös lähdekoodia tarkastelemalla, ja vertailemalla Google Analyticsin käyttöä haastattelussa esille tulleisiin tarpeisiin. Tämän jälkeen pääteltiin Visma Novan tarpeet, ja verrattiin niitä Snowplow'n mahdollisuuksiin. Yhdistelemällä tutkimusaineistoa teoriapohjaan voitiin alkaa tekemään yksityiskohtaista suunnitelmaa analytiikka-järjestelmän vaihdosta (Humak 2022). Tutkimuksellisessa kehittämistyössä menetelmä voi olla mikä tahansa systemaattinen keino päästä tavoitteisiin (Ojasalo ym. 2014). Vaikka opinnäytettä kirjoitettiin limittäin käytännön työn kanssa, koko työn ajan tehtiin päiväkirjamaista muistiinpanoa, jotta jälkikäteenkin olisi mahdollista tarkistaa tapahtumia tarpeen mukaan. Tämä opinnäytetyö on laadullinen, tutkimuksellinen kehittämistyö, ja metodistiseksi lähestymistavaksi on valittu konstruktiiivinen tutkimus, joka on yksi tapaustutkimuksen muotoja (Lukka 2001).

Kehittämisprosessin viides osa-alue on kehittämishankkeen toteuttaminen ja julkistaminen eri muodoissa (Ojasalo ym. 2014). Kun teoriapohja ja tutkimusaineisto oli analysoitu, ja oli tehty yksityiskohtainen suunnitelma vaihtoprosessista, päästiin toteutusvaiheeseen. Tässä vaiheessa korvattiin Google Analytics ERP-järjestelmässä Snowplow Analyticsilla. Tähän toteuttamisvaiheeseen kuului testauksia



järjestelmään, korjauksia sekä ongelmien ratkointaa, ja vasta hyväksytyjen testien jälkeen siirryttiin julkistamiseen. Toteutusvaiheesta raportoitiin opinnäytteen.

Viimeinen eli kuudes osa-alue kehittämissprosessissa Ojasalon ym. (2014) mukaan on kehittämissprosessin ja lopputulosten arviointi. Prosessin ja tulosten arviointia tehtiin ja tehdään jatkuvasti, myös sen jälkeen, kun Snowplow otettiin käyttöön Visma Novassa. Lopputulos on esitetty tämän opinnäytetyön luvussa viisi. Tutkimukselliseen kehitystyöhön kuuluu tärkeänä osana evaluointi eli arviointi. Prosessin eri vaiheita tarkastellaan ja arvostellaan kriittisesti alusta saakka. Eri vaiheiden onnistumista, tavoitteiden saavuttamista sekä vaikutuksia ja vaikuttavuutta tarkastellaan ja arvioidaan läpi työn. Myös toimeksiantajan palaute kuuluu tärkeänä osana evaluointiin. (Humak 2022.)

Tässä työssä arviointiin kuuluu pääasiassa kehittämissprosessin arviointia, sillä kauaskantoisia tuloksia Snowplow'n hyödyistä ja haitoista täytyisi odottaa jopa kuukausia. Kanasen (2014, 53–54) mukaan arviointi on työn peilaamista asetettuun tavoitteeseen; tässä tapauksessa tavoite oli saada ratkaisu toimimaan teknisesti, ja sitä on mahdollista arvioida välittömästi ratkaisun toteutuksen jälkeen.

## 4 TYÖN SUUNNITELMA JA TOTEUTUS

### 4.1 Visma Novan asiantuntijoiden haastattelut

Visma Novan asiantuntijatiimi on suhteellisen pieni; tiimi käsittää vain Visma Novan arkkitehdin, kolme kehittäjää, kaksi BA:ta (Business Analyst), yhden QA:n (Quality Assurer), sekä Team Managerin / PO:n (esihenkilö/Product Owner). Visma Novan asiantuntijoiden kanssa tehtiin yhteistyötä koko opinnäytteen ajan, ja haastattelujen ja tiedon lisäksi heiltä tuli ohjeet ja toiveet koskien opinnäytettä. Visma Novan asiantuntijat toivoivat, että itse koodipohjaan ei tarvitsisi koskea analytiikka-alustan vaihdossa vaan pyrittäisiin suoraan käyttämään aiemmin Google Analyticsin kanssa käytettyjä funktioita datan keräämiseen itse lähdekoodista.

Visma Novan tiimi päätti, että sovellukseen tulee Snowplow'n mukana käyttöön NPS-palautteen keruu, joka on uudistus Visma Novalle data-analytiikkajärjestelmässä. Koska toiminnanohjausjärjestelmä pitää sisällään useampia sovelluksia ja joitakin sovelluksia käyttää sama henkilö monesti päivässä, täytyy NPS:n keräystä suunnitella niin, ettei se ärsytä käyttäjiä, eikä kuormita vain yhtä henkilöä. Sovittiin, että NPS-kyselyt eivät tule kaikissa sovelluksissa samana päivänä kvartaalissa.

Visma-konsernilla on oma Snowplow-asiantuntija, ja tahtotila konsernissa olisi pitää Snowplow'n käyttö joiltain osin yhtenäisenä. Bryan Ye (2023) kertoo Snowplow'n tärkeimmän tarkoituksen Vismalla olevan seurata ohjelmistojen käyttäjien käyttäytymistä, säilöä käyttäjien käyttäytyminen BigQueryyn ja viimeisenä näyttää BigQueryn data Tableaussa. Asiantuntijan muistiinpanot, esitykset, luennot ja sähköpostit on otettu huomioon suunnittelussa ja toteutuksessa. Visma haluaa data-analytiikan osalta kustannustehokasta palvelua, uusia mitattavissa olevien KPI:den (Key Performance Indicator) luomista, teknisesti hyvin skaalautuvia alustoja, paikallisen omistuksen, vapaan pääsyn raakadataan ja yhtenäisen tietomallin eri toimialueille (Ye 2023).

Ennen vaihtoprosessin suunnittelua varmistettiin lakien ja asiantuntijoiden avulla GDPR:n asetusten toteutuminen. Visma Novan suhteen tehtiin päätöksiä koskien suoja-asetuksia: muun muassa käyttäjien toimintaa seurataan vain applikaation sisällä, näin ollen Visma Software Oy omistaa kaiken kerätyn datan itse. Ainoastaan Visma Novan lisenssin ostaneiden käyttäjien toimintoja voidaan seurata. Datasta nähdään lisenssin haltija (yritys), mutta spesifejä käyttäjiä ei voida datan tietojen perusteella tunnistaa.

Visma Novan Snowplow-implemентаaatiossa seurataan käyttödataa, eli esimerkiksi kuinka käyttäjät siirtyvät applikaation sisällä, ja mitä osia järjestelmästä käyttäjät käyttävät. Lisäksi seurataan статистиikkaa mikä näyttää, kuinka käyttäjät käyttävät järjestelmää ja järjestelmän tarjoamia toimintoja. Snowplow'lla ei tulla keräämään mitään sensitiivistä dataa, kuten esimerkiksi, nimiä, osoitteita, paikallistamistietoja, IP-osoitteita tai muuta vastaavaa. Dataa käytetään optimoimaan järjestelmän tarjoamia ominaisuuksia ja palveluita. Dataa käytetään myös статистиikkaan, jota analysoimalla voidaan tehdä parempia liiketoiminnallisia päätöksiä (esimerkiksi poistamaan ominaisuuksia, joita käyttäjät eivät käytä laisinkaan, tai lisätä rahan ja ajankäyttöä paljon käytettyjen osioiden kehitykseen). Myöhemmin dataa hyödynnetään myös profiloimaan erilaisia käyttäjärooleja; kuinka saadaan optimoitua järjestelmän toiminta kirjanpitäjän näkökulmasta, tai vaikkapa varastotyöntekijän näkökulmasta.

Mitään Snowplow'lla Visma Novasta saatua dataa ei jaeta muille yrityksille, asiakkaille tai ulkopuolisille osapuolille. Visman asiakkaat ovat hyväksyneet jo järjestelmän käyttöönottovaiheessa Visman oman TOS:in (Terms of Service), jota noudatetaan myös Visma Novassa. Tätä käytetään myös Snowplow'n kanssa. Visman TOS:issa on kirjattu datan maksimaalisen säilytysajan olevan kolme vuotta, joten myös BigQueryyn täytyy kehittää protokolla, jossa kolme vuotta vanhemmat tiedot poistetaan automaattisesti. Snowplow'n dokumentaatiosta selviää sen luovan automaattisen aikaleiman jokaiseen Eventtiin, joten protokollan tekeminen ei pitäisi olla ongelma. Vain Visma Novan tiimillä ja tuotepäälliköllä tulee olemaan pääsy Snowplow'n keräämään dataan.

## 4.2 Data-analytiikkajärjestelmän vaihtamisen suunnitelma

### 4.2.1 Tiimi

Ensimmäisenä luodaan kehittämistutkimuksen ideologian ja Visman ohjeistuksen mukaan tiimi. Visman ohjeen mukaan tiimissä on oltava mukana vähintään Visman Snowplow asiantuntija, tuotteen tuotepäällikkö, tuotteen arkkitehti sekä PO, ja luonnollisesti tässä tiimissä on myös opinnäytetyön tekijä, joka on vastuussa Snowplow'n implementoinnista (Ye, 2022).

Tiimin tehtävä on päättää, mikä tieto on tuotteelle tärkeää, mitä ja millaista tietoa pyritään keräämään. Tiimin päätökset ja toiveet kirjataan ylös, ja niiden mukaan muodostetaan Snowplow'n skeemat ja Eventit sekä kutsutaan Tracker. Tiimin luonnin jälkeen Snowplow asiantuntija luo tuotteelle oman tietovaraston BigQueryyn; yhden testauksarkoitukseen, ja toisen tuotteen käyttöön (Ye, 2022).

### 4.2.2 Kirjastot

Snowplow'n avoimen lähdekoodin analytiikka-alustan toiminta pohjautuu Snowplow Analyticsin ohjelmoimiin kirjastoihin. Kirjastoja käytetään ohjelmoinnissa antamaan valmiita toiminteita, ettei kaikkea tarvitse ohjelmoida alusta saakka. Snowplow'n rakentaminen Visma Novaan alkaa sillä, että asennetaan toiminnanohjausjärjestelmään kirjasto nimeltään Snowplow.Tracker.

Snowplow.Tracker on kirjasto, joka hoitaa Trackerin toiminnan. Snowplow'lla on tällä hetkellä tarjolla 19 erilaista kirjastoa, jotka on rakennettu eri koodikielten vuoksi (Snowplow, 2022). Koska Visma Novassa käytetään .NET -kieltä, käytetään siihen sopivaa kirjastoa. Kirjaston voi asentaa niin sanottuna NuGet-paketina, joka on Microsoftin kehitysympäristöön soveltuva pakettienhallintaohjelma, tai vaihtoehtoisesti erillisinä .dll-tiedostoina. Snowplow.Tracker pakettiin kuuluu kolme .dll-kirjastoa. Suunnitelmana on asentaa kirjasto ensin NuGet-paketina testausympäristöön, mutta lopulliseen versioon siirtää tiedostot yksittäisinä .dll:inä, jotta niitä olisi helpompi hallita.

### 4.2.3 Snowplow'n tutkinta

Tutkitaan Snowplow'n toimintaa; onko mahdollista olla koskematta jo olemassa oleviin GA:n tapahtumakutsuihin, ja rakentaa Snowplow'n Tracker niin, että GA:n tapahtumakutsuja voitaisiin käyttää myös Snowplow'ssa. Jos tämä osoittautuu mahdolliseksi, lähdetään rakentamaan Snowplow'n toimintaa ihan koodillisesti valmiiksi. Mikäli taas olemassa olevia funktiokutsuja ei voida käyttää tiedonvälittäjinä Trackerille, täytyy etsiä kaikki olemassa olevat tapahtumakutsut ja korvata ne Snowplow'n Eventeillä. Toisaalta tämä voisi olla teknillisesti helpompikin vaihtoehto, vaikka massiivisessa järjestelmässä työn määrä näin olisi suuri. Mutta jokainen muutos lähdekoodiin on riski uusille bugeille. Toinen ongelma tässä skenaariossa olisi se, että osa Visma Novan sovelluksista ei ole enää ylläpidossa ja osaa konvertoidaan uudelle koodikielelle. Näiden sovellusten kutsut jäisivät päivittämättä, ja arvokasta tietoa saamatta. Kun Snowplow on saatu lähettämään dataa eteenpäin, testataan toimivuutta testiympäristössä. Vasta kun on saatu aiemmin GA:lla kerätyt tiedot tallentumaan BigQueryyn, ryhdytään miettimään uusia jäljitettäviä asioita ja Product NPS:sää.

Asiaa lähdetään tutkimaan Visma Novan lähdekoodista; etsitään koodista Google Analyticsin tapahtumakutsut, ja tutkitaan niille annetut parametrit. Verrataan Snowplow'n vaatimiin tietoihin, riittävätkö annetut parametrit tai onko niitä liikaa. Mikäli kutsut ovat vaillinaisia Snowplow'n vaatimuksiin, tutkitaan, onko mahdollista tehdä Snowplow'n ja lähetyskutsujen välille apufunktiota, joka muuttaisi kutsut sopivaan muotoon Snowplow'n Trackerille tarpeellisine tietoineen.

Visma Novan lähdekoodia tutkiessa huomattiin, että data-analytiikan lähetyskutsut Google Analyticsille on rakennettu niin, että mikäli niitä muutettaisiin, vaikuttaisi se laajasti Visma Novan useisiin sovelluksiin. Koska samoja kutsuja käytetään sekä uudistetuissa sovelluksissa, aiemman koodikielen sovelluksissa, että jo ylläpidosta poistetuissa sovelluksissa (mutta joista kuitenkin tarvitaan tietoa) ja myös pilvipalvelussa, tultiin siihen lopputulokseen, että lähetyskutsujen muokkaaminen Visma Novan lähdekoodissa ei ole vaihtoehto. Data-analytiikalle on Visma Novassa olemassa oma projekti, joka sisältää data-analytiikkaan tarvittavat luokat. Paras vaihtoehto olisi saada kaikki muutokset luotua sinne. Tästä luo-

kasta luodaan instanssi Visma Novan käynnistyessä, jota käytetään joka soveluksessa. Mikäli kaikki muutokset saataisiin data-analytiikan omaan projektiin, voitaisiin olla varmoja, että data-analytiikka-alustan vaihto ei vaikuttaisi toiminnanohjausjärjestelmän toimintaan millään tavalla. Google Analyticsin huomattiin noudattavan kutsuissa sellaista kaavaa, että "välifunktiolla" (adapter function) muuttamalla ne voitaisiin saada toimimaan Snowplow'n Trackerin kanssa. Snowplow'hun täytyy rakentaa Trackeriin kustomoitu skeema, joka kerää samat tiedot kuin GA:n funktiot, ja silloin GA:n tapahtumakutsut saadaan toimimaan Snowplow'n kanssa. Snowplow voi kerätä huomattavasti enemmän tietoa kuin GA.

#### 4.2.4 Trackerin rakennus

Tracker rakennetaan Snowplow'n dokumentaation ohjeistuksen mukaan (Kuvio 3.). Trackeriin luodaan ensin Logger. Tämän jälkeen määritellään niin kutsuttu EndPoint, eli annetaan määrittelyt, minne Eventit lähetetään. Vismalla ne lähtevät Googlen BigQueryyn. Seuraavaksi määritellään väliaikainen tiedosto, joka säilyttää Eventin datan tietorakenteen jonossa siihen saakka, että data lähtee eteenpäin. Snowplow'n dokumentaatio pyytää käyttämään tähän .NET:in Trackerin kanssa erillistä .dll-tiedostoa; LiteDBStoragea. (Snowplow, 2022.) Tästä löytyi ensimmäinen mahdollinen ongelma Visma Novan suhteen.

```
// Create Logger
var logger = new ConsoleLogger();

// Controls the sending of events
var endpoint = new SnowplowHttpCollectorEndpoint(emitterUri, method: method, port: port, protocol: protocol,

// Controls the storage of events
// NOTE: You must dispose of storage yourself when closing your application!
var storage = new LiteDBStorage("events.db");

// Controls queueing events
var queue = new PersistentBlockingQueue(storage, new PayloadToJsonString());

// Controls pulling events of the queue and pushing them to the sender
var emitter = new AsyncEmitter(endpoint, queue, 1: logger);

// Contains information about who you are tracking
var subject = new Subject().SetPlatform(Platform.Mob).SetLang("EN");

Tracker.Tracker.Instance.Start(emitter: emitter, subject: subject, trackerNamespace: "some namespace", appId:
```

Kuvio 3. Esimerkki Snowplow'n Trackeristä (Snowplow 2022)

Snowplow luo väliaikaisen liteDB-tiedoston datalistaa varten. Visma Novan on-premise-sovelluksissa tämä tapahtuisi Visma Novan paikallisen kansion sisällä. Se ei häiritse millään tavalla toimintaa, koska tiedosto on paikallinen ja tuhoutuu heti datan lähtiessä eteenpäin. Mutta koska Visma Nova toimii myös SaaS-tuotteena sovelluspalvelussa, jossa eri asiakkaat käyttävät samaa palvelinta omilla tietokannoillaan, tämä voisi olla ongelma. LiteDB:n kansiopolkua ei voi määritellä Snowplow'n Trackeriin (Snowplow, 2022). Ainoastaan tiedoston nimi määritellään Storagea määriteltäessä. Yhtäaikaisia käyttäjiä saattaa olla kymmeniä, ja jos kaikilta lähtee Event samaan aikaan samannimiseen väliaikaistiedostoon samalle palvelimelle, miten voidaan olla varmoja, ettei ongelmaa synny? Koska Event pyyhkiytyy pois heti kun on jatkanut matkaa, suurta vaaraa tietoturvalle tästä tuskin olisi, mutta ratkaisu ei tunnu tarpeeksi varmalta. Tässä vaiheessa päätettiin, että kun Tracker on saatu toimimaan dokumentaation mukaisesti toteutettuna testiserverillä ja Tracker lähettää datan tietokantaan, selvitetään tarkemmin muita vaihtoehtoja tämän tiedoston suhteen.

Storagen jälkeen määritellään Eventien jonotus, ja tässä käytetään edellä mainittua liteDB:n väliaikaistiedostoa siihen tarkoitukseen. Seuraavaksi alustetaan Emitter, jonka vastuu on ottaa Eventejä Storagesta ja lähettää ne eteenpäin. Viimeisenä Subject määrittelee käyttäjädataa. Koska käytämme erillistä kustomoitua skeemaa Trackerissa, sen määrittelyt tehdään tässä vaiheessa ennen jäljityksen aloittamista. (Snowplow, 2022.)

#### 4.2.5 Snowplow'n toiminta

Määritysten jälkeen Tracker käynnistetään. Start()-funktio käynnistää Emitterin ja mahdollisen ajastimen (ei käytössä Visma Novassa) sekä valmistelee kaiken valmiiksi synkronoitua tapahtumaa varten. Startin jälkeen käytetään Track()-käskyä, joka tekee itse tiedon seurannan. Track()-funktioita on Snowplow'n .NET Trackerissa useita eri tarkoituksiin, esimerkiksi PageView, ScreenView ja Timing. Koska Visma Nova tarvitsee kuitenkin tiettyä dataa tietyistä tiedoista ja koska olemassa olevaan lähdekoodiin ei haluta koskea data-analytiikan osalta, luodaan oma Track(Structured)-funktio. Flush()-funktioilla tyhjennetään kaikki lähetyksen ja jäljityksen tiedot ennen seuraavaa jäljitystapahtumaa. Stop()-funktioilla pysäytetään koko tiedonkeruutapahtuma, tämä tapahtuu viimeisenä. (Snowplow, 2022.)

Trackeriin lähetetään Google Analyticsin olemassa olevat kutsut. Eli varsinaisessa Visma Novan lähdekoodissa on kutsu lähettää tietoa analytiikalle; halutaan lähettää esimerkiksi tieto, että laskutussovellus avattiin. Tämän lähdekoodin täytyy pysyä samana Visma Novassa. Tässä vaiheessa Snowplow'n luokassa luodaan samanniminen funktio, joka ottaa vastaan samat parametrit (sovellus avattu, laskutus) kuin Google Analyticsin funktio vastaanotti ja lähetti. Snowplow'ssa samannimisessä funktiossa muutetaan tiedot siihen muotoon, missä ne tarvitaan Snowplow'n Trackerissa, ja vastataan ne tästä välifunktiosta Trackerin Event-kutsuun. Näiden tietojen lisäksi Trackerissa kerätään tiedot asiakkaasta, ja sitä varten Trackerin sisään luodaan kustomoitu skeema. Koska Snowplow pystyy keräämään enemmän ja erityyppistä tietoa kuin Google Analytics, täytyy miettiä, mistä kyseiset tiedot haetaan. Tarvittavat parametrit päätettiin sisällyttää Snowplow-olion constructor-metodiin, jossa luodaan uusi olio luokasta.

#### 4.2.6 Tietovarasto

Visma käyttää tietovarastona Googlen BigQuerya. Visman Snowplow-asiantuntija on tehnyt samantyyppisen tietokantapohjan käyttöönotettavaksi kaikille Visman tuotteille. Käytössä on kaksi projektiympäristöä; testitietokanta, jossa Snowplow'n toimintaa voi testata, sekä varsinainen tietovarasto, johon oikea data ohjataan, kunhan implementaatio on testattu toimivaksi.

Kun implementaatio on todettu toimivaksi, se otetaan käyttöön Visma Novassa. Tutkitaan millaisessa muodossa ja miten tapahtumat tallentuvat tietokantaan. Tämän jälkeen korjataan tarpeen mukaan skeemoja tai Trackeria, kunnes tieto saadaan haluttuun muotoon tietokantaan. Uusien mittauspisteiden lisääminen on mahdollista, mikäli Snowplow-tiimistä tulee näistä toiveita ja tarpeita esille.

#### 4.2.7 NPS

Kun GA:sta aiemmin saatu data on onnistuttu kuljettamaan Snowplow'lla BigQueryyn, lähdetään hahmottelemaan NPS-mittauksen käytäntöä. Koska Visma Nova järjestelmässä on monia sovelluksia, ja kaikista halutaan palautetta, täytyy ottaa huomioon käyttäjät. Täytyy miettiä tarkkaan, miten kyselyt ajastetaan niin, ettei yksi ihminen kuormitu niiden vastaamisesta, ja saadaan mahdollisimman monipuolisesti vastauksia. Mikäli yrityksessä on esimerkiksi vain muutama



toimistotyöläinen, olisi ikävää, jos jokaisen avatun sovelluksen myötä tulisi NPS-lomake samana päivänä täytettäväksi samalle henkilölle. Tällaisessa tapauksessa voisi olla lähes varma, että kyselyihin ei jaksettaisi vastata, tai ainakin vastaaminen alkaisi ärsyttää vastaajia, mikä voisi vaikuttaa tulosten luotettavuuteen.

Otetaan selvää Snowplow-tiimiltä, miten kyselyt halutaan toteuttaa. Luodaan NPS-kyselylle ikkuna Visma Novan ulkonäköä mukaillen. Koodillisesti NPS on helppo toteuttaa, koska vastaajat valitsevat arvosanan väliltä 1–10, ja vastaus lähetetään Snowplow'iin.

#### 4.2.8 Käyttöliittymä

Snowplow'lla ei ole valmista käyttöliittymää analysoimaan dataa ja näyttämään tuloksia visuaalisesti. Visman Snowplow-tiimi on rakentanut yhtiölle oman yleis-pätevän dashboardin Tableaulla, jota käytetään datan visualisoinnissa BigQueryn kautta. Tableaun toimintaa testataan Visma Novassa, kunhan tietokantaan on saatu tarpeeksi dataa. Mikäli kyseinen Tableau osaa havainnollistaa Visma Novalle tärkeitä tietoja, otetaan tämä käyttöön. Mikäli jotain tarpeellista jää näkemättä, voi olla, että Tableautta joudutaan modifioimaan Visma Novaan sopivaksi.

Tableau-käyttöliittymää ei päästy modifioimaan opinnäytteen aikana, sillä se ei ole implementoitu Snowplow'n testin aikaiselle serverille. Snowplow-tiimi alkaa rakentamaan Visma Novan tietokantaan sopivaa Tableautta vasta siinä vaiheessa, kun dataa on kertynyt varsinaiselle käytössä olevalle serverille implementaation jälkeen.

#### 4.3 Toteutus, testit ja ongelmat

Jo ennen ohjelmoimisen aloittamista varmistettiin, että aiemmin suunnitelmissa kuvattu GDPR-politiikka toteutettiin. Täytettiin GDPR-varmistuslomake (Visman oma Use of Data Assessment -lomake), jossa käytiin läpi erilaiset skenaariot ja käytännöt, jotka voivat vaikuttaa GDPR:n toteutumiseen ja yleiseen tietoturvaan. Lomaketta täyttäessä nousi esiin datan säilytysaika; oli luotava protokolla, jolla kolme vuotta vanhemmat datat saadaan automaattisesti poistettua. Koska Snowplow'lla ei ole enää pääsyä dataan sen jälkeen, kun se on siirretty, tämä täytyi

toteuttaa BigQueryssa. Lomakkeen tarkisti Visman Security Specialist, ja lomakkeen hyväksymisen jälkeen sai käyttöönsä Visma Novan BigQueryn testikannan.

Visma Novasta luotiin erillinen testiprojekti, jonne siirrettiin tämänhetkiset Google Analyticsin kutsut Visma Novasta, ja testattiin millä tavoin ne saadaan nykyisessä muodossa lähettämään dataa BigQueryyn Snowplow'n kautta. Testiprojektiin liitettiin Snowplow.Tracker kirjastot NuGet-pakettina. Tehtiin luokka Snowplow'lle ja luokkaan rakennettiin Snowplow'n Tracker. Lopuksi Visma Novan analytiikka-luokan unit-testejä siirrettiin Snowplow-testiprojektiin, jotta datan lähetystä voitiin testata GA:n kutsuilla.

Kun testiprojektissa saatiin data kulkemaan BigQueryn testiserverille, lähdettiin miettimään vaihtoehtoja LiteDbStoragelle, ja yksi idea oli listan luominen. Huomattiin LiteDbStoragen implementoivan IStorage rajapintaluokkaa(interface). Tutkittiin Snowplow'n IStoragen ja LiteDBStoragen määritelmiä ja mitä ne sisältävät. Päätettiin kokeilla rakentaa LiteDbStoragen kaltainen luokka, joka sisältää samat käskyt ja joka implementoi IStoragea samanlailla kuin LiteDbStorage. Käytännössä yritettiin rakentaa identtistä luokkaa LiteDbStoragelle, joka kuitenkin väliaikaistiedoston sijaan luo listan. Tämä osoittautui kuitenkin vaikeammaksi kuin yksinkertaisuudessaan kuulostaa; pakolliset funktiot, jotka kuljettavat dataa paikasta toiseen oikealla hetkellä, pitää saada toimimaan samalla tavalla listan kanssa. Snowplow'n koodissa on niin monta osasta ja tarkkaa synkronointia, että se vaati paljon mietintää ja kokeilua. Jokaisessa funktiossa käytettiin LiteDb-kirjaston sisäisiä funktioita. Täytyi etsiä mitä funktiot tekevät, ja rakentaa uudet funktiot, jotka toimivat samalla tavalla listan kanssa. Useiden testien jälkeen päädyttiin käyttämään LiteDb-kirjastoa apuna, että voitiin hyödyntää muutamaa spesifiä funktiota. Funktioita käytettiin listan kanssa, ja lopulta saatiin tieto kulkemaan oikein.

Luonnollisesti, kun data saatiin kulkemaan jopa listan kanssa tietokantaan, kuviteltiin työn olevan valmis. Testejä tehdessä huomattiin kuitenkin, että lähetyksen performanssi oli huono. Tieto liikkui aivan liian hitaasti, ja sellaisenaan hidastaisi Visma Novan käyttöä. Ensin etsittiin syytä koodista, mutta lisää testatessa huomattiin saman koodin eri ympäristössä lyhentävän suoritusaikaa jopa 91,3%. Testattiin kirjastojen ja ohjelman eri versioiden vaikutusta nopeuteen, ja löydettiin

Snowplow.Tracker -dll:n version vaikuttavan nopeuteen. Valittiin Visma Novaan parhaiten sopiva versio. Tämä ei kuitenkaan vaikuttanut tarpeeksi lähetyksen nopeuteen, joten muutettiin Snowplow'n toimintamallia lähemmäs verkkosivujen toimintamallia. Kun muutettiin, että Tracker luodaan vain kerran ohjelman avautuessa, ja sama Tracker heittää Event-kutsuja koko ohjelman käytön ajan, lähetyksen nopeutui radikaalisti. Vasta ohjelman suljettaessa Tracker-olio hävitetään. Huomattavaa tässä oli, että ilman Flush()-komentoa data ei lähtenyt reaaliajassa listalta eteenpäin tietovarastoon.

NPS:n käyttöönotto ja lisämittauspisteiden lisäys päätettiin siirtää myöhempään ajankohtaan johtuen tiimin tiukoista aikatauluista; yhteinen kokous näistä saatiin järjestymään vasta reippaasti myöhempään ajankohtaan.

## 5 POHDINTA

Opinnäytetyön lähtökohtana olivat Google Analyticsin GDPR-ongelma, ja ongelma ratkaistiin; Visma Novassa ei enää ole GDPR-ongelmia. Yleistä tietosuojasetusta noudatetaan Snowplow'n avulla datan keräämisessä. Tietosuojasetus otettiin huomioon suunnitelmassa ja toteutuksessa. Data-analytiikkajärjestelmän vaihto suunniteltiin ja toteutettiin onnistuneesti niin, että Visma Novan varsinaiseen lähdekoodiin ei tarvinnut puuttua kuin analytiikkaolion luomisen muodossa.

Pääkysymyksenä oli, miten data-analytiikkajärjestelmän vaihtoprosessi tulisi toteuttaa. Visma Novan, asiantuntijoiden haastattelujen sekä Snowplow'n dokumentaation avulla tehtiin vaihtoprosessin suunnitelma. Suunnitelman teossa haasteita oli lähinnä tiedon kerääminen. Snowplow'n implementoinnista muualle kuin verkkosivustoihin oli yllättävänkin vähän dokumentaatiota. Myös Snowplow'n oma dokumentaatio .NET-kirjaston käytöstä oli hyvin rajallinen. Snowplow on hyvin laaja, mutta suurin osa laajasta dokumentaatiosta ja internetsivustoista koskee nimenomaan käyttöä verkkosivustoilla. Verkkosivustoille on paljon valmiita vaihtoehtoja, mutta Visma Novan käyttämää .NET-kirjastoa varten vaihtoehtoja ei käytännössä ollut.

.NET Trackeriin, jota Visma Novaan käytettiin, oli käytettävissä yksi tuettu Storage, yksi Emitter, yksi Endpoint, yksi Queue ja yksi Payload Serializer. Snowplow'n toimintaan tarvittiin kuitenkin niitä kaikkia. Lopuksi Storagesta päädyttiin tässä työssä ohjelmoimaan kokonaan uusi versio, jotta se saataisiin Novan tarpeiden mukaiseksi muokattua. Oli kuitenkin positiivista huomata, että vaikka Snowplow kertoo dokumentaatioissaan vaihtoehtoja olevan vain yksi, useampia vaihtoehtoja voi itse ohjelmoida lisää. Toki Snowplow'n idea on olla helppo ottaa käyttöön valmiiden komponenttien avulla, mutta yleisestikin tärkeä ja huomionarvoinen asia on, että tarvittaessa itse ohjelmoimalla uusia komponentteja voidaan Snowplow'hin saada lisää ulottuvuuksia sen toimintaan. Myös Snowplow'n valmiista Track()-tapahtumista huomaa verkkosivujen tuennan; Visma Novan käyttämästä kirjastosta valmiina löytyy muun muassa verkkosivuilla sivujen seurannan ja näytön katselun seurannan sekä verkkokauppatapahtuman seuranta. On-premise -sovellukseen täytyy itse rakentaa Eventit seuraamaan haluttuja kohteita. Verkkosivut toimivat toisaalta samankaltaisesti, joten valmiita Eventeja on

ehkä helpompi rakentaa yleisellä tasolla joka sivuun sopivaksi. Joka tapauksessa oli huomattavaa, että dokumentaatio on vajaampaa esimerkiksi juuri .NET-kirjastolle kuin esimerkiksi JavaScript-kirjastolle.

Snowplow'n käyttöönotto ja implementaatio ei ollut ihan niin suoraviivaista ja helppoa kuin oletettiin kerätyn tietoperustan perusteella. Tämä tosin riippune myös mitä kirjastoa käyttää; osalle, varsinkin verkkosivuista, on jo kaikki valmiina. Tuntuukin, että suurin osa Snowplow'sta kertovista lähteistä puhuvatkin nimenomaan Snowplow'n implementoinnista ja käyttäytymisestä verkkosivustoilla. Vaikka suunnitelma vaihtoprosessista tehtiin huolellisesti tietoperusta ja aineisto huomioonottaen, ongelmia esiintyi silti. Snowplow'n kirjastot implementoivat toisia kirjastoja, ja on lopulta hankalaa hahmottaa mitä mikäkin osanen tekee ohjelmassa, ja mikä kirjasto ja funktio on vastuussa mistäkin tapahtumasta. Esimerkiksi suoristuskyvyn ongelmat veivät huomattavan paljon aikaa ratkaista, koska ei ollut täysin selvää, mitä missäkin vaiheessa tiedon Trackausta tapahtuu. Ei myöskään ollut täysin selvää, mitä pilvipalvelussa olisi tapahtunut väliaikaiselle LiteDb-tiedostolle, mistä syystä sitä ei uskallettu Visma Novassa ottaa käyttöön. Kokeilujen ja testauksen avulla nämä ongelmat saatiin kuitenkin ratkaistua, jolloin tavoite täyttyi näiltäkin osin.

Muita tutkimuskysymyksiä olivat, miten vanhaa dataa käsitellään sekä onnistuuko NPS-mittaus Snowplow'n avulla. Google Analyticsillä kerätyn datan siirtäminen uuteen tietovarastoon olisi ollut liian hankalaa, että se kannattaisi ajankäytöllisesti. Google Analyticsin tili lakkautetaan, kun Snowplow'lla on saatu kerättyä dataa sen verran, että pystytään vertailemaan, että saatu tieto varmasti vastaa toisiaan. NPS-mittauksen toteuttamista tutkittiin, ja myös se tulee onnistumaan Snowplow'n kautta. Uuden toiminnallisuuden käyttöönotto tarvitsee kuitenkin yhtiössä sisäisiä päätöksiä, ja tiimin aikataulujen haasteista johtuen NPS:n käyttöönotto viivästyy hieman. Tässä myöhemmin olevassa kokouksessa päätetään myös uusista mittauspisteistä, jotka implementoidaan Visma Novaan yhtä aikaa NPS-mittauksen kanssa.

Data-analyysiin tarvittava tieto kerääntyy talteen Snowplow'n avulla valittuun BigQueryn tietovarastoon Eu:n yleisen tietosuoja-asetuksen mukaisesti. Visma Novan lähdekoodia ei muutettu ja matkalla olleet ongelmat ratkaistiin. Tämä oli

oikeastaan mittarina tulokselle, ja se toteutui. Visma hyötyi data-analytiikka-alustan päivityksestä ja GDPR-ongelman poistumisesta. Snowplow tuo myös tulevaisuudessa antamaan data-analytiikan suhteen enemmän mahdollisuuksia kuin aiemmin, myös analysoinnin osalta. Kun tiedot omistaa itse, niitä voi käyttää miten tarve on. Vaillinaisiksi jäivät Product NPS-mittaus, joka toteutetaan myöhemmin, sekä vanhan datan saatavuus. Vanhan datan osalta päätettiin, ettei siihen kannata käyttää aikaa. Kaiken kaikkiaan tulos oli onnistunut, kun opinnäytteen päätavoitteet saavutettiin.

Tässä opinnäytteessä ei ollut salassa pidettäviä tai muita sensitiivisiä suojeltavia tietoja, joten tämän suhteen eettisiä ongelmia ei työssä ollut. Toikko ja Rantanen (2009) totesivat että koska kehitystoiminnallisella opinnäytetyöllä pyritään aina positiivisiin tuloksiin, siinä sorrutaan herkästi liian positiiviseen raportointiin. Liian positiivinen raportointi saattaa viedä opinnäytetyötä epäluotettavaan suuntaan. Opinnäytetyön raportoinnissa täytyi pyrkiä olemaan realistinen ja todenmukainen korostamatta liikaa onnistumisia ja saavutuksia. Opinnäytteessä on kiinnitetty huomiota myös ongelmakohtiin, ja kirjoitusprosessissa on pyritty pysymään etäisenä. Tässä työssä haastateltavia ja kuunneltavia ei ollut montaa, ja tutkimusaineisto haastattelujen osalta on suppeahko. Tämä vaikuttaa työn luotettavuuteen hienoisesti, koska tämä on toistettavuuteen liittyvä ongelma. Haastattelujen määrään ei kuitenkaan voi vaikuttaa, sillä haastateltavien on oltava Visma Novan asiantuntijoita, ja Visma Novan tiimi on suhteellisen pieni.

Kehittämistoiminnassa luotettavuus tarkoittaa pääasiallisesti työn käyttökelpoisuutta. Todenmukainen tieto yksinään ei riitä, vaan sen tulee olla nimenomaan kyseiseen työhön soveltuvaa sekä ennen kaikkea olla hyödyllistä. (Toikko&Rantanen 2009.) Lopputulos oli hyödyllinen ja toimiva, juuri kyseiseen toimeksiantoon soveltuva. Toimeksiantaja pystyy implementoidulla analytiikka-alustalla seuraamaan kaikkea haluamaansa dataa. Implementaatio toteutettiin niin, että lähdekoodiin on helppo ja nopea tehdä Snowplow'n Tracker-kutsuja, ja saada uusia mittauspisteitä ja lisää dataa. Opinnäytteen aikana todettiin, että Snowplow'n omia kirjastoja voi tiettyjen raamien puitteissa ohjelmoida itse, kuten tässä työssä tehtiin LiteDb-tietokantakirjastolle (Storage). Se luo vielä enemmän mahdollisuuksia Snowplow'n käyttöön alustoille, joissa Trackeria ei ole kehitetty monipuol-

liseksi. Näin ollen jatkotutkimuksena voisi olla hyödyllistä tutkia Snowplow'n suppeampien kirjastojen laajentamista. Jos kaikkiin Trackerin osiin onnistuisi ohjelmoimaan lisävaihtoehtoja, se voisi tuoda helpotusta monellekin taholle.

## LÄHTEET

Ahuja, R., Asthana, S., Ahuja, A. & Agarwal, M. 2020. Data: Its Nature and Modern Data Analytical Tools. Gupta, D., Bhattacharyya, S., Khanna, A. & Sagar, K. (toim.) Intelligent Data Analysis. From Data Gathering to Data Comprehension. Hoboken, NJ: Wiley.

Castagna, R. 2021. General Data Protection Regulation (GDPR). TechTarget. Viitattu 24.11.2022 <https://www.techtarget.com/whatis/definition/General-Data-Protection-Regulation-GDPR>.

Datatilsynet. 2022. Press release: Use of Google Analytics for web analytics. 21.9.2022. Viitattu 14.10.2022 <https://www.datatilsynet.dk/english/google-analytics/use-of-google-analytics-for-web-analytics>

Frankenfield, J. 2020. General Data Protection Regulation (GDPR) Definition and Meaning. Investopedia 11.11.2020. Viitattu 24.11.2022 <https://www.investopedia.com/terms/g/general-data-protection-regulation-gdpr.asp>

– 2022. Data Analytics: What it is, How it's used, and 4 basic techniques. Investopedia. Investopedia 2022. Viitattu 11.11.2022 <https://www.investopedia.com/terms/d/data-analytics.asp>

Google Cloud, 2023. BigQuery Documentation - Guides. What is BigQuery? Viitattu 16.1.2023 <https://cloud.google.com/bigquery/docs/introduction>

Hokkanen, J. 2022. Snowplow-analytiikkaratkaisun käyttö Saas-alustalla. Opinnäytetyö, Jyväskylän ammattikorkeakoulu. Viitattu 16.11.2022. <https://urn.fi/URN:NBN:fi:amk-202205098091>

Humak 2022. Tutkimuksellisen kehittämistyön lähestymistavat ja menetelmät. Humanistinen ammattikorkeakoulu. Viitattu 14.10.2022 <https://humak.libguides.com/c.php?g=688355>

Igloo Analytics. 2019a. Google Analytics360 vs Snowplow Analytics. Viitattu 14.10.2022 <https://www.igloanalytics.com/google-analytics-360-vs-snowplow-analytics.html>

– 2019b. What is Snowplow Analytics - A new digital analytics paradigm. Viitattu 25.11.2022 <https://www.igloanalytics.com/what-is-snowplow-analytics.html>

Johnson, J. 2022. Snowplow Analytics Wins 2022 Data Breakthrough Award. Data Breakthrough. Los Angeles: PR Newswire US.

Jyväskylän yliopisto - Koppa 2015. Tapaustutkimus. Viitattu 16.11.2022 <https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tutkimusstrategiat/tapaustutkimus>

Kananen, J. 2012. Kehittämistutkimus opinnäytetyönä. Jyväskylän ammattikorkeakoulun julkaisuja -sarja. Jyväskylä: Jyväskylän ammattikorkeakoulu.



Key2Market 2016. Snowplow vs. Google Analytics. Viitattu 14.10.2022  
<https://medium.com/@key2market/snowplow-vs-google-analytics-5a03d3299e1e>

Kumar, P. 2020. Advantages and Limitations of Data Analytics. Rapid Sigma Solutions LLP 28.8.2020. Viitattu 14.10.2022  
<https://www.sigmamagic.com/blogs/analytics-advantages-and-limitations/>

Lindström, S. 2019. Google BigQuery on toimiva tietovarasto datan hallintaan ja työkalu markkinointistrategian toteutukseen. Ite Wiki 29.11.2019. Viitattu 16.1.2023 <https://www.itewiki.fi/blog/2019/11/google-bigquery-on-toimiva-tietovarasto-datan-hallintaan-ja-tyokalu-markkinointistrategian-toteutukseen/>

Lukka, K. 2001. Kari Lukka: Konstruktiivinen tutkimusote. Metodix – Metoditietämystä kaikille. Viitattu 12.1.2023 <https://metodix.fi/2014/05/19/lukka-konstruktiivinen-tutkimusote/>

Merlevede, J. 2020. An introduction to Snowplow. Towards Data Science 11.3.2020. Viitattu 25.11.2022 <https://towardsdatascience.com/what-is-snowplow-and-do-i-need-it-cbe30fcb302b>

Official Journal of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council. Viitattu 24.11.2022 <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e3265-1-1>

Ojasalo, K., Moilanen, T. & Ritalahti, J. 2009. Kehittämistyön menetelmät – Uudenlaista osaamista liiketoimintaan. Porvoo: WSOY pro Oy.

– 2014. Kehittämistyön menetelmät. Kotipaikkakunta: Sanoma Pro.

Oracle, 2022. What is ERP? Oracle. Viitattu 14.10.2022  
[https://www.oracle.com/erp/what-is-erp/#:~:text=started%20with%20ERP-,Definition%20of%20enterprise%20resource%20planning%20\(ERP\),compliance%2C%20and%20supply%20chain%20operations](https://www.oracle.com/erp/what-is-erp/#:~:text=started%20with%20ERP-,Definition%20of%20enterprise%20resource%20planning%20(ERP),compliance%2C%20and%20supply%20chain%20operations)

Poddar, A. 2021. Understanding Snowplow Analytics: 5 comprehensive Aspects. Hevo Data 6.8.2021. Viitattu 25.11.2022  
<https://hevodata.com/learn/snowplow-analytics/>

Pyyny, P. 2022. Viranomainen: Google Analytics on laitton Euroopassa. AfterDawn Oy 14.1.2022. Viitattu 14.10.2022  
<https://fin.afterdawn.com/uutiset/2022/01/14/google-analytics-laitton>

Salonen, K., Eloranta, S., Hautala, T. & Kinos, S. 2017. Kehittämistoiminta ja kehittämisen menetelmiä ammatillisessa korkeakoulutuksessa. Kotipaikkakunta: Turun ammattikorkeakoulu.

Santos, M. & Costa, C. 2020. Big Data. Concepts, warehousing, and analytics. Denmark: River Publishers.

Snowplow, 2022. Advanced analytics. Viitattu 14.10.2022  
<https://snowplow.io/advanced-analytics/>

Stedman, G. 2020. Data Analytics (DA). TechTarget. Viitattu 14.10.2022  
[https://www.techtarget.com/searchdatamanagement/definition/data-analytics#:~:text=Data%20analytics%20\(DA\)%20is%20the,of%20specialized%20systems%20and%20software](https://www.techtarget.com/searchdatamanagement/definition/data-analytics#:~:text=Data%20analytics%20(DA)%20is%20the,of%20specialized%20systems%20and%20software)

Tableau. 2023. Why choose Tableau? Viitattu 16.1.2023  
<https://www.tableau.com/why-tableau>

The Investopedia Team, 2022. Enterprise Resource Planning (ERP): Meaning, Components, and Examples. Viitattu 14.10.2022  
<https://www.investopedia.com/terms/e/erp.asp>

Tilastokeskus, 2015. Tutkimus- ja kehittämistoiminta. Viitattu 30.11.2022  
[https://www.stat.fi/meta/kas/t\\_ktoiminta.html](https://www.stat.fi/meta/kas/t_ktoiminta.html)

Tietosuojavaltuutetun toimisto. 2022. Euroopan tietosuojaviranomaiset ovat todenneet Google Analyticsin käytön verkkosivuilla tietosuojalainsäädännön vastaiseksi. Viitattu 14.10.2022 <https://tietosuoja.fi/-/euroopan-tietosuojaviranomaiset-ovat-todenneet-google-analyticsin-kayton-verkkosivuilla-tietosuojalainsaadannon-vastaiseksi>

Toikko, T. & Rantanen, T. 2009. Tutkimuksellinen kehittämistoiminta. Kotipaikkakunta: Tampere University Press.

Viita, P. 2022. Visma Software Oy. Software architectin haastattelu 24.11.2022.

Visma, 2022. Visma Nova Toiminnanohjausjärjestelmä (ERP). Viitattu 14.10.2022 <https://www.visma.fi/nova/toiminnanohjausjarjestelma/>

Vuori, J. 2021. Tapaustutkimus. Teoksessa J. Vuori (toim.) Laadullisen tutkimuksen verkkokäsikirja. Tampere: Yhteiskuntatieteellinen tietoarkisto. Viitattu 12.1.2023  
<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/tutkimusasetelma/tapaustudkimus/>

Wolford, B. 2022. What is GDPR, the EU's new data protection law? GDPR.EU. Proton AG. Viitattu 24.11.2022 <https://gdpr.eu/what-is-gdpr/?cn-reloaded=1>

Ye, B. 2022. Visma – Introduction to Snowplow. Norway: Visma.

– 2023. Implementing Snowplow to Visma Nova. Yksityinen sähköpostiviesti 26.1.2023. Viestin saaja: Malla Marttala.

Yin, R. 2018. Case study research and applications. Design and methods. UK: SAGE Publications Ltd.

Your Europe, 2022. Yleinen tietosuoja-asetus. Euroopan unionin virallinen verkkosivusto. Viitattu 24.11.2022  
[https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index\\_fi.htm](https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_fi.htm)