

Dataintegraatoratkaisut yritysten dataekosysteemeissä

LAB-ammattikorkeakoulu

Insinööri (AMK)

2023

Heidi Paananen

Tiivistelmä

Tekijä(t) Heidi Paananen	Julkaisun laji Opinnäytetyö, AMK Sivumäärä 72	Valmistumisaika 2023
Työn nimi Dataintegraatoratkaisut yritysten dataekosysteemeissä		
Tutkinto ja koulutusala Insinööri (AMK), tieto- ja viestintätekniikan koulutus		
Toimeksiantajaorganisaatio (jos opinnäytetyöllä on toimeksiantaja) LAB-ammattikorkeakoulu / Kyberturvalliset ekosysteemit Etelä-Karjala		
Tiivistelmä <p>Työn tavoitteena on tarjota yrityksille kattava yleiskuva eri dataintegraatiomallien ominaisuuksista, hyödyistä ja haasteista sekä niiden soveltuvuudesta eri liiketoimintaympäristöihin ja dataekosysteemeihin. Työ on toteutettu osana hanketta Kyberturvalliset ekosysteemit Etelä-Karjala.</p> <p>Aineisto koostuu kirjallisuuskatsauksesta ja kuuden anonymisoidun yrityksen haastatteluista, joiden avulla tutkittiin dataintegraatiomallien toimivuutta eri liiketoimintatarpeisiin ja kehitettiin suosituksia dataintegraatoratkaisujen valintaan.</p> <p>Tutkimustuloksissa korostetaan yritysten sisäisten dataintegraatioiden toimivuuden, järjestelmien yhteensopivuuden ja datan laadun varmistamista ennen laajentamista ulkoisiin dataintegraatioihin. Johtopäätöksenä todetaan, että yrityksen resursseihin, tarpeisiin ja kasvutavoitteisiin sopivan dataintegraatoratkaisun valitseminen on kriittistä, koska sillä on merkittävät pitkäkestoiset vaikutukset yrityksen liiketoimintaan.</p> <p>Jatkokehityksenä tutkimusta voitaisiin laajentaa kattamaan useampia yrityksiä ja toimialoja sekä tutkia syvällisemmin dataintegraatioiden taloudellisia vaikutuksia.</p>		
Asiasanat dataekosysteemi, dataintegraatio, dataintegraatiomalli		

Abstract

Author(s) Heidi Paananen	Type of Publication Thesis, UAS Number of Pages 72	Published 2023
Title of Publication Data Integration Solutions in Corporate Data Ecosystems		
Degree, Field of Study Engineer (UAS), Information and Communications Technology		
Organisation of the client (if the thesis work is commissioned by another party) LAB University of Applied Sciences / Cyber secure data sharing ecosystems of South Karelia		
Abstract <p>The aim of the thesis is to provide businesses with a comprehensive overview of the characteristics, advantages and challenges of various data integration models, and their applicability in different business environments and data ecosystems, as a part of the project Cyber secure data sharing ecosystems of South Karelia.</p> <p>The research consists of a literature review and interviews with six anonymized companies to evaluate data integration models for specific business needs and develop recommendations for selecting a data integration solution.</p> <p>The research findings highlight the importance of ensuring the functionality of internal data integrations, system compatibility, and data quality before expanding externally. Conclusively, choosing a data integration solution that suits a company's resources, needs and growth goals is critical due to its significant, long-term business impact.</p> <p>Future research could be expanded to encompass more companies, explore industry-specific differences and suitability, and further refine recommendations and best practices for companies seeking to enhance their businesses via data ecosystems.</p>		
Keywords data ecosystem, data integration, data integration model		

Sisällys

1	Johdanto.....	1
2	Dataintegraatio ja dataekosysteemi.....	2
2.1	Dataintegraation määritelmä	2
2.2	Dataekosysteemin määritelmä.....	2
3	Dataintegraatiomallit	3
3.1	Manuaalinen integraatio	3
3.1.1	Dataintegraatiomallin hyödyt.....	3
3.1.2	Dataintegraatiomallin haasteet.....	4
3.1.3	Dataintegraatiomallin tietoturva ja tietosuoja	4
3.1.4	Dataintegraatiomallin soveltuvuus	4
3.2	Robottivusteinen ohjelmistoautomaatio	5
3.2.1	Dataintegraatiomallin hyödyt.....	9
3.2.2	Dataintegraatiomallin haasteet.....	10
3.2.3	Dataintegraatiomallin tietoturva ja tietosuoja	11
3.2.4	Dataintegraatiomallin soveltuvuus	11
3.3	Pistemäinen integraatio (Point-to-point).....	12
3.3.1	Dataintegraatiomallin hyödyt.....	13
3.3.2	Dataintegraatiomallin haasteet.....	13
3.3.3	Dataintegraatiomallin tietoturva ja tietosuoja	14
3.3.4	Dataintegraatiomallin soveltuvuus	14
3.4	Datan jakeluun perustuva integraatio (Hub-and-spoke).....	14
3.4.1	Dataintegraatiomallin hyödyt.....	15
3.4.2	Dataintegraatiomallin haasteet.....	16
3.4.3	Dataintegraatiomallin tietoturva ja tietosuoja	16
3.4.4	Dataintegraatiomallin soveltuvuus	16
3.5	Yrityksen laajuinen integraatio	17
3.5.1	Dataintegraatiomallin hyödyt.....	20
3.5.2	Dataintegraatiomallin haasteet.....	20
3.5.3	Dataintegraatiomallin tietoturva ja tietosuoja	21
3.5.4	Dataintegraatiomallin soveltuvuus	21
3.6	Sovellusten ohjelmointirajapintaintegraatiot (API)	21
3.6.1	Dataintegraatiomallin hyödyt.....	27
3.6.2	Dataintegraatiomallin haasteet.....	29
3.6.3	Dataintegraatiomallin tietoturva ja tietosuoja	30

3.6.4	Dataintegraatiomallin soveltuvuus	31
4	Datan laadunhallinnan merkitys dataintegraatioissa	33
5	Yrityshaastattelut dataintegraatioiden käytöstä ja kokemuksista	34
5.1	Haastattelututkimuksen toteutus	34
5.2	Yritys A.....	34
5.3	Yritys B.....	36
5.4	Yritys C	38
5.5	Yritys D	40
5.6	Yritys E.....	42
5.7	Yritys F.....	44
6	Yrityksissä esiin tulleet dataintegraatioiden hyödyt ja haasteet	47
6.1	Yritysten esiin tuomat dataintegraatioiden hyödyt	47
6.2	Yritysten esiin tuomat dataintegraatioiden haasteet.....	47
7	Yrityksissä esiin tulleet parhaat käytänteet dataintegraatioissa.....	49
8	Dataintegraatiototeutusten vertailu	50
8.1	Dataintegraatiomallien ominaisuuksien vertailu.....	50
8.2	API-tyyppien ominaisuuksien vertailu	55
9	Ohjeita optimaalisen dataintegraatiototeutuksen valintaan.....	60
9.1	Yrityksen tarpeiden ja resurssien kartoittaminen	60
9.2	Dataintegraatiototeutuksen valinta.....	60
9.3	Yritysten kokemukset ja suositukset dataintegraatioon.....	62
10	Yhteenveto ja pohdinta	64
	Lähteet.....	66

Liite 1. Haastattelukirje

Liite 2. LAB_Opinnäytetyö_tietosuojailmoitus

Liite 3. Haastattelukysymykset

Liite 4. Opinnäytetyön_suostumuslomake

LYHENNELUETTELO

ACID	Atomisuus (engl. Atomicity), yhdenmukaisuus (engl. Consistency), erillisyys (engl. Isolation), kestävyys (engl. Durability) määrittelee periaatteet transaktioiden hallinnalle tietokannoissa.
AI	Artificial Intelligence, tekoäly on tekniikka ja tiede, jonka tavoitteena on luoda järjestelmiä, jotka kykenevät suorittamaan tehtäviä, jotka normaalisti vaativat ihmisen älykkyyttä.
AMQP	Advanced Message Queuing Protocol, viestijonojen protokolla, joka mahdollistaa monimutkaisten viestijärjestelmien luomisen.
API	Application Programming Interface, sovellusten ohjelmointirajapinta, jonka avulla eri ohjelmistot voivat keskustella keskenään.
ETL	Extract, Transform, Load.
ESB	Enterprise Service Bus.
Excel	Taulukkolaskentaohjelma, jota voidaan käyttää datan käsittelyyn ja analyysiin.
GraphQL	Application Programming Interface, kyselykieli API:lle, joka mahdollistaa tarkemman tiedonhaun.
GraphQL API	Ohjelmointirajapinta, joka noudattaa GraphQL-kyselykielen periaatteita ja mahdollistaa sovelluksien viestinvaihdon palvelimen kanssa käyttämällä GraphQL-kyselyitä.
gRPC	Avoimen lähdekoodin viestintäprotokolla ja työkalu, joka perustuu http/2-protokollaan ja joka käyttää Protocol Buffers -binäärimuotoista datansiirtoa mahdollistaen eri ohjelmistojen välisen tiedonvaihdon ja etäkutsujen tekemisen.
gRPC API	Ohjelmointirajapinta, joka noudattaa gRPC-protokollaa.
HTTP/s	HyperText Transfer Protocol/Secure, protokolla, jota käytetään verkkosivujen lataamiseen.

Hub-and-spoke	Datan jakeluun perustuva integraatio on integraatiomalli, jossa keskusyksikkö välittää tietoa useille kohdejärjestelmille.
IPA	Intelligent Process Automation on teknologia, joka yhdistää robottivusteisen ohjelmistoautomaation tekoälyyn ja koneoppimiseen parantaakseen automaation älykkyyttä ja tehokkuutta.
MQTT	Message Queuing Telemetry Transport, kevyt viestintäprotokolla, jota käytetään usein IoT (Internet of Things) -sovelluksissa.
RPA	Robotic Process Automation, robottivusteinen ohjelmistoautomaatio on teknologia, joka automatisoi rutiininomaisia tehtäviä ohjelmistorobottien avulla.
ROI	Return on Investment, sijoitetun pääoman tuotto, joka kertoo, kuinka kannattava sijoitus on ollut.
REST	Representational State Transfer, HTTP-protokollaan perustuva arkkitehtuurimalli ohjelmointirajapintojen toteuttamiseen.
REST API	Representational State Transfer Application Programming Interface on ohjelmointirajapinta, joka noudattaa REST-arkkitehtuurin periaatteita
Sitra	Suomen itsenäisyyden juhlarahasto on ajatushautomo ja rahasto, joka tutkii ja kehittää yhteiskunnallisia ilmiöitä.
SOAP	Simple Object Access Protocol, kieli ja viestintäprotokolla.
SOAP API	Simple Object Access Protocol Application Programming Interface on ohjelmointirajapinta, joka käyttää SOAP-protokollaa tiedonvaihtoon
SOCKETS	Rajapinta TCP- ja UDP-pohjaiseen tiedonsiirtoon, mikä mahdollistaa kahden ohjelmiston välisen reaaliaikaisen kommunikoinnin
TCO	Total Cost of Ownership, kokonaiskustannus, joka kuvaa tuotteen, palvelun tai järjestelmän hankinnan, käytön, ylläpidon ja poistamisen kustannuksia koko sen elinkaaren aikana.

WSDL	Web Service Description Language on XML-pohjainen kieli, jota käytetään web-palvelujen rajapintojen kuvaamiseen ja niiden käyttöliittymän määrittämiseen.
XML	EXtensible Markup Language, merkintäkieli, jota käytetään tietojen jäsentelemiseen ja vaihtamiseen.

1 Johdanto

Yritysten liiketoiminnan digitalisoituminen ja datan kasvu ovat johtaneet yhä suurempaan tarpeeseen hyödyntää dataa päätöksenteossa, innovaatioissa ja liiketoiminnan kehittämässä. Dataintegraatio, eli eri tietojärjestelmistä ja lähteistä peräisin olevan datan yhdistäminen, on välttämätöntä yrityksille, kun ne pyrkivät yhdistämään ja analysoimaan kasvavaa datamäärää monimutkaisissa toimintaympäristöissään. Samalla yhteistyön ja datan jakamisen arvo laajemmassa mittakaavassa tunnustetaan yhä enemmän. (Hurja Solutions 2022.) Dataekosysteemeissä toimijat tekevät yhteistyötä luodakseen arvoa ja synergiaa jakamalla dataa. Suomen itsenäisyyden juhlarahasto Sitran määritelmän mukaisesti jokaisen osallistujan tulee hyötyä yhteisestä dataekosysteemistä. (Sitra 2023a.) Koska dataekosysteemit tarjoavat potentiaalia liiketoiminnan kasvattamiseen, kilpailukyvyyn edistämiseen ja tehokkuuden lisäämiseen, yritysten tulisi harkita, miten ne voivat hyödyntää dataekosysteemejä liiketoimintansa kehittämässä (Hurja Solutions 2022).

Opinnäytetyön tavoitteena on tutkia dataintegraatioiden teknisiä ratkaisuja, jotka mahdollistavat dataekosysteemien toiminnan ja arvon luomisen yrityksille. Tekniset ratkaisut ovat keskeisessä roolissa, kun pyritään luomaan toimivia ja tehokkaita dataekosysteemejä. Työn päämääränä on antaa yrityksille kattava käsitys eri dataintegraatiomallien vaihtoehtoista, auttaa yrityksiä hahmottamaan eri vaihtoehtojen hyödyt ja haasteet sekä valitsemaan parhaiten omiin tarpeisiinsa ja resursseihinsa soveltuvat ratkaisut heidän tulevaisuuden liiketoiminnan vahvistamiseksi. Opinnäytetyö on toteutettu osana hanketta Kyberturvalliset ekosysteemit Etelä-Karjala. Hankkeen tavoitteena on edistää reiluun datanjakoon perustuvia ekosysteemejä ja vahvistaa pk-yritysten kykyä hyödyntää datataloutta.

Opinnäytetyön teoriaosuus perustuu kirjallisuuskatsaukseen erilaisista dataintegraatitoteutuksista. Empiirinen osuus on toteutettu haastattelututkimuksena, jossa anonymisoidut yritykset kertovat dataintegraatoratkaisuistaan. Lopuksi on tarjottu yrityksille suosituksia ja työkaluja tukemaan yrityksiä sopivan dataintegraatoratkaisun valinnassa.

Opinnäytetyön tutkimuskysymykset on rajattu seuraaviin:

- Millaisia dataintegraatoratkaisuja on olemassa ja miten erilaiset dataintegraatiomallit mahdollistavat tietojen liikkumisen dataekosysteemeissä?
- Millaisia mahdollisuuksia ja haasteita liittyy dataekosysteemeihin dataintegraatiomallien näkökulmasta?
- Miten eri dataintegraatiomallit soveltuvat erilaisten yritysten tarpeisiin ja liiketoimintamalleihin?

2 Dataintegraatio ja dataekosysteemi

2.1 Dataintegraation määritelmä

Dataintegraatio on prosessi, jossa eri tietolähteistä peräisin oleva data yhdistetään ja yhdenmukaistetaan mahdollistaen tietojen saumattoman siirron ja yhteistoimivuuden eri järjestelmien välillä. Tämä luo yhtenäisen ja luotettavan tietopohjan päätöksenteon, liiketoiminnan kehittämisen sekä eri järjestelmien välisen datan hyödyntämisen tueksi. (Heavy AI 2023.) Integraatioiden avulla voidaan helpottaa tiedon siirtoa, automatisoida tehtäviä, lisätä asiakasyymmärrystä ja tietoturvaa, sekä varmistaa, että sama tieto on käytettävissä koko organisaatiossa (Hurja Solutions 2022).

Dataintegraatio mahdollistaa myös yritysten välisen datan jakamisen dataekosysteemeissä, luoden verkostoon yhteistyötä, innovointia ja uutta liiketoimintaa. Yritysten välinen datan jakaminen dataintegraation avulla mahdollistaa yrityksille paremman tiedon hyödyntämisen ja päätöksen teon, asiakaspalvelun parantamisen sekä kilpailukyvyyn ylläpidon ja parantamisen. (Talend 2023a.)

Dataintegraatio voidaan toteuttaa eri tavoin riippuen tarpeista ja käytettävissä olevista resursseista. Erilaisia dataintegraatiomalleja ovat manuaalinen integraatio, robottiaivusteinen automaatio (engl. Robotic Process Automation, RPA), pistemäinen integraatio (engl. Point-to-point), datan jakeluun perustuva integraatio (engl. Hub-and-spoke), yrityksen laajuinen integraatio ja sovellusten ohjelmointirajapintaintegraatiot (API-integraatiot). (Business Technology Standard 2023.)

2.2 Dataekosysteemin määritelmä

Dataekosysteemillä tarkoitetaan useista dataverkoista koostuvaa verkostoa, jossa toimijat tekevät yhteistyötä jakamalla ja hyödyntämällä dataa (Sitra 2023a). Dataekosysteemiin voi kuulua useita yrityksiä ja organisaatioita, tietojärjestelmiä, prosesseja sekä dataa tuottavia ja käyttäviä henkilöitä. Dataekosysteemissä verkoston jäsenet ymmärtävät yhdessä tuotetun arvon ja jakavat sitä yhteisin pelisäännöin. (Malkamäki 2022.)

Dataekosysteemin tavoitteena on edistää innovointia, uutta liiketoimintaa ja yhteistyötä dataekosysteemiverkoston sisällä (Sitra 2023a). Sitran dataekosysteemin määritelmän mukaan tavoitteena on luoda reilu datatalous, joka edistää hyvinvointia, kilpailukykyä ja yhteiskuntien kehittymistä. Sitra korostaa, että reilussa datataloudessa pyritään luomaan arvoa ja hyötyä kaikille osapuolille: yksilöille, yrityksille ja yhteiskunnalle. (Malkamäki 2022.) Reilun datatalouden periaatteisiin kuuluu datan luotettavuus, saatavuus, ihmiskeskeisyys, arvонуonti sekä osaamisen ja datan jakaminen (Sitra 2023b).

3 Dataintegraatiomallit

3.1 Manuaalinen integraatio

Manuaalisessa integraatiossa tiedon siirto järjestelmästä toiseen ja datan käsittely suoritetaan manuaalisesti. Manuaalisia integraatioita käytetään tämän vuoksi yleensä pienimuotoisissa integraatioissa, joissa tietomäärä on vähäinen tai tiedon siirtotiheys on alhainen. (Business Technology Standard 2023.)

Manuaalinen integraatio on yksinkertainen toteuttaa. Manuaalisen integraation toteuttamisessa tarvitaan yrityksen henkilöresursseja, koska tiedon siirto tapahtuu käyttäjän toiminnan kautta, sekä jonkin verran tietoteknistä osaamista. Manuaalisessa integraatitoteutuksessa selkeät prosessit ja ohjeistukset tiedon siirtoon ovat tärkeitä tietojen eheyden ja luotettavuuden varmistamiseksi. (Talend 2023b.)

Manuaalisessa integraatiossa käyttäjän suorittamalla tiedon siirrolla voidaan tarkoittaa esimerkiksi tietojen lataamista ja lähettämistä, tiedon kopioimista ja liittämistä järjestelmästä toiseen tai jopa tietojen uudelleen kirjoittamista käsin. Yksi yleisimmistä manuaalisen datan integraation muodoista on Excel-tiedostojen hyödyntäminen. Data voidaan ladata käyttäjän toimesta Excel-tiedostona järjestelmästä, käsitellä Excelissä ja siirtää toiseen järjestelmään. Datan kopioiminen ja liittäminen järjestelmästä toiseen on hyvin yksinkertaisesti toteutettava manuaalisen datan integraation muoto eikä vaadi erityisen paljon teknistä osaamista. Myös manuaalinen tiedon syöttö esimerkiksi verkkolomakkeiden kautta voi olla osa manuaalista integraatiota. (Dearmer 2021.)

3.1.1 Dataintegraatiomallin hyödyt

Manuaalinen integraatio on yksinkertainen ja kustannustehokas vaihtoehto pienille yrityksille, joilla on vähän tietoja integroitavana ja joilla ei ole resursseja tai tarvetta automatisoiduille integraatoratkaisuille. Manuaalisen integraation toteutus ei vaadi käyttäjiltä syvällistä teknistä osaamista tai erityistä integraatitietämystä. (Dearmer 2021.)

Manuaalinen integraatio voi tarjota paremman kontrollin tietojen siirrosta, sillä käyttäjä voi tarkistaa ja validoida datan ennen sen siirtoa. Lisäksi tarvittaessa dataan voidaan tehdä korjauksia ja muutoksia, mikä ei olisi mahdollista automatisoidussa integraatioprosessissa. (Dearmer 2021.)

3.1.2 Dataintegraatiomallin haasteet

Manuaaliset integraatiot ovat aikaa vieviä, vaativat runsaasti henkilöstöresursseja ja ovat alttiita inhimillisille virheille, koska manuaaliset integraatiot suoritetaan ihmisten toimesta (Talend 2023b). Virheet voivat johtaa tietojen menetykseen tai virheelliseen tietoon. Manuaalisessa integraatiossa tietojen ylläpito on aikaa vievää ja haastavaa. Koska manuaalinen integraatio on hidasta, se ei välttämättä tarjoa reaaliaikaista tietoa. (Prevalova 2023.)

Manuaalisen integraation skaalaaminen suurille tietomäärille tai usein toistuville integraatioille voi olla hankalaa, koska se edellyttää manuaalista muutosta jokaisen integraation osalta. Manuaalisessa integraatiossa suuret tietomäärät voivat lisätä virheiden todennäköisyyttä ja tehdä virheiden havaitsemisesta ja korjaamisesta vaikeampaa. Jos tiedon ja integraatioiden määrä yrityksessä kasvaa, manuaaliset integraatiot voivat tulla yritykselle kalliiksi ja hitaiksi. Tällöin integraation toteuttamiseen tarvitaan lisää henkilöstöresursseja, mikä lisää kustannuksia ja voi heikentää tehokkuutta. (Prevalova 2023.)

3.1.3 Dataintegraatiomallin tietoturva ja tietosuoja

Manuaalisessa integraatiossa käyttäjien suorittamassa tiedon siirrossa avautuu mahdollisuus tietovuodoille, inhimillisille virheille ja väärinkäytöksille. Inhimillisen virheen riski korostuu manuaalisessa integraatiossa. Tietoa siirrettäessä manuaalisesti riski tiedon päätymisestä väriin käsiin kasvaa. Fyysisten tallennusvälineiden turvallinen säilytys ja hallinta on olennaista. (Rosencrance 2023). Säännöllisen koulutuksen ja tietoturvatietoisuuden lisäämisen avulla virheiden mahdollisuutta voidaan vähentää. (Khando 2021; Kost 2023).

Tietoturvassa ja tietosuojassa vaaditaan rajoitettua pääsyä sekä kaksivaiheista tunnistautumista. Arkaluonteisen tiedon siirrossa salaus on välttämätöntä. Henkilö- ja luottamuksellisiin tietoihin pääsyä tulee rajoittaa ja valvoa tietosuojan varmistamiseksi. Kaikista tietojen siirrosta ja muutoksista saatavilla olevien lokitietojen avulla ongelmia voidaan jäljittää. (Kost 2023; National Center for Education Statistics 2023.)

Säännölliset tietoturva- ja tietosuojatarkastukset ovat tärkeitä tietoturva ja tietosuojariskien ehkäisemiseksi. Näin voidaan varmistua siitä, että kaikki prosessit ja toimenpiteet ovat ajantasaisia ja tehokkaita. (Pratt 2022.)

3.1.4 Dataintegraatiomallin soveltuvuus

Manuaalinen integraatio voi olla sopiva pienille yrityksille pienimuotoisiin integraatioihin, joissa tietomäärä on vähäinen ja tiedon siirtotiheys alhainen. Se voi myös olla sopiva yrityksille, joilla on tiukat tietoturvan vaatimukset ja jotka tarvitsevat täyden kontrollin tietojen siirrosta. Manuaalinen integraatio voi hyvä vaihtoehto myös sellaisille pienille yrityksille, jotka ovat vasta aloittamassa dataintegraation hyödyntämistä, koska se ei vaadi suuria investointeja teknologioihin. (Dearmer 2021.)

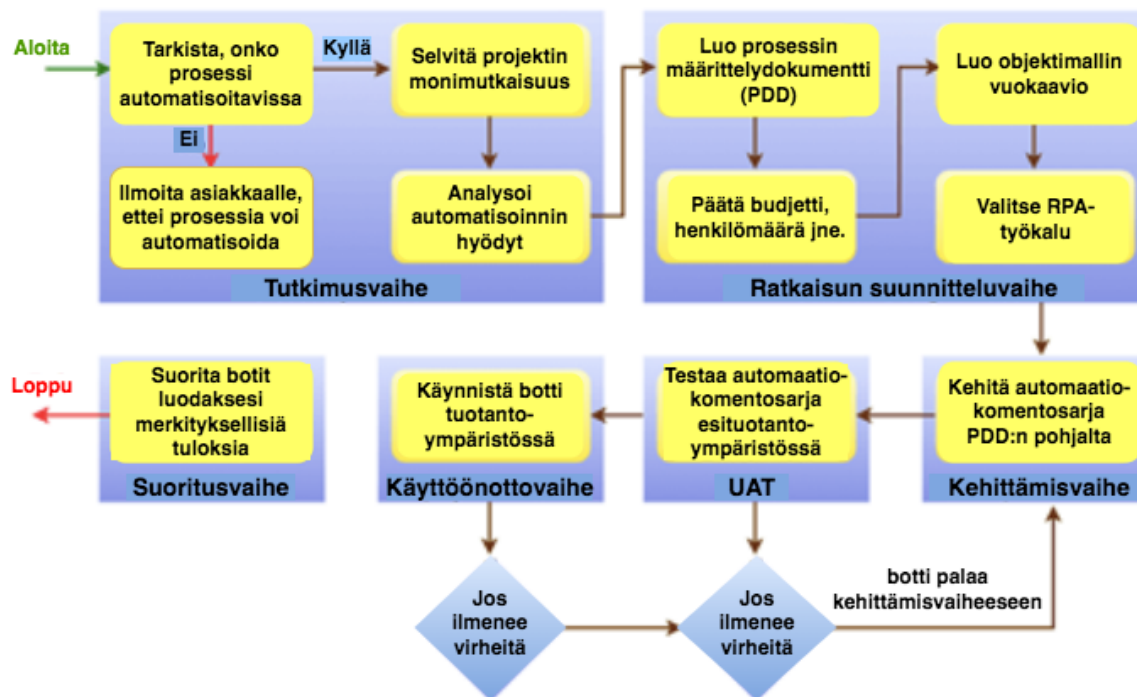
Manuaalinen integraatio ei sovellu suurille yrityksille, jotka tarvitsevat usein toistuvia tai suurten tietomäärien integraatioita. Se ei myöskään sovellu yrityksille, jotka luottavat reaaliaikaiseen tai lähes reaaliaikaiseen tietoon liiketoiminnan päätöksenteossa. Vaikka manuaalinen integraatio voi tarjota alhaisen kynnyksen pääsyn dataintegraatioon, se ei ole pitkän aikavälin ratkaisu kasvavalle yritykselle, kun datan määrä ja integraatioiden tarve lisääntyvät. (Prevalova 2023.)

3.2 Robottivusteinen ohjelmistoautomaatio

Robottivusteinen ohjelmistoautomaatio (Robotic Process Automation, RPA) on teknologia, jolla toistuvat, sääntöpohjaiset tehtävät, kuten datan syöttäminen, voidaan automatisoida. Robottivusteisen ohjelmistoautomaation avulla monimutkaiset integraatiotehtävät, jotka aiemmin vaativat manuaalista työtä, voidaan automatisoida. (Inscribe 2023.)

Robottivusteisessa ohjelmistoautomaatiossa robotteja eli botteja luodaan tehtävien, kuten datan syöttämisen, tiedostojen siirtämisen tai lomakkeiden täyttämisen, suorittamiseen. Botilla voidaan jäljitellä ihmisen toimia, kuten hiiren napsautuksia tai näppäimistön komentoja, ja se voi toimia useissa eri järjestelmissä ja sovelluksissa. (Lawton 2021.)

Robottivusteisen ohjelmistoautomaation toteutus on jaettu kuvion 3 mukaisesti tutkimusvaiheeseen (engl. Discovery Phase), ratkaisun suunnitteluvaiheeseen (engl. Solution Design Phase), kehittämissvaiheeseen (engl. Development Phase), käyttäjän hyväksymistestivaiheeseen (engl. User Acceptance Tests, UAT) sekä käyttöönotto- ja ylläpitovaiheeseen (engl. Deployment and Maintenance Phase), jonka osana on suoritusvaihe (engl. Execute Bots) (Javapoint 2023).



Kuvio 1. Robottivusteisen ohjelmistoautomaation elinkaari (mukailtu Javapoint 2023)

Robottivusteisen ohjelmistoautomaation toteutus alkaa tutkimusvaiheen (engl. Discovery Phase) prosessien analysoinnilla sekä määrittelyllä, mitkä prosesseista voidaan automatisoida. Lisäksi tutkimusvaiheessa arvioidaan prosessin monimutkaisuus ja hyödyt automatisoinnille. Prosessin pitää olla hyvin määriteltä, toistettava ja sääntöihin perustuva, jotta botti pystyy sen suorittamaan. (Javapoint 2023.)

Ratkaisun suunnitteluvaiheessa (engl. Solution Design Phase) laaditaan suunnitelma automatisoitavasta prosessista sekä päätetään esimerkiksi budjetista ja aikataulusta. Teknisen arkkitehdin ja prosessiarkkitehdin yhteistyönä tehdään prosessin määrittelydokumenttiasiakirja (engl. Process Definition Document, PDD), joka kuvaa yksityiskohtaisesti jokaisen vaiheen automatisointiprosessissa. (Javapoint 2023.)

Kun prosessi on valittu, botin kehittämisvaiheessa (engl. Development Phase) hyödynnetään RPA-työkaluja, jotka tarjoavat käyttöliittymän bottien ohjelmointiin ja hallintaan. (SS&C Blue Prism 2023.) Usein automatisoinnin toteutus ei vaadi perinteistä koodausta, vaan prosessin eri tehtävävaiheet voidaan "opettaa" boteille graafisen käyttöliittymän avulla (Javapoint 2023).

Kuvio 2 esittää kansainvälisen tutkimusyhtiö Gartnerin analyysin ohjelmistorobotiikan kilpailutilanteesta, minkä mukaan suosituimpia ohjelmistorobotiikan tarjoajia ovat muun muassa UiPath, Automation Anywhere, Blue Prism, WorkFusion ja Microsoft (Kirkwood 2020).

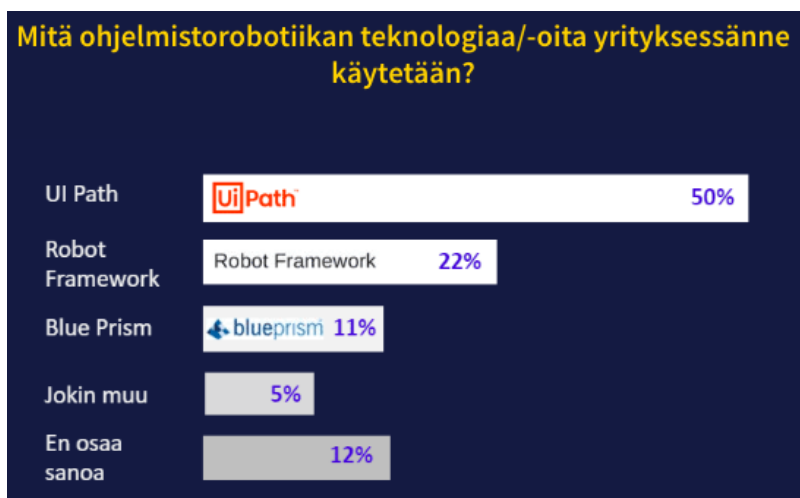
Magic Quadrant for Robotic Process Automation



Source: Gartner (July 2020)

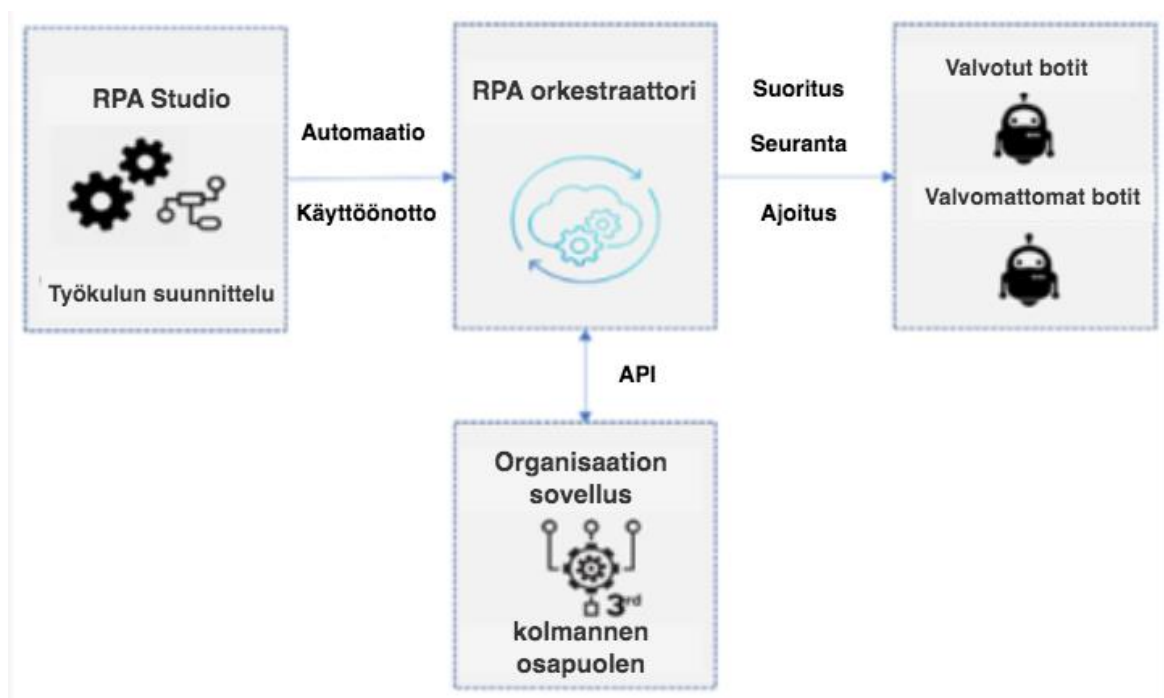
Kuvio 2. Gartnerin analyysi ohjelmistorobotiikan kilpailutilanteesta (Kirkwood 2020)

Kuvion 3 esittämän Midpointedin (2021) tutkimuksen mukaan suomalaisissa yrityksissä UiPath on selvästi suosituin RPA-teknologia, halliten 50 prosenttia markkinaosuudesta. Tämä vahva asema suomalaisissa yrityksissä kuvastaa Gartnerin analyysin maailmanlaajuisia analyysejä.



Kuvio 3. RPA-integraatioiden käyttö suomalaisissa yrityksissä (Midpointed 2021.)

Robottiaivusteisen automaation tekninen toteutus voi vaihdella eri RPA-työkalujen ja -alustojen välillä. Choi ym. (2021) antavat kuviossa 4 yleiskuvan siitä, miten monet RPA-ratkaisut on rakennettu ja mitkä ovat niiden keskeiset komponentit.



Kuvio 4. Robottiaivusteisen ohjelmistoautomaation arkkitehtuuri ja komponentit (mukailtu Choi ym. 2021)

RPA Studio (engl. myös Designer) on graafinen käyttöliittymä, jonka avulla automaatioita voidaan rakentaa, testata ja etsiä niistä virheitä. RPA Studio tarjoaa työkaluja, joiden avulla komponentteja voidaan helposti raahata ja pudottaa paikoilleen automaation luomiseksi. RPA Studion tai Designerin avulla luodaan työnkulkusuunnitelmia (engl. Workflow Design) eli prosesseja, jotka suoritetaan tietyssä järjestyksessä. RPA:ssa työnkulun suunnittelu on keskeinen osa automaatioprosessia. RPA orkestraattori (engl. Orchestrator) on palvelinperusteinen yksikkö, joka hallitsee, valvoo ja ajoittaa robottien suorituksia. (Choi ym. 2021.)

Robottiohjelmat (engl. bots) ovat automaatiotehtäviä suorittavia yksiköitä, joita on usein kolmea eri tyyppiä. Valvotut botit (engl. attended bots) suorittavat tehtäviä käyttöliittymän kautta simuloimalla käyttäjän toimintoja, kuten klikkaamista ja tiedon syöttämistä. Valvotut botit toimivat ihmisen ohjauksessa tai valvonnassa, ja ne on suunniteltu toimimaan yhdessä käyttäjän kanssa. Valvomattomat botit (engl. unattended bots) toimivat itsenäisesti ilman ihmisen väliintuloa, ja integroituvat suoraan järjestelmiin ja palveluihin API:en kautta. Hybridibotit (engl. hybrid bots) yhdistävät sekä valvottujen bottien frontendin että valvomattomien bottien backendin toiminnot. (Choi ym. 2021; Raju 2021.)

Käyttäjän hyväksymistestivaiheessa (engl. User Acceptance Tests, UAT) testataan bottien toimintaa, että botit toimivat odotetulla tavalla. Jos testaus epäonnistuu, botit palautuvat kehitysvaiheeseen, jossa havaitut virheet korjataan ennen uutta testausta. (Javapoint 2023.)

Käyttöönotto- ja ylläpitovaiheessa (engl. Deployment and Maintenance Phase) botit otetaan käyttöön tuotantoympäristössä. Tällöin niiden suorituskykyä on tärkeää seurata ja kehitys- ja testausvaiheesta tulee tehdä tarvittavat säädöt niiden toiminnan parantamiseksi. Bottien käynnistämistä ja niiden suorituskyvyn tarkistamista kutsutaan suoritusvaiheeksi (engl. Execute Bots), joka on osa käyttöönotto- ja ylläpitovaihetta. (Javapoint 2023.)

Nykyisin robottivusteisen ohjelmistoautomaation mahdollisuuksia laajennetaan yhä enemmän yhdistämällä se tekoälyyn ja koneoppimiseen parempien tulosten saavuttamiseksi. Tästä käytetään nimitystä älykäs ohjelmistoautomaatio, IPA (Intelligent Process Automation). Tekoälyn ja koneoppimisen avulla RPA-botit voivat oppia ja tehdä itsenäisiä päätöksiä, eivätkä vain toistaa ennalta määriteltyjä tehtäviä. (Blueprint 2021.)

3.2.1 Dataintegraatiomallin hyödyt

Robottivusteinen ohjelmistoautomaatio vähentää manuaalisen työn tarvetta ja parantaa liiketoiminnan tuottavuutta ja tehokkuutta. Robotteja käyttäen tehtävät suoritetaan nopeammin kuin ihmiset niitä suorittaisivat, ja robotit voivat toimia ympäri vuorokauden ilman taukoja tai sairaspäiviä. Robotteja käyttäen tehtävät suoritetaan myös tarkemmin kuin ihmiset niitä suorittaisivat. Robottivusteinen ohjelmistoautomaatio vähentää inhimillisten virheiden määrää, mikä tekee prosesseista luotettavampia. Näistä syistä johtuen robottivusteinen ohjelmistoautomaatio tuo yritykselle merkittäviä kustannussäästöjä. Vaikka RPA-järjestelmän käyttöönottoon liittyy alkukustannuksia ja ylläpitokuluja, nämä ovat usein huomattavasti pienempiä kuin vastaavan määrän ihmistyövoiman palkkaaminen. (SS&C Blue Prism 2023.)

Robottivusteisen ohjelmistoautomaation ansiosta henkilöstöresursseja vapautuu muihin, korkeamman arvon tuottaviin tehtäviin, kun rutiinimaiset työt siirtyvät roboteille. Tämän on todettu lisäävän työntekijöiden tyytyväisyyttä ja sitoutumista. (Kirkwood 2019). Robottivusteinen ohjelmistoautomaatio parantaa myös asiakastytyväisyyttä nopeuttamalla prosesseja, parantamalla asiakaspalvelua ja lyhentämällä vastausaikoja (SS&C Blue Prism 2023, PRP Business Solutions 2023). Lisäksi robottivusteisen ohjelmistoautomaation etuna on sen joustavuus, sillä se soveltuu monenlaisiin sovelluksiin ja ympäristöihin (PRP Business Solutions 2023).

Kuvion 5 esittämän Midpointedin (2021) analyysin mukaan RPA:n tärkeimpinä etuina yritykset pitävät manuaalisten rutiinitöiden siirtymistä roboteille, virheiden vähenemistä, syötetyn tiedon laadun parantumista ja mahdollisuutta prosessien jatkuvaan operointiin kellon ympäri, riippumatta ihmisten työajoista. Yritykset katsoivat keskeisiksi hyödyiksi myös mahdollisuuden kasvattaa työntekijöiden tyytyväisyyttä ja säästää henkilökustannuksissa. (Midpointed 2021.)



Kuvio 5. RPA:lla tavoitellut hyödyt suomalaisissa yrityksissä (Midpointed 2021)

3.2.2 Dataintegraatiomallin haasteet

Robottiaivusteisen ohjelmistoautomaation käyttöönotto voi aiheuttaa pelkoa työpaikkojen menettämisestä. Tämän vastustuksen lieventämiseksi on tärkeää viestiä selkeästi RPA:n eduista myös työntekijöille ja siitä, miten se voi lisätä työtyytyväisyyttä. (SS&C Blue Prism 2023.)

Robottiaivusteisen ohjelmistoautomaation käyttö vaatii merkittäviä investointeja aikaan ja resursseihin. Botit vaativat myös jatkuvaa ylläpitoa, seuranta ja valvontaa. Mahdolliset viat tai ongelmat voivat pysäyttää botin tai aiheuttaa virheitä, jotka vaikuttavat liiketoimintaprosesseihin. (Blueprint 2020.)

Koska robotit käsittelevät usein arkaluonteista tietoa, on tärkeää varmistaa niiden noudattavan tietoturvastandardeja, ja että niiden pääsy on rajoitettu vain tarvittaviin järjestelmiin ja tietoihin. Lisäksi on oltava valmiudet reagoida mahdollisiin tietoturvaloukkauksiin tai muihin ongelmiin. (Lawton 2021.)

Automatisoitavaksi valittavien prosessien on myös tärkeää tulla tunnistetuiksi, koska robottivusteisessa ohjelmistoautomaatiossa kaikki prosessit eivät sovellu automatisointiin. Robottivusteisen automatisoinnin onnistuminen riippuu paljon siitä, kuinka hyvin automatisoitavat prosessit valitaan. (SS&C Blue Prism 2023.)

3.2.3 Dataintegraatiomallin tietoturva ja tietosuoja

Suurimmat tietoturvariskit robottivusteisessa ohjelmistoautomaatiossa liittyvät tietovuotoihin ja petoksiin. Tietoturvan ja tietosuojan edistämiseksi jokaiselle RPA-botille tulisi antaa yksilöllinen tunniste. Kaksivaiheisen tunnistautumisen käyttö käyttäjätunnuksen ja salasanan yhteydessä auttaa torjumaan tietoturva- ja tietosuojauhkia.

Robottivusteisen ohjelmistoautomaation käyttö voi lisätä tilioikeuksia, mikä lisää tietoturva- ja tietosuojauhan riskiä. Kuten manuaalisessa tietojen siirrossa tai integraatiossa, on tärkeää rajoittaa ihmisten käyttöoikeuksia. RPA-boteille myönnettävät käyttöoikeudet tulisi rajoittaa vain niihin tehtäviin ja tietoihin, joita ne tarvitsevat määritetyn tehtävän suorittamiseksi. Lisäksi on suositeltavaa käyttää istunnonhallintatyökaluja petoksen estämiseksi.

Yritysten tulisi varmistaa, että RPA-työkalu tarjoaa kattavat ja aukottomat lokit kaikesta tietojen siirrosta ja muutoksista. Näin mahdollisten tietoturva- tai tietosuojaloukkausten sattuessa tietoturvtiimi voi tarkastella tapahtumia tarkemmin. (Sakpal 2021.)

Robottivusteisen ohjelmistoautomaation kehittäminen on jatkuva prosessi, ja tietoturva- ja tietosuoja-asiat on huomioitava jo kehitysvaiheessa yhteistyössä tietoturvtiimin kanssa. Robottivusteinen ohjelmistoautomaatio edellyttää myös jatkuvaa seuranta- ja ohjelmakoodien tarkistusta. (Sakpal 2021.)

3.2.4 Dataintegraatiomallin soveltuvuus

Robottivusteinen ohjelmistoautomaatio soveltuu erityisen hyvin yrityksille, joilla on paljon rutiininomaisia, sääntöpohjaisia toistuvia tehtäviä, jotka voidaan suorittaa tietyllä, ennalta määritellyllä tavalla. Robottivusteinen automaatio toimii parhaiten, kun se käsittelee selkeästi järjestettyä dataa. Robottivusteinen ohjelmistoautomaatio on hyvä ratkaisu myös vanhojen järjestelmien välisten integraatioiden hoitamiseen, jotka eivät tue moderneja integraatiotekniikoita. (Laserfiche 2023.)

Robottivusteinen ohjelmistoautomaatio ei sovellu tehtäville tai prosesseille, jotka vaativat suurta ihmisen harkintakykyä, luovuutta tai ovat erittäin monimutkaisia ja epäsäännöllisesti toistuvia. Harvoin tapahtuvissa ja epäsäännöllisissä tehtävissä robottivusteisen automaation hyöty voi jäädä vähäiseksi. Myös rakenteettoman datan automatisointi voi olla haastavaa tai jopa mahdotonta. (SS&C Blue Prism 2023.)

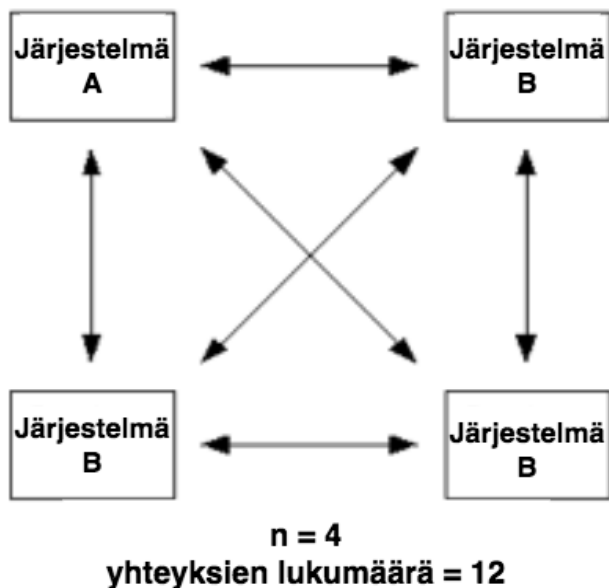
Midpointedin (2021) tutkimustulosten mukaan RPA on saavuttanut merkittävän aseman suomalaisten yritysten toiminnassa, ja noin 40 prosenttia vastaajista ilmoitti käyttävänsä tätä teknologiaa. Kiinnostus RPA:ta kohtaan on myös suurta niissä yrityksissä, joissa sitä ei vielä ole otettu käyttöön, ja peräti 38 prosenttia vastaajista piti sitä erittäin tai melko kiinnostavana. (Midpointed 2021.)

3.3 Pistemäinen integraatio (Point-to-point)

Pistemäinen integraatio (engl. Point-to-point) on yksinkertainen, edullinen ja ketterä vaihtoehto pienimuotoisissa integraatioissa. Pistemäisessä integraatiossa kaksi järjestelmää on yhdistetty suoraan toisiinsa ilman välittäjää tai välikäsiä, ja viesti siirretään lähettäjältä yhdelle vastaanottajalle. (Pearson Education 2022). Vaikka pistemäinen integraatio on yksi vanhimmista integraatiomuodoista, se on edelleen yleinen ratkaisu yksinkertaisissa ympäristöissä, joissa on vain muutama järjestelmä. Kuitenkin sen rajoitukset tulevat esille, kun järjestelmien määrä ja monimutkaisuus kasvavat. (GreenVulcano Technologies 2019.)

Pistemäisessä integraatiossa yhteys kahden järjestelmän välille voidaan toteuttaa räätälöidyn koodauksen, API:ien tai näiden yhdistelmän avulla (Kouzmanov 2015). Räätälöity koodaus vaatii syvällistä ymmärrystä molemmista järjestelmistä ja niiden teknologioista. API:t (Application Programming Interfaces) ovat toinen yleinen tapa toteuttaa pistemäinen integraatio. API:t ovat joukko protokollia, jotka mahdollistavat kahden eri ohjelmistojen välisen kommunikaation. Lisäksi pistemäinen integraatio voi tapahtua myös API:ien ja räätälöidyn koodauksen yhdistelmänä, mikä tarjoaa joustavuutta ja mahdollistaa monipuolisemman kommunikaation eri järjestelmien välillä. Esimerkkinä tästä voi olla kolmannen osapuolen palvelun, kuten Google Mapsin API:n, käyttö verkkosivujen karttatietojen integroimiseen. (SnapLogic 2023.)

Pistemäisen integraation toteutuksessa käytetään usein jonoja ja integraatiovälittäjiä. Viestit sijoitetaan jonoon, ja integraatiovälittäjä ohjaa viestit oikeille vastaanottajille, jotka käsittelevät ne. Järjestelmien määrän kasvaessa, yhteyksien määrä kasvaa eksponentiaalisesti $n(n-1)$ -sääntönä. Kuvio 6 havainnollistaa tätä sääntöä ja pistemäistä integraatiota neljän eri järjestelmän välillä. (Pearson Education 2022.)



Kuvio 6. Pistemäisen integraation $n(n-1)$ -sääntö (mukailtu Pearson Education 2022)

3.3.1 Dataintegraatiomallin hyödyt

Pistemäinen integraatio tarjoaa useita etuja erityisesti pienimuotoisissa integraatioissa. Se on helppo toteuttaa ja nopea ottaa käyttöön. Pienimuotoisissa integraatioissa myös ylläpito on vaivatonta, koska se ei vaadi lisäinfrastruktuuria tai ohjelmistoja. Tämä tekee pistemäisestä integraatiosta kustannustehokkaan vaihtoehdon pienissä mittakaavoissa. (SnapLogic 2023.)

Lisäksi pistemäinen integraatio mahdollistaa joustavuuden, koska jokainen integraatio voidaan räätälöidä tarkasti sen tarpeisiin. Tämä voi olla hyödyllistä tilanteissa, joissa tarvitaan nopeaa ja yksinkertaista tapaa yhdistää kaksi järjestelmää. (SnapLogic 2023.)

3.3.2 Dataintegraatiomallin haasteet

Pistemäisten integraatioiden kasvaessa ajan myötä, lyhyen aikavälin edut voivat muuttua pitkän aikavälin haasteiksi. Kasvava järjestelmien määrä ja niiden keskinäinen yhdistäminen lisäävät nopeasti integraatioiden monimutkaisuutta ja sekavuutta (Sythoff 2022; SnapLogic 2023). Pistemäisten integraatioiden monimutkaistuessa integraatioista tulee myös vaikeasti ylläpidettäviä (SnapLogic 2023).

Lisäksi yksittäiset integraatiot voivat olla riippuvaisia tietyn kehittäjän osaamisesta, mikä voi aiheuttaa ongelmia ylläpidossa ja kehityksessä. Tämä voi tehdä yrityksen haavoittuvaksi, jos avainhenkilöitä lähtee tai vaihtuu. Pistemäinen integraatio voi johtaa myös tietovirtojen

hajanaisuuteen ja tiedonhallinnan haasteisiin, jos käytetään useita erillisiä integraatioita eri osissa organisaatiota. (Kouzmanov 2015.)

3.3.3 Dataintegraatiomallin tietoturva ja tietosuoja

Pistemäisessä integraatiossa keskitetyn tietoturvahallinnan toteuttaminen voi olla haastavaa, koska järjestelmiä yhdistetään suoraan. Tämä voi johtaa turvallisuusriskeihin, kun integraatiot laajenevat ja monimutkaistuvat. (Kouzmanov 2015.)

Päivitysprosessissa on otettava huomioon, että uusien korjauspäivitysten yhteensopivuus vanhojen yhteyksien kanssa voi tuoda haasteita. Tämä voi hidastaa päivityksiä tai estää ne kokonaan joissakin tilanteissa. Turvallisuusriskien vähentämiseksi päivitykset ja integraatiot on suunniteltava huolellisesti. (Sythoff 2022).

3.3.4 Dataintegraatiomallin soveltuvuus

Pistemäinen integraatio soveltuu ainoastaan pienimuotoisiin integraatioihin. Se on hyvä valinta tilanteissa, joissa tarvitaan nopea, edullinen ja yksinkertainen yhteys kahden sovelluksen välille, kuten valmiiden kolmannen osapuolen ominaisuuksien lisäämiseen. (SnapLogic 2023.)

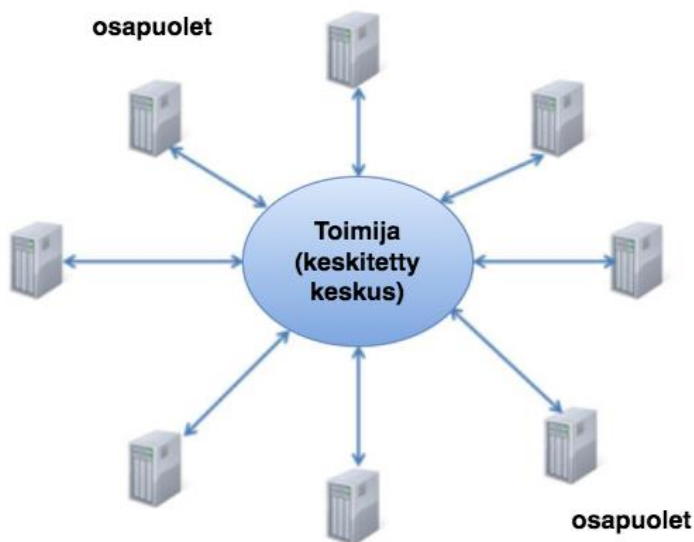
Yrityksen tarpeiden kasvaessa ja integraatioiden määrän lisääntyessä, pistemäinen integraatio voi kuitenkin johtaa liian monimutkaisiin ja vaikeasti hallittaviin integraatioihin. Tällöin kannattaa harkita muita integraatiomalleja, kuten datan jakeluun perustuvaa integraatiototeutusta (engl. Hub and Spoke) tai yrityksen laajuisia integraatiototeutuksia, jotka tarjoavat paremman skaalautuvuuden ja hallittavuuden suuremmissa ympäristöissä. (Kouzmanov 2015; SnapLogic 2023.) Pistemäinen integraatio ei siis sovellu nopeasti kasvaville yrityksille tai yrityksille, jotka tarvitsevat laajamittaisia integraatioita. Se ei myöskään sovellu yrityksille, joiden liiketoimintaympäristö muuttuu nopeasti. (Sythoff 2022.)

3.4 Datan jakeluun perustuva integraatio (Hub-and-spoke)

Datan jakeluun perustuva integraatio poikkeaa pistemäisestä integraatiomallista siten, että jokainen järjestelmä ei ole suoraan yhteydessä muihin järjestelmiin, vaan kommunikoi keskitetyn toimijan kautta. Datan jakeluun perustuva integraatiomalli tarjoaa selkeän ja keskitetyn tavan hallita ja valvoa tietovirtoja eri järjestelmien välillä.

Datan jakeluun perustuvassa integraatiossa (engl. hub-and-spoke) tieto ei liiku suoraan yhdestä pisteestä toiseen, vaan se kulkee toimijan eli keskitetyn keskuksen (engl. hub tai message broker) kautta. Tämä vähentää reittien määrää ja mahdollistaa tietojen ryhmittelyn

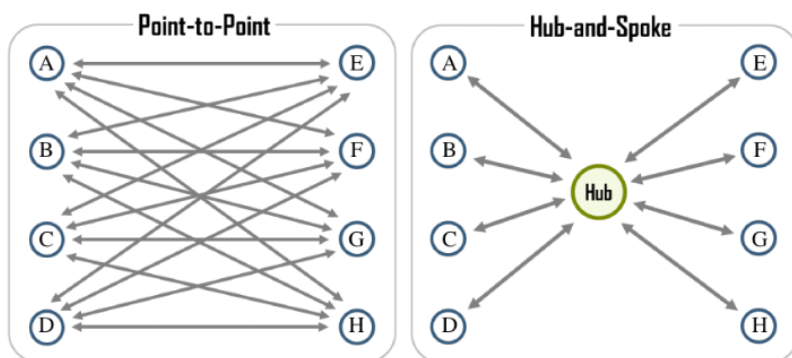
ja tehokkaan jakelun. Kuvio 7 ilmenee, kuinka toimija toimii keskusyksikkönä, johon kaikki muut järjestelmät, eli osapuolet (engl. spoke), yhdistyvät. Kuviossa tiedot lähtevät liikkeelle toimijalta, josta ne jaetaan edelleen osapuolille. Näissä tiedot käsitellään ja jaetaan edelleen. Osapuolet voivat olla esimerkiksi sovelluksia, tietokantoja tai järjestelmiä, joilla on omat tehtävänsä. (Santos 2016.)



Kuvio 7. Toimija ja osapuolet (mukailtu Santos 2016)

3.4.1 Dataintegraatiomallin hyödyt

Datan jakeluun perustuvassa arkkitehtuurissa on useita etuja verrattuna pisteestä pisteeseen integraatioon. Integraation laajentaminen on helpompaa, koska uusien sovellusten lisääminen ei vaikuta muihin sovelluksiin. Jokainen uusi sovellus voidaan yhdistää toimijaan vain yhdellä integraatiolla ilman vaikutusta muihin osapuoliin. Kuvio 8 osoittaa, että integraatioiden määrä on näin ollen vähemmän kuin suorassa pisteestä pisteeseen -integraatiossa, mikä vähentää kustannuksia. (Santos 2016).



Kuvio 8. Pistemäinen integraatio vs datan jakeluun perustuva integraatio (Soleymanifar 2018)

Datan synkronointi ja jakelu tapahtuvat tehokkaasti toimijan kautta, mikä vähentää tiedon siirtämisen kustannuksia ja optimoi tietoreittien suunnittelua (Santos 2016; Habas 2019). Datan jakeluun perustuvan integraation etuna verrattuna pistemäiseen integraatioon on myös parempi tietoturva keskitetyn toimijan ansiosta, ja se soveltuu monenlaisiin liiketoimintatarpeisiin (Santos 2016; Giuliani 2020).

3.4.2 Dataintegraatiomallin haasteet

Kun dataa kuljetetaan toimijan kautta, voi aiheutua pieni viive tietojen liikkumisessa järjestelmien välillä verrattuna suoraan pistemäiseen integraatioon. Yhteensopivuuden varmistaminen datan jakeluun perustuvassa integraatiossa voi olla haasteena erityisesti vanhempien järjestelmien kanssa, koska kaikkien osapuolten toimesta on noudatettava yhteisiä standardeja ja protokollia, jotta integraatio toimii. (Tähtinen 2015)

Koska kaikki integraatiot kulkevat toimijan kautta, jos toimijan toimesta ei pystytä käsittelemään suurta viestimäärää raskaan kuormituksen aikana tai toimija ei ole toiminnassa esimerkiksi teknisten ongelmien vuoksi, se voi muodostua integraation pullonkaulaksi (Santos 2016). Tällaiset tilanteet voivat vaikuttaa kaikkiin integraatioihin ja aiheuttaa häiriöitä tietojen liikkuvuudessa ja jakelussa eri sovellusten välillä. Tämä voi heijastua organisaation toimintaan. (Tähtinen 2015)

3.4.3 Dataintegraatiomallin tietoturva ja tietosuoja

Datan jakeluun perustuva integraatio tarjoaa paremmat mahdollisuudet tietoturvan hallintaan kuin pistemäinen tai manuaalinen integraatio. Keskitetyn toimijan hallinnan ansiosta on helpompaa määrittää, mitä tietoa ja kenelle jaetaan, mikä mahdollistaa turvallisen ja luotettavan tiedon jakamisen eri sovellusten välillä. (Giuliani 2020.)

Toisaalta, vaikka toimija mahdollistaa keskitetyn pääsynhallinnan, se voi myös tuoda haasteita, jos eri sovellusten välillä tarvitaan erilaisia pääsynhallintasääntöjä. Tämä voi vaikuttaa integraation monimutkaisuuteen ja edellyttää tarkkaa hienosäätöä. Koska datan jakelu kulkee toimijan kautta, toimijasta tulee yksi piste, joka voi altistua tietoturvariskeille. Jos toimija altistuu hyökkäykselle tai tietomurrolle, se voi vaikuttaa koko integraatoratkaisun turvallisuuteen. (Giuliani 2020.)

3.4.4 Dataintegraatiomallin soveltuvuus

Datan jakeluun perustuva integraatitoteutus soveltuu moniin teollisuudenaloihin, kuten ilmaliikenne, logistiikka ja vähittäiskauppa. Erityisesti yritykset, jotka jatkuvasti kehittävät ja päivittävät järjestelmiään vastatakseen muuttuviin liiketoiminnan tarpeisiin, voivat hyötyä

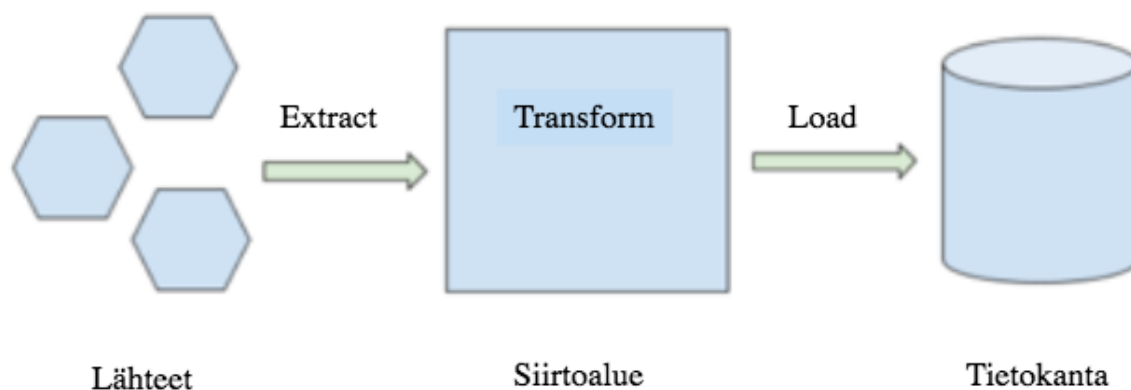
datan jakeluun perustuvasta integraatiosta, koska se mahdollistaa uusien sovellusten lisäämisen ja niiden pitämisen synkronoituna ilman vaikutusta muihin järjestelmiin. (Habas 2019.)

Datan jakeluun perustuva integraatio soveltuu hyvin myös tilanteisiin, joissa tarvitaan tehokasta tiedon jakelua ja synkronointia eri sovellusten välillä, koska integraatiomalli mahdollistaa tehokkaan synkronoinnin ja jakelun keskitetyn toimijan kautta. Lisäksi datan jakeluun perustuva integraatio voi olla hyvä valinta yrityksille, jotka haluavat keskitetyn tietoturvan hallinnan, jonka avulla voidaan helposti määrittää, kuka saa pääsyn mihinkin tietoihin. (Habas 2019.)

3.5 Yrityksen laajuinen integraatio

Yrityksen laajuinen integraatio toteutetaan integraatioalustalla, joka mahdollistaa eri sovellusten, tietolähteiden ja toimintojen saumattoman yhdistämisen ja tiedonvaihdon. Yrityksen laajuisen integraation toteutuksessa kaksi yleistä mallia ovat Extract-Transform-Load (ETL) ja Enterprise Service Bus (ESB). (Business Technology Standard 2023.)

ETL-prosessi koostuu kuvion 8 mukaisesti datan keräämisestä eri lähteistä (extract), sen puhdistamisesta ja muokkaamisesta (transform), kuten puuttuvien arvojen poistamisesta tai datan uudelleenjärjestelystä, sekä lopulta datan tallentamisesta tietovarastoon (load), kuten data warehouseen tai data martiin (Keebola 2022). Data warehouse on kattava ja laaja-alainen tietovarasto, kun taas data mart on pienempi ja erikoistuneempi tietovarasto, joka keskittyy tiettyyn tarpeeseen tai käyttötarkoitukseen (Tobin 2020). Tietovarastossa data on valmis analysoitavaksi ja hyödynnettäväksi.



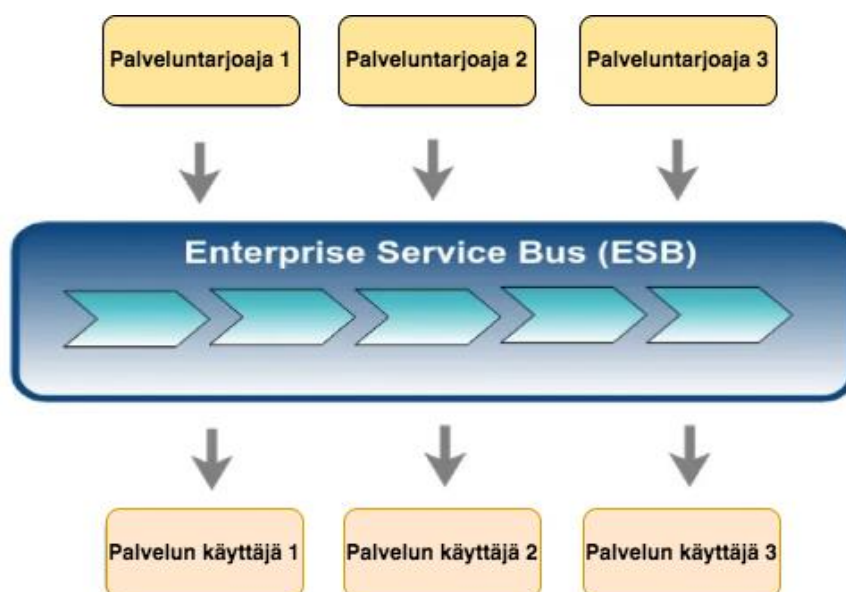
Kuvio 9. ETL-arkkitehtuuri (mukailtu Keebola 2022)

ETL-arkkitehtuurissa data haetaan aktiivisesti joko määritellyn aikataulun mukaisesti tai vastauksena tiettyyn kysyntään (Smallcombe 2020). Lähestymistapaa, jossa data haetaan

aktiivisesti lähteistä joka kerta ja siirretään prosessointiin, kutsutaan täysi lähestymistavaksi (engl. full-extraction). Vaiheittaisessa lähestymistavassa (engl. incremental extraction) uusi data kerätään joka kierroksella yhdessä edellisestä kierroksesta muuttuneen datan kanssa. Lähteeseen perustuvassa lähestymistavassa (engl. source-driven extraction) lähdejärjestelmä ilmoittaa muutoksista, ja prosessi suoritetaan näiden muutosten perusteella. (Keebola 2022.)

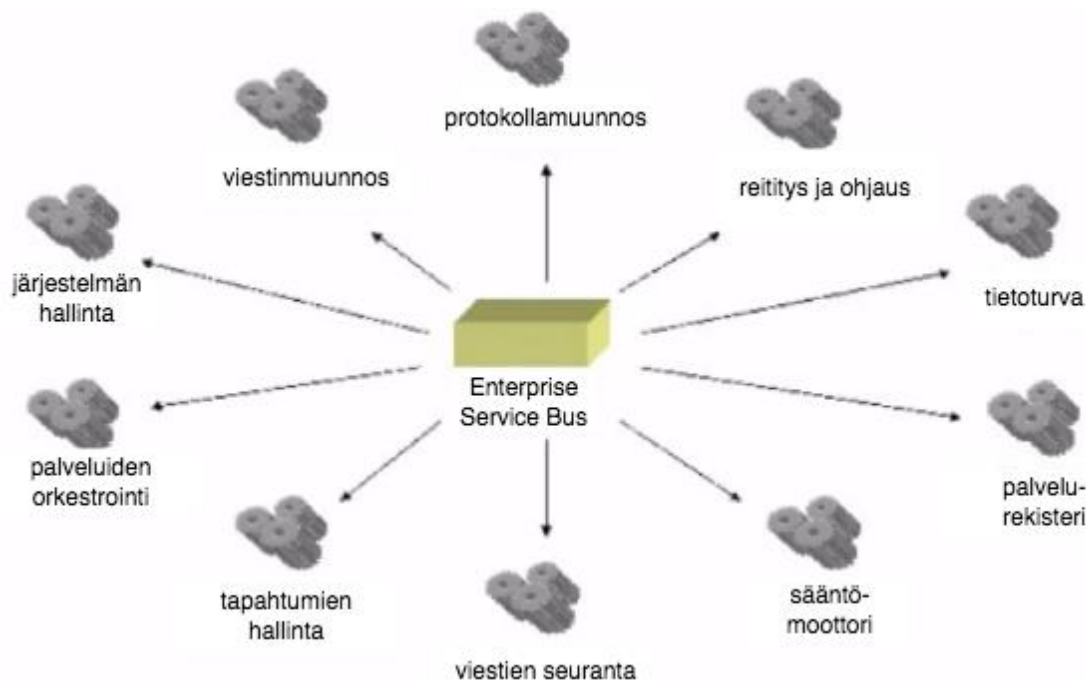
ELT eli Extract-Load-Transform on rinnakkainen integraatiomalli ETL-mallille. ELT-arkkitehtuurissa raakadata kerätään ensin tietovarastoon (engl. data warehouse) tai datajärveen (engl. data lake). Tämän jälkeen datan muokkaus ja jalostaminen tapahtuvat joustavasti vasta tarpeen mukaan ennen datan analysointia. (Keebola 2022.)

Enterprise Service Bus käyttää viestiväylää (engl. Message Bus) sovellusten välisen tiedon vaihtamiseen. Viestiväylä toimii keskuksena, joka ohjaa viestien kulkua eri komponenttien välillä. ESB on valmiudessa ja tietokantapalvelimilta seurataan tietovirtaa ja vastaanotetaan tietoa, minkä jälkeen saapuvat tehtävät priorisoidaan ja osoitetaan oikeille prosesseille tai henkilöille. Jos useita tehtäviä saapuu yhtä aikaa, ESB pystyy määrittämään, mitkä tehtävät käsitellään ensin ja mitkä odottavat vuoroaan. Tämä mahdollistaa tehokkaan ja reaaliaikaisen tiedonkäsittelyn sekä tapahtumavetoisen toiminnan. (Smallcombe 2020.) Viestiväylän avulla sovelluksille mahdollistetaan kommunikointi toistensa kanssa ilman riippuvuutta tai tietämättömyyttä muista järjestelmistä viestiväylällä. Kuvio 10 kuvastaa ESB:n arkkitehtuuria, joka perustuu palvelukeskeiseen arkkitehtuuriin (SOA). Palvelukeskeisessä arkkitehtuurissa sovellukset voivat tarjota ja kuluttaa palveluita toisiltaan. (Ariyarathne 2019.)



Kuvio 10. ESB-arkkitehtuuri (mukailtu Ariyarathne 2019)

Kuvio 11 esittää ESB:n keskeiset toiminnallisuudet, jotka kattavat viestinmuunnoksen (engl. message transformation), reitityksen (engl. routing), ohjauksen (engl. mediation), protokollamuunnos (engl. protocol transformation), tietoturvan (engl. security), tapahtumienhallinnan (engl. transaction management tai event management), palvelurekisterin (engl. service registry), palveluiden orkestrointi (engl. service orchestration), sääntömoottorin (engl. rules engine), viestien seurannan (engl. message monitoring) ja järjestelmänhallinnan (engl. system management) (Richards 2006).



Kuvio 11. ESB:n keskeiset toiminnallisuudet (mukailtu Richards 2006)

ESB-arkkitehtuurissa viestinmuunnos-komponentilla vastataan viestien rakenteen, formaatin tai protokollan muuntamisesta, jotta eri järjestelmät voivat käsitellä viestejä oikein. Protokollamuunnoksella tarkoitetaan kykyä muuntaa viestien protokollat toisiin, jotta eri järjestelmien väliset kommunikaatiot voivat tapahtua. Reitityksellä ja ohjauksella määritellään, miten viestit ohjataan lähteestä kohdesovellukselle. Tietoturva-komponentilla vastataan ESB:n tietoturvaominaisuuksista, kuten viestien salaamisesta, käyttöoikeuksien hallinnasta ja käyttäjän tunnistautumisesta varmistaakseen, että tiedonsiirto on suojattua. Palvelurekisterissä pidetään kirjaa käytettävissä olevista palveluista, niiden sijainnista ja rajapinnoista. Siinä autetaan sovelluksia löytämään tarvittavat palvelut reaaliaikaisesti ja automaattisesti. Sääntömoottorilla mahdollistetaan liiketoimintasääntöjen määrittämisen ja hallinnan. Viestienseurannan avulla voidaan seurata viestien liikkumista järjestelmien välillä ja pyrkiä havaitsemaan mahdolliset ongelmat tai viivästyksset. Tietojen eheyttä voidaan varmistaa eri järjestelmien välisten tapahtumien- eli transaktioiden hallinnan avulla. Siinä

tallennetaan ja valvotaan eri järjestelmistä tulevia tapahtumia, kuten virheitä ja onnistuneita viestien toimituksia. Palveluiden orkestrointi tarkoittaa kykyä hallita useita palveluita ja prosesseja kokonaisuutena. Järjestelmänhallinnassa on työkaluja ja toiminnollisuuksia, joilla helpotetaan ESB-ympäristön hallintaa, seurantaa ja konfigurointia. (Richards 2006.)

3.5.1 Dataintegraatiomallin hyödyt

ETL-prosessit mahdollistavat suurten datamäärien tehokkaan siirron ja muokkauksen eri lähteistä moniin kohteisiin. Ne mahdollistavat datan muokkaamisen, puhdistamisen ja uudelleenjärjestämisen ennen sen tallentamista tietovarastoihin. Tämä takaa datan paremman laadun ja yhtenäisyyden sekä tekee datan helposti analysoitavaksi. ETL-prosessit ovat skaalautuvia, mikä mahdollistaa erilaisten datamäärien käsittelyn. Lisäksi niitä voidaan aikatauluttaa automaattisesti toimiviksi. (Smallcombe 2020; Keebola 2022.)

Enterprise Service Busin vahvuutena pidetään nopeaa ja luotettavaa viestintää eri sovellusten välillä. Se mahdollistaa reaaliaikaisen tiedonvälityksen ja tapahtumien käynnistämisen. Sovellusten riippuvuus vähenee, kun ESB toimii välittäjänä, mikä tekee järjestelmästä joustavamman ja ylläpidettävämmän. Palvelukeskeinen arkkitehtuuri helpottaa palvelujen keskittämistä ja jakamista eri sovellusten välillä. ESB:a ei ole sidottu tiettyyn teknologiaan, vaan sillä tuetaan erilaisia kuljetusprotokollia, kuten AMQP, HTTP/s ja MQTT. (Ariyarathne 2019; Smallcombe 2020.)

3.5.2 Dataintegraatiomallin haasteet

ETL-prosessit ovat tehokkaita datan siirtämisessä suurina erinä, mutta niiden luomaa tietovarastoa on ylläpidettävä ja tarkasti seurattava. Datan lähteiden tai vaatimusten muuttuessa ETL-prosessit on suunniteltava ja muokattava uudelleen. Jatkuvat muutokset voivat vaikuttaa prosessien tehokkuuteen ja tarkkuuteen. Suurten datamäärien käsittely ETL-prosesseissa voi kestää aikaa. Tämä voi aiheuttaa viiveitä tiedon saatavuudessa. (Smallcombe 2020; Keebola 2022.)

Enterprise Service Busin haasteena on ylikuormittuminen ja hidastuminen, jos siinä käsitellään liian monta viestiä tai tehtävää samanaikaisesti. Tämä vaarantaa ESB:n suorituskyvyn ja saattaa vaikuttaa negatiivisesti reaaliaikaisten sovellusten ja prosessien toimintaan. Skaalautuvuuden hallinta voi olla ESB:ssa haaste, kun järjestelmää laajennetaan suuremmille kuormille. Näiden syiden vuoksi ESB ei ole ihanteellinen ratkaisu suurten datamäärien siirtämiseen, sillä siihen on parempi vaihtoehto ETL. ESB-ratkaisujen ylläpitäminen ja hallinta voivat olla monimutkaisia erityisesti suurissa organisaatioissa tai monimutkaisten sovellusten integroinnissa, mikä lisää riskiä virheille. Tämä riski kasvaa

erityisesti silloin, kun järjestelmässä on paljon erilaisia viestikanavia ja integraatioita. (Nilsen 2018; Smallcombe 2020.)

3.5.3 Dataintegraatiomallin tietoturva ja tietosuoja

ETL- ja ESB-integraatioprosesseissa on tärkeää varmistaa turvalliset yhteydet datan siirrossa ja estää haitalliset sisällöt. Tärkeää on myös rajoittaa pääsyä vain tarvittaville henkilöille ja sovelluksille. (Panhelainen 2008.)

Datan siirron aikana ja sovellusten välillä on tärkeää käyttää vahvaa salausmenetelmää. Erityisesti ESB:n kautta kulkevan datan salaus on tärkeää, koska se auttaa pitämään tiedot turvassa, kun ne kulkevat keskitetyn väylän läpi. (Panhelainen 2008.)

3.5.4 Dataintegraatiomallin soveltuvuus

ETL on erityisen sopiva yrityksille, joissa käsitellään suuria datamääriä ja jossa tarvitaan tehokasta tapaa siirtää ja valmistella dataa tietovarastoja varten sekä tuottaa erilaisia raportteja ja analyyskejä. Esimerkkinä tästä voivat olla suuret yritykset, joissa tarvitaan päivittäisiä analyyskejä ja raportointia liiketoiminnan päätöksenteon tueksi. (Smallcombe 2020.) ELT voi olla hyödyllinen erityisesti silloin, kun yrityksellä on tarve käsitellä todella suuria määriä dataa ja ELT sopii erityisen hyvin monimuotoisen ei-strukturoidun datan käsittelyyn (Keebola 2022).

Enterprise Service Bus on hyvin soveltuva organisaatioille, joissa hyödynnetään palvelukeskeistä arkkitehtuuria ja jossa sovellukset voivat tarjota ja kuluttaa palveluita toisiltaan. Enterprise Service Busia käytetään pääasiassa sovellusten integrointiin eikä suurten datamäärien siirtämiseen. ESB voi olla sopiva ratkaisu, kun tarvitaan joustavia integraatioita eri järjestelmien ja sovellusten välillä, erityisesti tilanteissa, joissa tietoja pitää siirtää reaaliajassa. (Smallcombe 2020.)

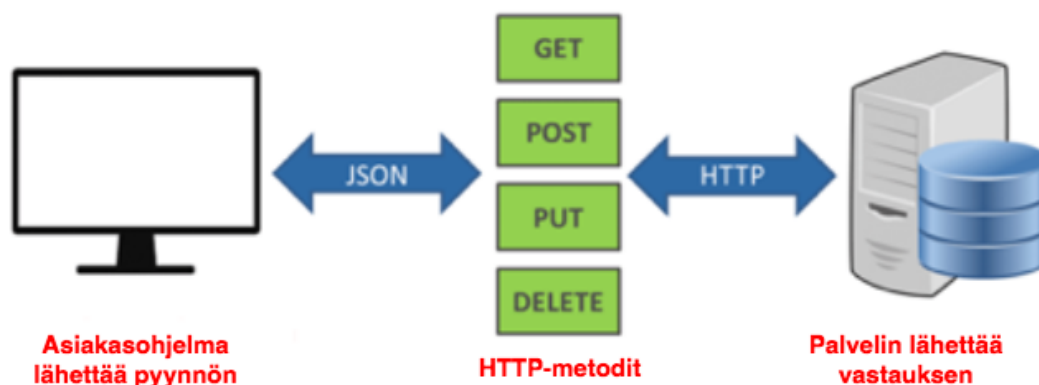
3.6 Sovellusten ohjelmointirajapintaintegraatiot (API)

Sovellusten ohjelmointirajapintaintegraatioiden (engl. Application Programming Interface, API) avulla eri sovellukset voivat kommunikoida keskenään ja vaihtaa tietoa. Erilaiset API-formaatit ja -protokollat tarjoavat erilaisia tapoja toteuttaa API-integraation. Yleisimpiä ovat REST (Representational State Transfer), SOAP (Simple Object Access Protocol), GraphQL ja gRPC. (Reselman 2020).

REST API

REST (Representational State Transfer) API perustuu RESTful-arkkitehtuuriin, joka pohjautuu HTTP-protokollaan (Reselman 2020). RESTful API on tilaton (engl. stateless), mikä tarkoittaa, että jokainen pyyntö itsenäinen ja sisältää kaiken tarvittavan tiedon pyynnön käsittelyyn, eikä palvelimen tarvitse muistaa aiempia pyyntöjä. Palvelin ei säilytä mitään tietoa pyynnöstä sen suorittamisen jälkeen. Tilattomuuden etuna on yksinkertaisuus ja skaalautuvuus. Palvelimet eivät kuormitu säilyttämällä tietoja, mikä mahdollistaa suuremman määrän samanaikaisia yhteyksiä. (Gupta 2022.) Tilatietoja voidaan säilyttää asiakkaalla esimerkiksi selaimessa, tai ne voidaan lähettää evästeiden ja tokenien muodossa (Ram 2021).

GraphQL-arkkitehtuurissa käytetään asiakas-palvelin-mallia, jossa asiakas tekee HTTP-pyyntöjä palvelimelle erilaisten HTTP-metodien avulla (Reselman 2020). Näiden metodien toimintaa ja vuorovaikutusta asiakkaan ja palvelimen välillä havainnollistetaan kuviossa 12. Metodit kuten GET, POST, PUT ja DELETE vastaavat eri operaatioita resurssien hallinnassa, kuten luomista, lukemista, päivittämistä ja poistamista. Palvelin puolestaan vastaa pyyntöihin toimittamalla dataa yleisimmin JSON-muodossa, mutta myös XML ja muut formaatit ovat mahdollisia. (Tuli 2021.)



Kuvio 12. REST API:n asiakas-palvelin-vuorovaikutus ja HTTP-metodit (mukailtu Tuli 2021)

REST-arkkitehtuuri käsittelee jokaista komponenttia resurssina (Tutorials Point 2023a). Jokainen resurssi on yksilöitävissä URL-osoitteen avulla (Reselman 2020), ja pyyntöjen ja vastausten metadatatiedot kulkevat http-otsikoiden mukana. Vastauksen otsikot voivat sisältää tietoja, kuten sisällön tyyppin (engl. Content-Type) ja sisällön pituuden (engl. Content-Length). (Tuli 2021).

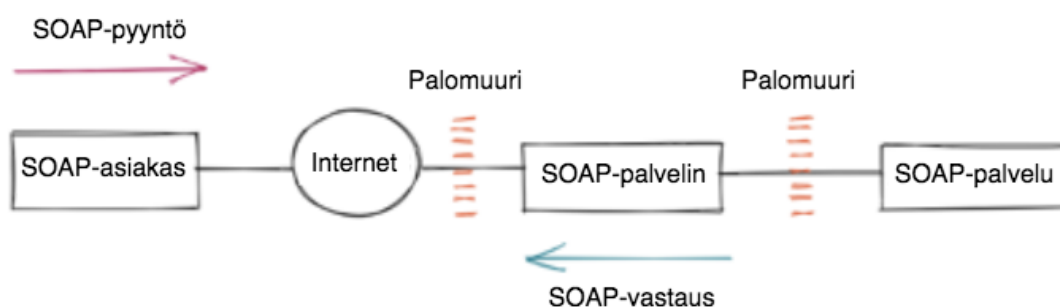
SOAP API

SOAP (Simple Object Access Protocol) API perustuu XML-formaattiin ja voi käyttää erilaisia kuljetusprotokollia, kuten HTTP, FTP ja SMTP (Reselman 2020). SOAP tukee

monimutkaisempia toimintoja, kuten transaktioita ja ACID-ominaisuuksia. Transaktiot mahdollistavat useiden operaatioiden suorittamisen yhtenä kokonaisuutena, joko kaikki onnistuen tai kaikki epäonnistuen. Tämä on tärkeää liiketoimintakriittisissä sovelluksissa, kuten pankkijärjestelmissä, joissa varmuus ja eheys ovat ensiarvoisen tärkeitä. (AltexSoft 2019.)

SOAP-palvelut määritellään yleensä WSDL-tiedostoilla (Web Service Description Language). WSDL on XML-pohjainen kieli, joka kuvaa palvelun tarjoaman rajapinnan, mukaan lukien sen tarjoamat toiminnot ja niiden parametrit. Tämä tekee SOAP-palvelujen integroinnista helpompaa, koska asiakas voi tarkistaa WSDL-tiedoston ja ymmärtää, mitä palvelulta voi odottaa. (Reselman 2020.)

Kuvio 13 havainnollistaa SOAP-arkkitehtuurin peruseriaatetta, kun SOAP-viesti liikkuu verkon eri osien läpi palvelimen ja asiakkaan välillä. Asiakas (engl. SOAP client) lähettää SOAP-pyyntöviestin Internetin kautta palvelimelle. Tämä voi olla esimerkiksi sovellus, joka tarvitsee tietoa etäpalvelulta. Palvelin (engl. SOAP server) vastaanottaa SOAP-pyyntöviestin Internetin kautta. Tämä voi olla esimerkiksi verkkopalvelin, joka tarjoaa tietoa tai suorittaa toimintoja asiakkaan puolesta. Palvelin käsittelee SOAP-pyyntöviestin ja tarjoaa siihen vastauksen. SOAP-palvelu (engl. SOAP service) voi suorittaa monenlaisia tehtäviä, kuten tietokantakyselyjä tai liiketoimintalogiikkaa. Tietoturvamenetelmät, kuten WS-Security, voidaan integroida SOAP-viestien suojaksi matkan varrella, ja palomuurit auttavat suojaamaan verkon turvallisuutta. (Janani 2022.)



Kuvio 13. SOAP-viestin matka verkon läpi (mukailtu Janani 2022)

SOAP-protokolla voi olla sekä tilaton (engl. stateless) että tilallinen (engl. stateful) riippuen siitä, miten sitä käytetään ja millaisia ominaisuuksia se hyödyntää. Kun SOAP-kommunikaatio on tilaton, jokainen pyyntö-vastaus-pari käsitellään erillisenä ja riippumattomana tapahtumana. Tämä tarkoittaa, että palvelin ei säilytä tietoa aiemmista pyynnöistä, ja jokainen pyyntö on itsenäinen. Tämä voi olla hyödyllistä tilanteissa, joissa jokainen pyyntö on erillinen ja riippumaton. Kun SOAP-kommunikaatio on tilallinen (engl. stateful), palvelin säilyttää tiedon aiemmista pyynnöistä ja niiden tilasta. Tämä mahdollistaa

monimutkaisemmat vuorovaikutukset palvelimen ja asiakkaan välillä. Esimerkkejä tilallisista sovelluksista voivat olla pankkijärjestelmät, joissa tarvitaan tietoa edellisistä tapahtumista, kuten tilitapahtumista, jotta voidaan suorittaa uusia tapahtumia, kuten tilisiirtoja. (Janani 2022.)

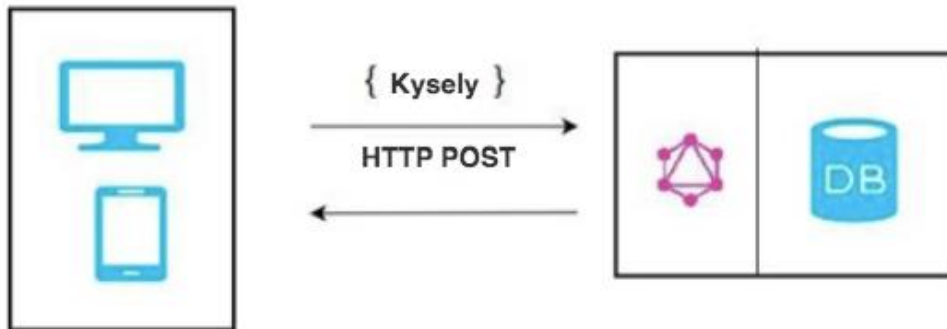
GraphQL API

GraphQL perustuu GraphQL-kielen käyttöön ja on erityisesti suunniteltu datan hakemiseen ja muokkaamiseen. GraphQL on määrittely, joka kuvaa, miten GraphQL-palvelimen käyttäytyy. Se tarjoaa ohjeita siitä, miten pyynnöt ja vastaukset tulee käsitellä, kuten tuetut protokollat, hyväksyttävän datan muoto ja palvelimen palauttaman vastauksen muoto. (Tutorials Point 2023b.)

GraphQL API:ssa käytetään HTTP POST -metodia palvelimen kanssa kommunikointiin. GraphQL API on kuljetuskerroksesta riippumaton, mikä tarkoittaa, että GraphQLia voi käyttää eri verkkoprotokollien, kuten TCP:n tai Websocketin, kanssa (Burk 2017). Lisäksi GraphQL on tietokantateknologiasta riippumaton, joten sitä voi käyttää sekä relaatiotietokantojen että NoSQL-tietokantojen kanssa (Tutorials Point 2023b).

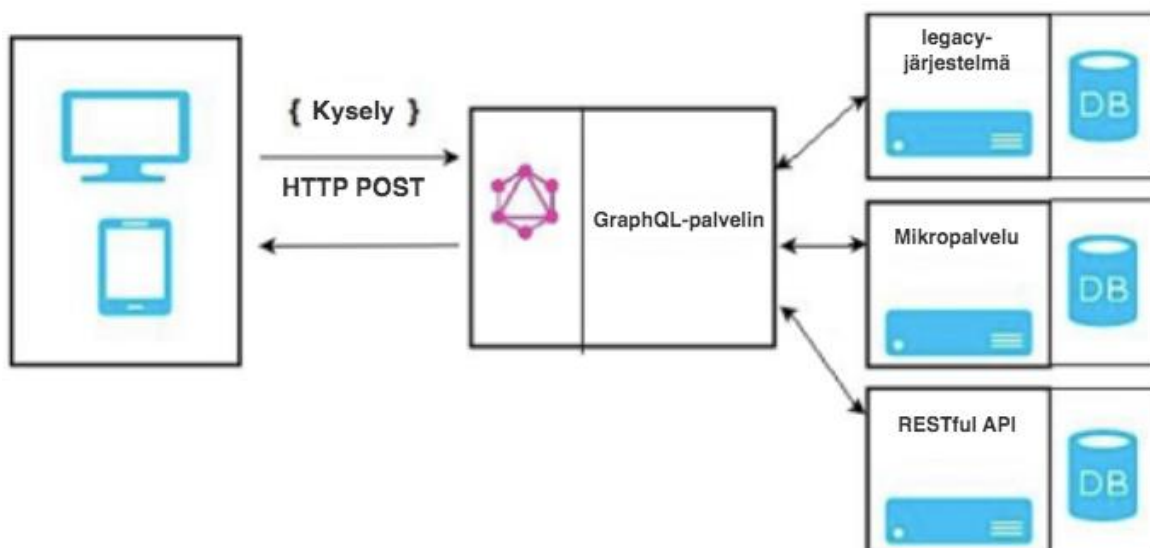
GraphQL API:n toteutuksessa määritellään GraphQL skeema (engl. schema), joka kuvaa käytettävät tyypit, kyselyt ja mutaatiot. Skeema määrittää, millaista dataa voi hakea ja millaisia muutoksia voi tehdä API:n kautta. GraphQL-pyyntö voidaan tehdä palvelimelle käyttäen tätä skeemaa, ja vastauksena voidaan saada tarkasti haluttu tieto. (Reselman 2020.)

Kolme yleisintä tapaa toteuttaa GraphQL-palvelin ovat GraphQL-palvelin liitetyllä tietokannalla, olemassa olevat järjestelmät integroiva GraphQL-palvelin, ja hybridilähestymistapa. Kuvio 14 kuvaa arkkitehtuuri, jossa GraphQL-palvelin integroituu tietokantaan. Se on usein hyvä valinta uusille projekteille. Kun palvelin saa kyselyn (engl. query), se lukee pyyntöpayloadin ja hakee tarvittavan datan tietokannasta. Tämä tieto palautetaan asiakkaalle noudattaen GraphQL-spesifikaation määrittelemää formaattia. (Tutorials Point 2023b.)



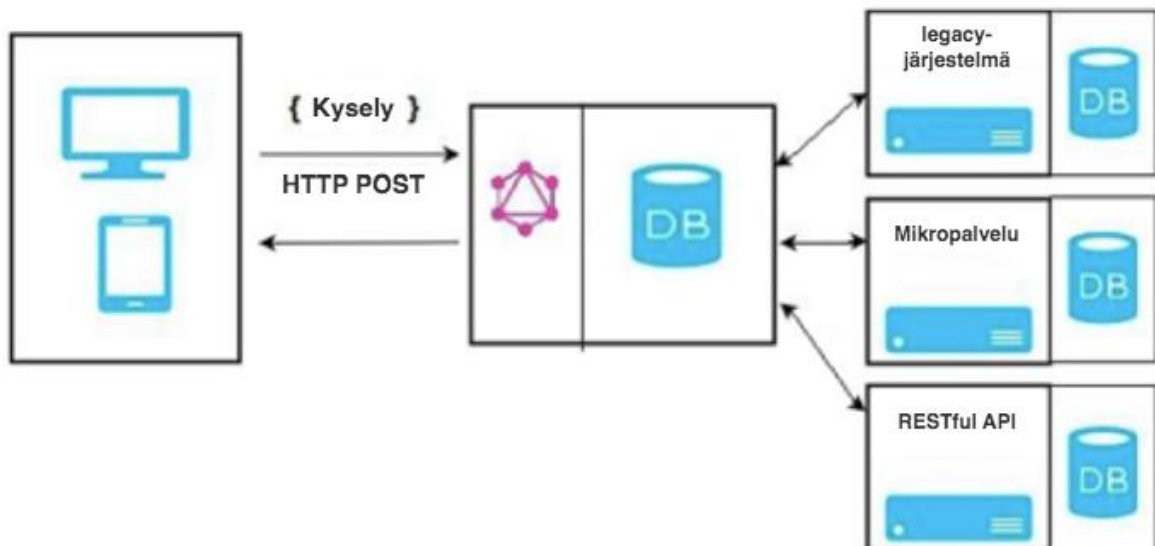
Kuvio 14. Tietokantaan integroitava GraphQL-palvelin (mukailtu Tutorials Point 2023b)

Kuvio 15 kuvaa arkkitehtuurimallia, jossa GraphQL-palvelin integroi olemassa olevat järjestelmät. Tämä sopii yrityksille, joilla on vanhaa infrastruktuuria ja erilaisia API:ita. GraphQL voi toimia yhdistävänä kerroksena, joka yhdistää mikropalvelut, vanhan infrastruktuurin ja kolmannen osapuolen API:t nykyiseen järjestelmään. (Tutorials Point 2023b.)



Kuvio 15. Olemassa olevat järjestelmät integroiva GraphQL-palvelin (mukailtu Tutorials Point 2023b)

Hybridi-arkkitehtuurissa yhdistyvät edellä mainitut lähestymistavat kuvion 16 mukaisesti. GraphQL-palvelin voi ratkaista kyselyt joko liitetystä tietokannasta tai integroidusta API:sta riippuen, mikä palvelee parhaiten pyyntöä. (Tutorials Point 2023b.)

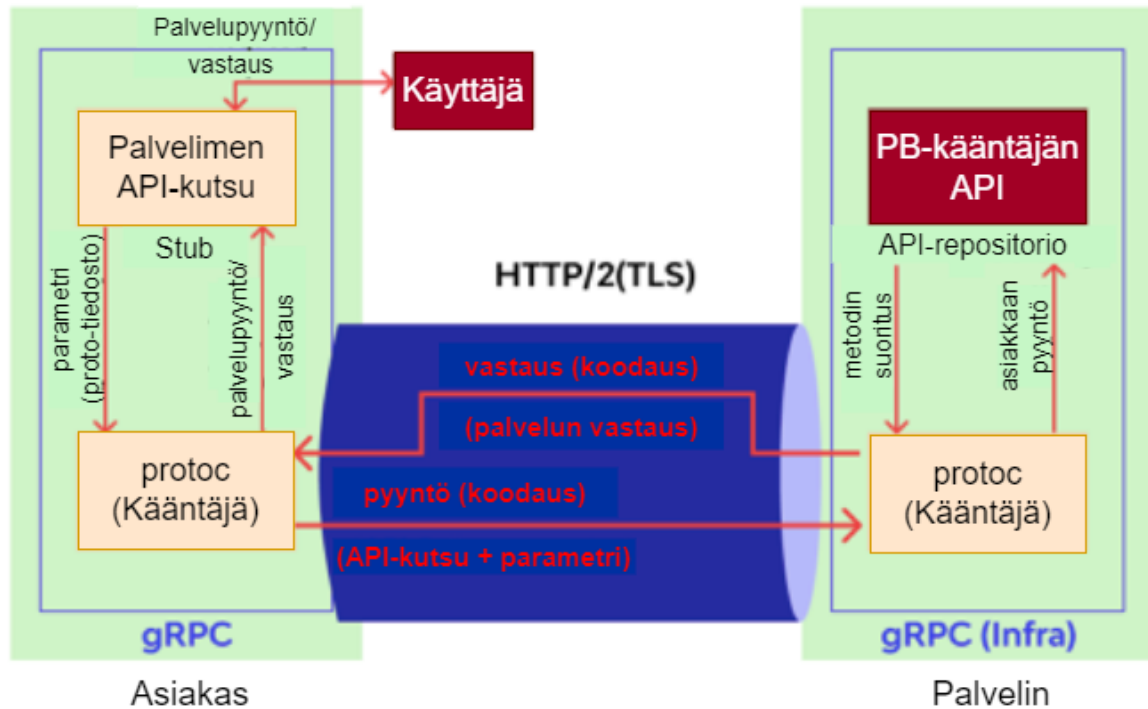


Kuvio 16. Hybridilähestymistapa GraphQL-palvelimen arkkitehtuurissa (mukailtu Tutorials Point 2023b)

gRPC API

gRPC API on Googlen kehittämä moderni avoimen lähdekoodin RPC-kehys (Remote Procedure Call), joka perustuu Protocol Buffers -binäärimuotoon ja HTTP/2-protokollaan. Tämä mahdollistaa nopean ja tehokkaan tiedonsiirron sekä tukee sekä yksisuuntaista että kaksisuuntaista tietojen vaihtoa. Saman Protocol Buffers -skeeman käyttöä edellytetään kehysten käytössä sekä asiakas- että palvelinpuolella. (Reselman 2020.)

Kuviossa 17 esitellään gRPC:n arkkitehtuurin keskeiset komponentit sekä tiedon kulku asiakkaan ja palvelimen välillä. Palvelupyynnön käsitteleminen aloitetaan lähettämällä asiakkaan toimesta pyyntö, joka kulkee palvelimen API:n kautta Protobufin protokompilaattorille. Protoc-kompilaattori luo asiakas- ja palvelinkoodia, joka lukee .proto -tiedoston muistiin ajon aikana ja käyttää muistissa olevaa skeemaa binääri viestien serialisointiin ja deserialisointiin. Tämä mahdollistaa huomattavasti tehokkaamman tiedonsiirron verrattuna JSONiin tai XML:ään. Palvelinpuolella protoc-kompilaattori vastaanottaa pyynnön ja ohjaa sen Protocol Buffers -kompilaattorin API:n kautta palvelumetodin suoritukseen. Tiedonsiirto asiakkaan ja palvelimen välillä tapahtuu HTTP/2-protokollan ja TLS-salauksen avulla, mikä varmistaa tehokkaan ja turvallisen tiedonvaihdon. (Wallarm 2023.)



Kuvio 17. gRPC API:n arkkitehtuuri ja tiedonvaihto (mukailtu Wallarm 2023)

3.6.1 Dataintegraatiomallin hyödyt

API-integraatiot tarjoavat yrityksille joustavuutta ja skaalautuvuutta, ja ne mahdollistavat joustavan integraation eri järjestelmien ja palveluiden välillä. API-integraatiot mahdollistavat myös nopean reagoinnin muuttuviin tarpeisiin ja voivat vähentää päällekkäistä työtä sekä tehostaa kehitystä. (Guthrie 2023.) Lisäksi kolmannen osapuolen APIen käyttö voi nopeuttaa sovelluskehitystä merkittävästi (Gail 2021). API-integraatiot ovat yleensä kustannustehokkaampia ja nopeampia toteuttaa kuin monimutkaisemmat integraatoratkaisut kuten ETL tai ESB (Performance Software 2023). Eri API-formaattien spesifiset hyödyt vaihtelevat.

REST API:n hyödyt

REST API on yksinkertainen ja kevyt integraatoratkaisu, joka on helppo toteuttaa (Reselman 2020). Sen nopeus on yksi sen merkittävimmistä eduista, koska siinä ei ole tiukkoja spesifikaatioita kuten SOAPissa, ja se kuluttaa vähemmän kaistaa ja resursseja (JavaTPoint 2021). REST API on myös kieli- ja alustariippumaton, tukee useita dataformaatteja ja voi integroitua SOAP-webpalveluihin. Lisäksi REST API on suunniteltu skaalautumaan helposti, koska jokainen pyyntö sisältää tarvittavan tiedon eikä vaadi palvelimelta tilan säilyttämistä. Tämä mahdollistaa sen, että uusia resursseja tai palvelimia voidaan lisätä tarpeen mukaan ilman suurta monimutkaista konfigurointia. (Janani 2022.)

REST API onkin laajasti käytetty monenlaisissa sovelluksissa ja arkkitehtuureissa (Reselman 2020).

SOAP API:n hyödyt

SOAP API tukee useita kuljetusprotokollia. WSDL-kuvaus helpottaa palvelimen toiminnan ymmärtämistä. Language neutral -mahdollisuus tukee eri ohjelmointikieliä. (Reselman 2020.)

GraphQL API:n hyödyt

GraphQL API on joustava ja monipuolinen integraatoratkaisu (Reselman 2020). GraphQL API:n avulla on mahdollisuus määrittellä tarkasti haluttu data kyselyssä, mikä tarkoittaa, että asiakas voi pyytää vain tarvitsemansa tiedot (Reselman 2020; The GraphQL Foundation 2023a). GraphQL on myös vahvasti tyyppiin perustuva, mikä tarkoittaa, että kyselyissä on oltava oikeanlaiset tyypit. Tämä ominaisuus auttaa vähentämään virheitä ja helpottaa vianetsintää. (The GraphQL Foundation 2023b; Tutorials Point 2023c.)

GraphQL tukee asynkronista tiedonvaihtoa tilausmekanismin (engl. GraphQL Subscriptions, GQLS) avulla, mikä mahdollistaa reaaliaikaisen datan päivityksen ja tapahtumien seurannan. (Reselman 2020.) GraphQL API mahdollistaa useiden liitettävien liiketoimintaobjektien hakemisen yhdellä kyselyllä, mikä voi vähentää tarvetta tehdä useita pyyntöjä palvelimelle saman datan saamiseksi. Tämä parantaa suorituskykyä ja vähentää verkkoliikennettä. (The GraphQL Foundation 2023c.) GraphQL tarjoaa myös monipuoliset kehittäjätyökalut, kuten automaattisen kyselyn dokumentaation ja älykkään koodin täydennyksen, mikä helpottaa API:n käyttöä ja testausta (The GraphQL Foundation 2023d).

gRPC API:n hyödyt

gRPC API mahdollistaa nopean ja tehokkaan tiedonvaihdon Protocol Buffers -binäärimuodolla (Reselman 2020). gRPC:n binäärimuoto tehostaa datan käsittelyä, mikä edistää sen nopeaa suorituskykyä. gRPC tarjoaa jopa 10-kertaisesti paremman suorituskyvyn ja API-tietoturvan verrattuna REST:n json-kommunikointiin. gRPC tukee myös erilaisia suoratoistotyypppejä, kuten palvelinpuoleista, asiakaspuoleista ja kaksisuuntaista suoratoistoa (engl. Bidirectional streaming). Kaksisuuntaisella suoratoistolla mahdollistetaan asynkroninen tiedonvaihto. (Wallarm 2022.) gRPC tukee myös monia eri ohjelmointikieliä, ja sen kautta toteutettujen sovellusten on helppo olla yhteensopivia eri järjestelmien kanssa (gRPC Authors 2023a). Tietoturva on myös vahva, sillä gRPC käyttää http/2:n yli TLS:n päästä päähän salauksen yhteyttä (Wallarm 2022).

3.6.2 Dataintegraatiomallin haasteet

API-integraatioiden yleisiin haasteisiin kuuluu yhteensopivuusongelmat, erityisesti kun uusia API:eja yhdistetään vanhoihin järjestelmiin. Tietoturva on myös huolenaihe, sillä API:n turvallisen käytön varmistaminen tunnistautumis- ja käyttöoikeusmekanismien avulla voi olla haastavaa. Versiointi ja ylläpito tuovat omat ongelmansa, kuten API-muutosten hallinta ja mahdollisen dokumentaation puute. Kustannukset voivat nousta, jos API-kutsujen määrä ylittää sallitun rajan. Monimutkaiset integraatiot saattavat lisätä virheriskiä, ja kolmannen osapuolen API:ien kanssa on aina olemassa luotettavuusriskejä. (Healthie 2022.) Lopuksi, kun käsitellään henkilötietoja, tulee varmistaa, että integraatiot noudattavat tietosuojalainsäädäntöä, kuten GDPR:a EU:ssa (Art 2020).

REST API:n haasteet

REST API:n haasteena on sen rajoittunut joustavuus datan rakenteessa. Lisäksi samaan HTTP/1.1 -yhteyteen liittyvät pyynnöt ja vastaukset hidastavat tiedonvaihtoa. (Reselman 2020.) Se, että REST ei sisällä ACID (Atomicity, Consistency, Isolation, Durability) -ominaisuuksia, jotka ovat tärkeitä tietokannoissa ja transaktiohallinnassa, voi olla haitaksi esimerkiksi liiketoimintakriittisissä sovelluksissa, kuten pankkijärjestelmissä. Tämä tarkoittaa, että RESTin avulla ei ole mahdollista varmistaa, että monimutkaiset tietokantatransaktiot suoritetaan oikein ja varmuudella. (Som 2018.)

SOAP API:n haasteet

SOAP API:n haasteena on sen monimutkainen XML-muoto ja ulkoiset namespace-määrittelyt, jotka voivat tehdä viesteistä raskaita ja vaikeita hallita. (Reselman 2020.)

GraphQL API:n haasteet

GraphQL API:n käyttö vaatii GraphQL-kielen ja -spesifikaation tuntemusta sekä asynkronisen tiedonvaihdon toteuttamista Subscriptions-mekanismien avulla, mikä voi olla monimutkaista verrattuna perinteisiin synkronisiin RESTful API:n palveluihin (Reselman 2020.)

gRPC API:n haasteet

gRPC API:n käyttö edellyttää Protocol Buffers -spesifikaation tuntemusta ja http/2-protokollan tukea sekä asiakkaalta että palvelimelta. Tämä tekee siitä vähemmän ihanteellisen ratkaisun suoraan selaimesta tapahtuvaan kutsuun. Protobuf-muodon luettavuuden takia viestien vianetsintä ja manuaalisten pyyntöjen tekeminen voi olla monimutkaisempaa. gRPC:n nopea tiedonvaihto vaatii tarkkaa optimointia. (Reselman 2020.)

3.6.3 Dataintegraatiomallin tietoturva ja tietosuoja

SOAP ja gRPC tarjoavat vahvaa sisäänrakennettua tietoturvaa, kun taas REST ja GraphQL vaativat usein enemmän kehittäjän huomiota ja lisätoimenpiteitä tietoturvan varmistamiseksi.

REST API:n tietoturva ja tietosuoja

REST ei tarjoa sisäänrakennettuja tietoturvamekanismeja, joten sen turvallisuuden varmistaminen jää suurelta osin kehittäjän vastuulle. Lisätoimenpiteitä, kuten SSL/TLS-protokollaa, on käytettävä turvalliseen tiedonsiirtoon. Lisäksi käyttäjien tunnistamisesta sekä oikeuksien hallinnasta sovellustasolla on huolehdittava. (Neray & Scott 2021.)

RESTin turvallisuus voi olla riippuvainen kehittäjän taidoista ja huolellisuudesta, mikä saattaa johtaa tietoturva-avoittuvuuksiin. Tästä syystä on tärkeää noudattaa parhaita käytäntöjä, kuten kaikkien syötteiden tarkistamista palvelimella, välttää arkaluonteisten tietojen, kuten käyttäjienimien ja salasanojen, sisällyttämistä URL-osoitteisiin ja hallita HTTP-metodien käyttöä asianmukaisesti. Lisäksi on tarpeen tarkistaa syötteen oikeellisuus sekä käyttää http-virhesanomiamäärittämiä kieltämiseen. (Tutorials Point 2023d.)

SOAP API:n tietoturva ja tietosuoja

SOAP integraatioissa turvallisuus on usein vahvempi. SOAP sisältää erilaisia turvallisuusprotokollia, kuten WS-Security, jotka mahdollistavat viestinnän salaamisen, käyttäjien tunnistuksen ja virheiden hallinnan. Näitä ominaisuuksia pidetään usein osana standardia, mikä tarkoittaa, että ne ovat käytettävissä integraatioissa ilman erityistä konfigurointia. SOAP tarjoaa vahvan tietoturvan erityisesti liiketoimintakriittisissä sovelluksissa. (Walker 2023.)

GraphQL:n tietoturva ja tietosuoja

GraphQL:ssä tietoturva riippuu sovelluksen toteutuksesta. GraphQL-protokolla itsessään ei tarjoa vahvaa sisäänrakennettua tietoturvaa, mutta se antaa kehittäjille mahdollisuuden määrittellä tietoturvamekanismeja sovelluksen tarpeiden mukaisesti. GraphQL:ssä voidaan määrittellä, mitkä kentät ja resurssit ovat näkyvissä ja muokattavissa jokaiselle käyttäjälle. Tämä voi parantaa tietoturvaa ja tietosuojaa, mutta se vaatii myös huolellista suunnittelua. (Stemmler 2021.)

gRPC:n tietoturva ja tietosuoja

gRPC sisältää useita tietoturvamekanismeja, kuten määritykset TLS-salaamisesta ja käyttäjien tunnistuksesta. gRPC mahdollistaa myös käyttäjien oikeuksien hallinnan ja

tietoturvapäivitysten helpon integroinnin. gRPC:ssä voidaan käyttää erilaisia autentikaatiomenetelmiä, kuten JWT-tokeneita. Tämä tekee gRPC:stä erittäin turvallisen vaihtoehdon. (gRPC Authors 2023b.)

3.6.4 Dataintegraatiomallin soveltuvuus

API-integraatiot soveltuvat eri kokoisille, eri toimialoilla toimiville yrityksille, jotka etsivät joustavuutta, nopeutta ja skaalautuvuutta. Ne ovat erityisen hyödyllisiä start-upeille ja nopeasti kasvaville yrityksille, jotka tarvitsevat kustannustehokkaita ja nopeita integraatoratkaisuja. API-integraatiot sopivat myös yrityksille, joille tietoturva on kriittinen huolenaihe, sillä ne tarjoavat vahvoja tietoturvamekanismeja ja suojattua tiedonsiirtoa. (Reselman 2020.)

Valittu API-tyyppi riippuu projektin monimutkaisuudesta ja tarpeista. Yksinkertaisissa projekteissa RESTful API voi olla riittävä, mutta monimutkaisemmissa tapauksissa SOAP, GraphQL tai gRPC voivat olla parempia vaihtoehtoja. Mikropalveluarkkitehtuuriin siirtyminen voi myös olla harkinnan arvoinen vaihtoehto, erityisesti suurissa järjestelmissä, joissa on useita riippuvaisia integraatioita. Mikropalveluarkkitehtuuri voi tarjota joustavuutta ja hajautettua arkkitehtuuria, joka tukee monimutkaisia integraatioita ja skaalautuvuutta. (Janani 2022.)

REST API:n soveltuvuus

REST API soveltuu hyvin julkisiin sovelluksiin ja yksinkertaisiin, resurssipohjaisiin sovelluksiin. REST API:n helppokäyttöisyys ja laaja selaintuki tekevät siitä hyvän valinnan verkkosivujen ja mobiilisovellusten taustajärjestelmiin. (AltexSoft 2020.) Lisäksi REST API soveltuu yksinkertaisiin ja laajasti käytettyihin integraatioihin, kuten perusdatan siirtämiseen eri järjestelmien välillä tai palveluiden välisiin kommunikaatioihin. Sen helppokäyttöisyys ja skaalautuvuus tekevät siitä erinomaisen valinnan sovelluksille, jotka eivät vaadi tiukkoja spesifikaatioita kuten SOAP-webpalveluissa. REST API:a pidetään hyvänä valintana, kun tarvitaan nopeaa tiedonvaihtoa JSON- tai XML-formaatissa, mikä on yleistä käytäntöä web-sovellusten ja palveluiden välillä. REST ei kuitenkaan ole ihanteellinen monimutkaisten transaktioiden ja tietokantajärjestelmien hallintaan, sillä se vaatii usein useita peräkkäisiä pyyntöjä ja voi olla haasteellinen monimutkaisissa tietokantavaatimuksissa. (Janani 2022.)

SOAP API:n soveltuvuus

SOAP API soveltuu tilanteisiin, joissa tarvitaan tarkkaa XML-formaattia ja tukea useille kuljetusprotokollille. Se on hyvä valinta vanhemmissa järjestelmissä ja monimutkaisemmissa integraatioissa, koska se tukee monimutkaisempia toimintoja, kuten

transaktioita ja ACID-ominaisuuksia. SOAP:a pidetään hyvänä vaihtoehtona erityisesti liiketoiminnallisesti kriittisiin ja turvallisuuden kannalta herkkiin integraatioihin, jotka vaativat transaktioiden ja tietokannan eheyden hallintaa. (Janani 2022.) SOAP on myös yleinen valinta maksuliikenteen, identiteetinhallinnan, asiakkuudenhallinnan (CRM) sekä finanssi- ja telekommunikaatiopalveluiden aloilla. Se tukee myös legacy-järjestelmiä, mikä tekee siitä joustavan ratkaisun monenlaisiin tarpeisiin. (AltexSoft 2020.)

GraphQL API:n soveltuvuus

GraphQL API soveltuu tilanteisiin, joissa tarvitaan joustavaa ja monipuolista tiedonvaihtoa frontendin ja backendin välillä (Janani 2022). Se on erityisen hyödyllinen monimutkaisissa järjestelmissä, joissa on tarve suurelle joustavuudelle datan suhteen. GraphQL API sopii erinomaisesti mobiilisovelluksille, jotka tarvitsevat räätälöityä dataa, sekä mikropalveluarkkitehtuureille, joissa yksittäisten palveluiden välillä saattaa olla monimutkaisia datariippuvuuksia. (AltexSoft 2020.) Lisäksi se on hyvä valinta tilanteissa, joissa tarvitaan asynkronista tiedonvaihtoa ja tarkkaa datakontrollia. GraphQL API:a pidetään hyvänä valintana myös silloin, kun halutaan välttää tarpeetonta datan noutamista tai monien pyyntöjen tekemistä saman datan saamiseksi. (Janani 2022.)

gRPC API:n soveltuvuus

gRPC API soveltuu erinomaisesti hajautettuihin järjestelmiin, pilvipalveluihin ja mikropalveluarkkitehtuureihin, joissa nopeus, tehokkuus ja tietoturva ovat keskeisiä vaatimuksia. Erityisesti backend-sovellusten väliseen viestintään suunniteltuna gRPC tarjoaa nopean ja tehokkaan tiedonvaihdon, korkean suorituskyvyn sekä tuen monikielisille ratkaisuille (Reselman 2020). gRPC:n binäärimuotoiset viestit ja HTTP/2-tuki mahdollistavat erittäin nopeat ja tehokkaat kutsut, mikä tekee siitä erityisen sopivan suurten yritysten tarpeisiin, ja sitä käyttävät esimerkiksi Google, Facebook ja Twitch. (AltexSoft 2020.)

Lisäksi tietoturvan ja eristyksen näkökulmasta gRPC on luotettava valinta, mikä tekee siitä hyvin sopivan pilvipohjaisiin mikropalveluarkkitehtuureihin, joissa tietoturva ja eristys ovat tärkeitä. Kuitenkin rajoitettu selaintuki ja oppimiskynnys voivat muodostua haasteeksi gRPC:n käytölle. (Janani 2022.)

4 Datan laadunhallinnan merkitys dataintegraatioissa

Datan laadunhallinta on keskeinen osa dataintegraatioita nykypäivän dataohjautuvassa liike-elämässä. Eri lähteistä peräisin olevaa tietoa yhdistettäessä on tärkeää varmistaa tiedon yhteensopivuus, tarkkuus ja luotettavuus. Tämä tarjoaa oivalluksia kuluttajakäyttäytymiseen, markkinoiden suuntauksiin ja yrityksen suorituskykyyn. Lisäksi se auttaa pienentämään dataintegraatioon liittyviä kustannuksia ja vähentämään resurssien käyttöä virheellisen tai puutteellisen tiedon korjaamiseen. Tiedon laadukas hallinta on avainasemassa, sillä se vaikuttaa suoraan analyysin onnistumiseen ja päätöksenteon tehokkuuteen. (Otieno 2023.)

Datan laadunhallintaan liittyy monia haasteita, erityisesti, kun dataa integroidaan useista eri lähteistä. Puutteellinen tai virheellinen tieto, epä johdonmukaisuudet ja turhat toistot voivat johtaa virheellisiin johtopäätöksiin. Tämän vuoksi on välttämätöntä jatkuvasti seurata ja arvioida datan laatua sekä tehdä tarvittaessa puhdistusta ja validointia. Kun dataintegraatiot muuttuvat yhä monimutkaisemmiksi, yritysten on priorisoitava nämä prosessit välttääkseen virheellisen analyysin riskit ja varmistaakseen parhaat mahdolliset tulokset ja tehokkaan dataohjautuvan päätöksenteon. (Otieno 2023).

5 Yrityshaastattelut dataintegraatioiden käytöstä ja kokemuksista

5.1 Haastattelututkimuksen toteutus

Haastattelututkimuksen tarkoituksena oli syventää ymmärrystä dataintegraatiomallien teknisestä toteutuksesta sekä saada käsitys hyvistä dataintegraatoratkaisuista, toimintatavoista ja haasteista ja niiden ratkaisuista erilaisissa Suomessa toimivissa dataekosysteemeissä. Haastattelu kohdistettiin kuudelle erilaiselle ja erikokoiselle yritykselle, jotka edustavat eri toimialoja ja joilla on kokemusta dataintegraatoratkaisuista. Haastattelut mahdollistivat paremman käsityksen yritysten hyviksi koetuista ratkaisuksista ja toimintatavoista sekä yritysten kohtaamista haasteista ja yritysten ratkaisuksista näihin haasteisiin.

Haastatteluprosessi suunniteltiin ja toteutettiin tiiviissä yhteistyössä kollegani Marja Airaksisen kanssa, koska opinnäytetyömme aiheet sivuavat toisiaan ja liittyvät samaan hankkeeseen Kyberturvalliset ekosysteemit Etelä-Karjala. Airaksisen opinnäytetyö keskittyy erityisesti Suomessa toimivien yritysten dataekosysteemien selvittämiseen ja tarkasteluun.

Haastattelut toteutettiin puolistrukturoituina haastatteluina. Tämä tarkoitti, että vaikka kysymykset oli valmiiksi määritelty, jätettiin tilaa avoimille kysymyksille ja keskustelulle. Tällä menetelmällä pyrittiin ymmärtämään syvällisesti yritysten näkemyksiä ja kokemuksia. Jokainen haastattelu kesti keskimäärin tunnin, ja haastattelut toteutettiin videoyhteyden välityksellä. Kaikki haastattelut nauhoitettiin ja litteroitiin myöhemmin analyysia varten. Tämä auttoi varmistamaan, että kaikki yritysten esittämät näkökulmat ja informaatio huomioitiin opinnäytetöissämme.

Haastateltavien yritysten identiteettiä suojellaksemme kaikki haastatteluaineistot anonyymisoitiin. Kaikki haastatteluissa saatu tieto on käsitelty luottamuksellisesti.

5.2 Yritys A

Yritys A on yksittäisen asiantuntijan perustama IT-konsultointiyritys, joka on tehnyt merkittävän työn datan integroinnin saralla. Yrityksen erikoisalaa on tietojen käsittely, hyödyntäminen ja tietovarastointi. Yritys tarjoaa datan jakamista, dataintegraatioiden rakentamista ja tiedon visualisointia. Yritys on työskennellyt useille suurille toimijoille.

Integraatoratkaisut

Yrityksen asiantuntija on valinnut dataintegraatiomalliksi API-integraatiot, joista hän on käyttänyt REST-, SOCKETS- ja SOAP-rajapintoja.

REST-rajapinta valittiin aluksi yrityksen omaan käyttöön, ja se mahdollisti monenlaisten toimintojen seurannan. Tämä rajapinta avattiin kehittäjien käyttöön, kun eräs toimija maksoi rajapinnan laajentamisesta. SOCKETS-rajapinta tarjottiin asiakkaan toimesta.

Integraatoratkaisujen hyödyt ja haasteet

Dataintegraatiot ovat mahdollistaneet monimutkaisten tietolähteiden parantamisen ja uusien palveluiden tarjoamisen. Dataekosysteemin luominen, jossa kehittäjillä on pääsy erilaisiin API-rajapintoihin, on avannut mahdollisuuksia ulkopuolisille toimijoille, jotka ovat luoneet innovaatioita ja uusia liiketoiminta-alueita yhdistämällä eri tietolähteitä. Yritys on säästännyt, kun se ei ole joutunut kehittämään kaikkia sovelluksia itse, ja ulkopuolisten sovellukset ovat tuoneet monipuolista lisäarvoa ilman lisäinvestointeja. Nämä sovellukset eivät ainoastaan rikasta dataekosysteemiä, vaan tukevat ja täydentävät yrityksen palveluita, laajentavat käyttömahdollisuuksia ja tuottavat taloudellista hyötyä. Datan integraation kannattavuus ei aina ole suoraviivaista, mutta sen epäsuorat hyödyt, kuten ulkopuolisten kehittäjien innovaatiot, voivat olla merkittäviä.

Yrityksen suurimpana haasteena datan integroinnissa on ollut datan laadun hallinta ja reaaliaikaisuuden ylläpitäminen, koska pienetkin viiveet voivat näkyä nopeasti vaikuttaen merkittävästi datan laatuun ja sovellusten toimivuuteen. On tärkeää miettiä, mitä dataa tallennetaan ja mitä ei, sillä tulevaisuuden datan käyttötarpeiden ennustaminen voi olla haastavaa.

Tietoturvan ja tietosuojan huomioiminen datan jakamisessa

Yrityksellä A tietoturva ja tietosuoja ovat keskeisiä datan integroinnissa. Yritys korostaa lakien noudattamisen tärkeyttä ja sensitiivisen datan eristämistä muista tiedoista. Yrityksen ratkaisu on fyysisten rajoitusten käyttö datan suojaamisessa.

Dataintegraatioiden vaikutukset yritykselle

Yritys A:n päätös integroida datansa REST-, SOAP- ja SOCKETS-rajapintojen avulla on ollut hyödyllinen. Tämä on mahdollistanut julkisen tiedon jakamisen, taloudelliset hyödyt, reaaliaikaisen seurannan ja ennustetoiminnon. On kuitenkin tärkeää, että yritys jatkaa tietoturvan ja tietosuojan korostamista ja datan laadun varmistamista, jotta yrityksen dataintegraatiostrategia voi jatkua onnistuneesti.

Yrityksen suositukset muille yrityksille

Yritys A korostaa liiketoimintalähtöisen datan merkitystä. Yritys A:n mukaan dataintegraation ei pitäisi olla tavoite itsessään, vaan sen tulisi aina palvella liiketoiminnan tarpeita. Ennen integraatioprojektien aloittamista on tärkeää määritellä ja ymmärtää, mitä

konkreettisia hyötyjä datan avulla halutaan saavuttaa. Pelkkä datan kerääminen ilman selkeää tavoitetta ei ole kustannustehokasta eikä tuota toivottua lisäarvoa.

Yritys A suosittelee panostamaan datan laatuun ja turvallisuuteen. Yritys painottaa datan laadunvalvonnan ja ylläpidon merkitystä, koska jälkikäteen datan laadun korjaaminen voi olla sekä aikaa vievää että kallista. Yritys suosittelee suojaamaan arkaluonteisen tiedon aina asianmukaisesti, ja tähän voidaan käyttää sekä fyysisiä että loogisia rajoituksia. Yritys varoittaa myös liiallisen datan keräämisen riskeistä, jotka voivat ylikuormittaa järjestelmiä ja aiheuttaa niiden toimintahäiriöitä. Tämän lisäksi ylimääräinen tieto voi tuoda mukanaan tarpeettomia kustannuksia.

5.3 Yritys B

Yritys B on monessa maassa toimiva iso konserni, joka käsittelee suuria datamääriä päivittäin ja hyödyntää laajasti erilaisia dataintegraatoratkaisuja ja automatisointia datan jakamisessa ja liiketoiminnan tehostamisessa. Heidän tarjoamansa dataintegraatoratkaisut ovat aina räätälöityjä asiakaskohtaisesti. Asiakkaat voivat ostaa valmiita integraatioita tai löytää muiden yritysten tekemiä integraatioita tai palveluja yrityksen verkkokaupasta.

Integraatoratkaisut

Yritys B:n paljon käyttämiä dataintegraatoratkaisuja ovat muun muassa Friends ja FlashNode, jotka mahdollistavat integraatioiden nopean toteutuksen olemassa olevien rajapintojen päälle. Friends on suomalainen integraatioalusta, joka mahdollistaa eri järjestelmien ja sovellusten saumattoman yhteistoiminnan ja tiedonvaihdon (HiQ Finland Oy 2022).

Lisäksi yritys on kehittänyt omia räätälöityjä integraatioita, jotka helpottavat datan siirtoa sisäisesti yrityksen omien ohjelmistojen välillä. Robotiikkaa, kuten robotiikkaprosessiautomaation työkalua UiPathia, hyödynnetään integraatioiden automatisoinnissa, mikä on tuonut huomattavia hyötyjä aikasäästönä, vähentyneinä virheinä ja tehokkuuden kasvuna.

Yritys B kehittää rajapintoja GraphQL API:n avulla, joka tukee hierarkkisia rakenteita ja mahdollistaa tarvittavan datan tehokkaan pyytämisen ja hyödyntämisen. API-liikennettä seurataan tarkasti sen varmistamiseksi, että se pysyy optimaalisella tasolla. Lisäksi yritys käyttää SnowFlow-työkalua käyttöanalytiikassa ja Tableau-työkalua datan visualisoinnissa.

Integraatoratkaisujen hyödyt ja haasteet

Dataintegraatiot ovat olleet keskeinen tekijä yrityksen monimutkaisten tietolähteiden parantamisessa ja uusien palveluiden tarjoamisessa. Yrityksen luoma dataekosysteemi, jossa kehittäjillä on pääsy erilaisiin API-rajapintoihin, on avannut mahdollisuuksia ulkopuolisille toimijoille, jotka ovat luoneet innovaatioita ja uusia liiketoiminta-alueita yhdistämällä eri tietolähteitä. Tämä on tuottanut monipuolista lisäarvoa ilman merkittäviä lisäinvestointeja. Lisäksi yritys B on saavuttanut säästöjä, kun se ei ole joutunut kehittämään kaikkia sovelluksia itse. Ulkopuolisten sovellukset eivät ainoastaan rikasta dataekosysteemiä, vaan tukevat ja täydentävät yrityksen palveluita, laajentavat käyttömahdollisuuksia ja tuottavat taloudellista hyötyä. Lisäksi datan integroinnin epäsuorat hyödyt ovat osoittautuneet merkittäviksi ja ovat vahvistaneet yrityksen menestystä.

Haasteita datan jakamisessa ja integroinnissa tuovat esimerkiksi robottikäyttäjien toimivuus ja käyttöoikeuksien hallinta sekä automaatioon liittyvät poikkeukset, joihin automaattiset hälytykset ovat toivottuja. Yritys B pyrkii jatkuvasti kehittämään ja optimoimaan datayhteistyön ja integraatioiden prosesseja, reagoiden nopeasti liiketoiminnan vaatimuksiin. Datan laadukkuus ja käyttökelpoisuus sekä automatisoitujen prosessien oikeellisuus ja luotettavuus varmistetaan testiautomaation ja seurannan avulla. Yritys B seuraa, miten integraatiot on tehty ja miten ne toimivat kumppaneiden kanssa tehdyissä sopimuksissa.

API-liikennettä seurataan tarkasti sen varmistamiseksi, että se pysyy optimaalisella tasolla. Rajapintojen käyttöä rajoitetaan tarvittaessa suorituskyvyn optimoimiseksi ja myös tietoturvasyistä. Asiakkaille tarjotaan tukea ja työkaluja datan integrointiin. Yrityksen asiakkaat haluavat yleensä valmiita integraatioita, jotka toimivat saumattomasti heidän järjestelmiensä kanssa. Uusien tietolähteiden lisääminen yrityksen dataekosysteemiin on suhteellisen nopeaa, jos toinen osapuoli tietää tarvittavat asiat ja tarpeet.

Tietoturvan ja tietosuojaan huomioiminen datan jakamisessa

Yritys B korostaa, että suorituskyvyn lisäksi dataintegraation teknisessä toteutuksessa on tärkeää varmistaa turvallisuus. Yritys rajoittaa rajapintojen käyttöä tarvittaessa suorituskyvyn optimoimisen lisäksi myös tietoturvasyistä. Yritys painottaa myös lainsäädännön ja tietosuoja-asetuksen (GDPR) huomioimista datan integroinnissa. Asiakkaiden tulee säilyttää tietoa määrätty ajat, ja yritys B varmistaa, että dataa käsitellään ja säilytetään oikein.

Dataintegraatioiden vaikutukset yritykselle

Yritys B:n laaja dataekosysteemi, vahvat kumppanuudet, tekninen osaaminen sekä laadukkaat, turvalliset ja monipuoliset integraatoratkaisut ovat olleet avainasemassa

yrittäjien menestyksessä. Dataintegraatiot mahdollistavat laajentumisen uusille toimialoille ja avaavat ovia uusille liiketoimintamahdollisuuksille.

Yrittäjien asiakkaat arvostavat tarjottua tukea ja työkaluja datan integrointiin. He odottavat valmiita integraatioita, jotka toimivat saumattomasti omien järjestelmiensä kanssa. Yrittäjien jatkuvat panostukset teknisiin ratkaisuihin ja yhteistyökumppanuuksiin varmistavat menestyksen jatkumisen. He tavoittelevat myös tekoälyn hyödyntämistä tulevaisuudessa.

Yrittäjien suositukset muille yrittäjille

Yrittäjä B korostaa datan tarpeellisuuden harkintaa ja tarpeettoman datan keräämisen välttämistä. Ylimääräisen datan keräämisestä voi aiheutua turhia kustannuksia ja GDPR-haasteita. Yrittäjä B suosittelee integraatioiden jatkuvaa kehittämistä sekä automatisointia, koska se tuo merkittävää ajan säästöä ja virheiden vähentämistä. Yrittäjä suosittelee erilaisten teknologioiden käyttöä integraatioiden ja automaation helpottamiseksi. Yrittäjä B muistuttaa, että ulkoistaessa datan integrointia, yrittäjien tulisi huomioida kumppanin asiantuntemus, palvelun kattavuus ja tietoturva.

5.4 Yrittäjä C

Yrittäjä C on suuri yritys, joka on omalla alallaan edelläkävijä datan hyödyntämisessä ja kansallisella tasolla yritys on pitkällä dataekosysteemien rakentamisessa. Yrittäjällä on melko suljettu, mutta moninainen ja laaja dataekosysteemi. Suurin osa yrittäjien datan jakamisesta tapahtuu yrittäjien sisäisen dataekosysteemin puitteissa, mutta yritys on integroitunut myös kansallisiin järjestelmiin. Yrittäjien tavoitteena on tarjota laadukkaita palveluita asiakkailleen.

Integraatoratkaisut

Yrittäjä pyrkii toteuttamaan dataintegraatioitaan uudelleen käytettävien palveluiden, API-integraatioiden ja integraatioalustojen avulla. Yrittäjien lähtökohtana on "API first" -ajattelu, jossa rakennetaan uudelleen käytettäviä komponentteja ja palveluita. Näin yritys pyrkii skaalautuvuuteen, tehokkaaseen kehitykseen ja nopeaan markkinamuutoksiin reagointiin.

Datan hallinta on keskeisessä roolissa, jotta varmistetaan tietojen yhdenmukaisuus ja oikeellisuus eri järjestelmissä. Integraatiot ja yhteydet kaikista järjestelmistä on rakennettu tietovarastoon (engl. data warehouse), jossa datan keskitetty käsittely tapahtuu. Yrittäjä huomioi toimialakohtaiset standardit. Yrittäjä pyrkii helpottamaan integraatiokyvykkyyttään pyrkimällä varautumaan tulevaisuuden haasteisiin.

Integraatoratkaisujen hyödyt ja haasteet

Yritys C kokee dataintegraatioiden hyödyt erittäin merkittäviksi ja integraatiot välttämättömiksi heidän liiketoiminnallensa, koska dataintegraatiot ovat keskeisessä roolissa digitaalisten palveluiden kehittämisessä. Kustannustehokkuus on yksi tärkeimmistä hyödyistä, kun kehittäjien työt tehostuvat ja nopeutuvat ja digitaaliset palvelut skaalautuvat helpommin. Lisäksi integraatiot ovat tuoneet yritykselle kilpailuetua, koska integraatioiden avulla yritys pystyy kehittämään asioita nopeammin ja reagoimaan markkinamuutoksiin tehokkaammin.

Yksi dataintegraatioiden suurimmista haasteista on toimittajien välinen yhteistyö, sillä integraatioiden onnistumisen kannalta vaaditaan kaikkien toimittajien tiivistä yhteistyötä ja sitoutuneisuutta. Toimittajien järjestelmien on oltava yhteensopivia, jotta integraatiot toimivat sujuvasti. Teknologiaavaintojen tekeminen on merkittävä ja haastava päätös, koska teknologia kehittyy nopeasti, ja päätökset sitovat yritystä pitkällä aikavälillä ja vaikuttavat kauaskantoisesti yrityksen toimintaan ja muutoskyvykkyyteen. Muutoskyvykkyyden parantaminen onkin yksi tulevaisuuden suunnitelmista integraatioarkkitehtuurin osalta.

Tietoturvan ja tietosuojaan huomioiminen datan jakamisessa

Yrityksessä C:ssä tietoturva ja tietosuoja ovat erittäin tärkeitä, ja niitä pidetään kilpailuetuna. Heillä on tarkat prosessit ja toimintamallit tietoturvan ja tietosuojaan varmistamiseksi. Jokainen yrityksen työntekijä on vastuussa tietoturvan ja tietosuojaan noudattamisesta ja toimii vastuullisesti tässä asiassa, eikä tietoturvaa ole ulkoistettu. Tietoturvan ja tietosuojaan huomioiminen on erityisen tärkeää, kun jaetaan sensitiivistä dataa. Yritys noudattaa tarkasti lainsäädäntöä ja pitää asiakasdatan turvassa.

Dataintegraatioiden vaikutukset yritykselle

Tehokkaasti ja järkevästi toteutetut dataintegraatiot ovat mahdollistaneet yritykselle kustannustehokkaan liiketoiminnan ja liiketoiminnan kasvun sekä kilpailuedun. Integraatioiden avulla yritys on pystynyt kehittämään digitaalisia palveluita ja reagoimaan nopeasti markkinamuutoksiin, mikä on mahdollistanut uusien palveluiden ja innovaatioiden kehittämisen. Asiakkaille integraatiot ja niiden kehittäminen ovat tuoneet sujuvampaa ja parempaa asiakaskokemusta.

Yrityksen suositukset muille yrityksille

Yritys C suosittelee lähtökohtaisesti dataintegraatiohankkeiden rakentamista kestäväälle ja skaalautuvalle arkkitehtuurille. Yritys suosittelee pistemäisen integraatiomallin hyödyntämistä ainoastaan alustavassa arvontuoton todentamisessa, ja neuvoo heti sen

jälkeen siirtymään muihin joustavampiin dataintegraatiomalleihin, jotka tukevat laajentumista.

Yritys C korostaa tiedolla johtamisen kulttuuria, joka vaatii sekä teknologista että hallinnollista asiantuntemusta kaikilla organisaation tasoilla. Yritys painottaa, että johdon on ymmärrettävä ja näytettävä esimerkkiä siitä, kuinka tärkeää datan hyödyntäminen on liiketoiminnan menestykselle. Yritys C pitää tärkeänä, että tiedon jakamisen ideologia on syvälle juurtunut yrityskulttuuriin ja toiminnan ytimeen.

Kumppanuutta harkittaessa yritys C korostaa saman arvomaailman, visioiden, toimintatapojen ja kulttuurien jakamisen merkitystä. Yritys C:n mukaan kumppanin ei välttämättä tarvitse noudattaa identtisiä toimintamalleja, mutta kyky toimia muutosjohtajana ja ohjata organisaatiota kulttuurimuutoksessa on kriittisen tärkeää. Yritys painottaa myös teknologisten ratkaisujen yhteensopivuuden tärkeyttä. Yritys neuvoo valitsemaan ratkaisuja, jotka tukevat liiketoiminnan tavoitteita ja ovat yhteensopivia olemassa olevien järjestelmien kanssa.

5.5 Yritys D

Yritys D on pieni logistiikan alan yritys, joka operoi logistiikan dataekosysteemissä. Vaikka yritys on yksi pienimmistä aktiivisista toimijoista nykyisessä dataekosysteemissään, sillä on merkittäviä näkemyksiä ja kokemuksia datan integraatiosta ja sen haasteista. Yritys tarjoaa maksullista dataa dataekosysteemissä yrityksille, mutta samalla se tarjoaa loppukäyttäjille ilmaista dataa, jos he käyttävät yrityksen sovellusta.

Integraatoratkaisut

Yrityksen integraatoratkaisujen ytimessä on ArcGIS Online, joka mahdollistaa tietojen siirtymisen eri järjestelmiin ja tarjoaa monipuolisia työkaluja. Yrityksen valinta perustuu alustan kykyyn tarjota niin kutsuttuja low code -ratkaisuja, mikä minimoi koodauksen tarpeen.

ArcGIS Online on Esrin pilvipohjainen kartta- ja paikkatietopalvelu, joka mahdollistaa karttojen ja geotietosovellusten luomisen, jakamisen ja hallinnoinnin verkkoympäristössä. ArcGIS Online mahdollistaa interaktiivisten web-karttojen ja sovelluksien luomisen ilman koodausta. ArcGIS Onlinessa on myös Living Atlas, joka on kokoelma globaaleja, valmiiksi tehtyjä kartta-aineistoja ja tietoja, jotka ovat Esrin ja muiden käyttäjien tuottamia. ArcGIS Online on helppo integroida muihin Esrin tuotteisiin ja se integroituu myös ulkopuolisiin sovelluksiin ja palveluihin.

Yritys käyttää muun muassa ArcGis Online -paikkatietopalvelusta löytyviä WMS- ja WFS-rajapintoja sekä GeoJson-dataformaattia datan siirtämiseen. Yritys pyrkii valitsemaan järjestelmästään jaettavan datan siten, että se palvelee yrityksen omia liiketoimintatarpeitaan ja asiakasrajapintoja.

Integraatoratkaisujen hyödyt ja haasteet

Yritys kokee, että tällä hetkellä dataintegraatioiden hyödyt eivät ole vielä täysin realisoituneet yritykselle itselleen, vaikka yritys on dataekosysteemissä tuonut paljon hyötyjä sekä isommille toimijoille että loppukäyttäjille. Yksi merkittävä datan jakamisen tuoma hyöty on kuitenkin näkyvyys, joka mahdollistaa yrityksen D:n ratsastamisen dataekosysteemin aallonharjalla ja houkuttelemaan asiakkaita. Näkyvyyden eteen yritys pyrkii tekemään paljon töitä.

Yrityksen mukaan suurin haaste on datan laatu, joka ei aina ole heidän hallinnassaan. Huonolaatuinen data voi johtaa asiakasmenetyksiin. Lisäksi osa toimijoista saattaa käyttää datan integroinnissa edelleen manuaalista integraatiota, kuten esimerkiksi toimittaa CSV-tiedostoja, vaikka rajapintojen kautta datan siirto olisi tehokkaampaa. Dataintegraatioiden toteuttamiseen haasteita tuo myös erilaiset teknologiat ja järjestelmät. Yritys mainitsee haasteeksi myös API-integraatioiden dokumentaatioiden vaikeasti ymmärrettävyyden, sillä dokumentaatiot ovat usein vaikeasekoisia eivätkä kovin käyttäjäystävällisiä.

Yritys kokee, että dataekosysteemissä kilpailu voi aiheuttaa ongelmia erityisesti pienille ja pienenä yrityksenä toimiminen dataekosysteemissä on haastavaa, koska suuret yritykset saattavat viedä suurimman osan hyödyistä. Yritys korostaa, että dataekosysteemin tulisi olla vastavuoroista hyötyä kaikille osapuolille, eikä se saisi suosia suuria yrityksiä. Yrityksen kokemuksen mukaan isot pelaajat saattavat keskittyä toisten isojen toimijoiden kanssa yhteistyöhön ja jättää pienet yritykset vaille huomiota. Pienet toimijat joutuvat tekemään enemmän työtä ekosysteemin eteen. Yritys kertoo haasteeksi myös hyötyjen näyttämisen etukäteen potentiaalisille kumppaneille ja koko dataekosysteemin saamisen toimimaan yhteisen tavoitteen eteen. Yritys kokee erityisesti pienille yrityksille haasteelliseksi myös projektiluontoisuuden, ja yritys itse pyrkikin lisäämään toiminnassaan säännöllistä kuukausilaskutusta. On ollut haastavaa näyttää hyötyjä etukäteen potentiaalisille kumppaneille ja saada koko dataekosysteemi toimimaan yhteisen tavoitteen eteen.

Tietoturvan ja tietosuojan huomioiminen datan jakamisessa

Yritys on tarkka siitä, minkälaista dataa se jakaa. Yritys ei koskaan jaa aineistoja, joita se käyttää oman datan luomiseen. Yrityksellä on myös paljon asiakastietoa, jota he eivät jaa eteenpäin.

Yritys huomioi tietosuojan ja tietoturvan Esri:n tuotteiden avulla. Esri noudattaa useita alan standardeja ja parhaita käytäntöjä varmistaakseen palvelun ja sen sisällön tietoturvan. Esri pyrkii suojaamaan palveluitaan tunkeutumisilta, haittaohjelmilta ja uhilta. ArcGIS Onlinessa käyttäjät voivat määrittää, kuka näkee heidän sisältönsä ja määrittää sen julkisuustason. ArcGis Online kerää vähimmäismäärän henkilökohtaisia tietoja tarjotakseen palveluitaan, eikä jaa näitä tietoja ilman käyttäjän suostumusta. Vaikka ArcGIS Online on pilvipalvelu, Esri tarjoaa tietoa siitä, missä datakeskuksissa tietoja säilytetään.

Dataintegraatioiden vaikutukset yritykselle

Dataintegraatioiden kautta yritys D on saanut mahdollisuuden kasvattaa liikevaihtoaan dataekosysteemissä, mutta se on vaatinut paljon aktiivista työtä ja ajatustyötä, erityisesti, kun otetaan huomioon yrityksen pieni koko. Kuitenkin haasteet, kuten datan laatu ja kilpailu isojen toimijoiden kanssa, ovat vaikuttaneet heidän kykyynsä hyötyä täysin dataekosysteemistä. Yritys toivoo, että tulevaisuudessa dataekosysteemin hyödyt tulisivat paremmin esille ja realisoituisivat tehokkaammin. Yritys näkee tulevaisuudessa mahdollisuuden laajentaa toimintaansa ulkomaille ja tarjota samoja palveluja kuin Suomessa.

Yrityksen suositukset muille yrityksille

Yritys D suosittelee, että yrityksen tulisi ensin miettiä, onko sillä jaettavaksi sopivaa dataa ja onko jakaminen teknillisesti mahdollista. Yritys painottaa datan laadun merkitystä ja suosittelee yrityksiä varmistamaan, että jaettava data on riittävän laadukasta.

Yritys korostaa merkitystä ymmärtää, miten saavuttaa hyötyä dataintegraatioiden avulla dataekosysteemissä ja kuinka rakentaa toimiva liiketoimintamalli sen ympärille. Yrityksen mukaan on tärkeää tunnistaa ja määritellä molemminpuoliset hyödyt yhteistyössä. Yritys kehottaa pieniä yrityksiä olemaan aktiivisia ja innovatiivisia datan jakamisessa ja integraatiossa sekä etsimään aktiivisesti yhteistyömahdollisuuksia isojen toimijoiden kanssa.

Yritys korostaa, kuinka tärkeää on pitää oma datansa hallussaan. Yritys pitää tärkeänä myös olla teknologiariippumaton, jotta se voi sopeutua ja tehdä yhteistyötä eri toimijoiden kanssa. Lisäksi yritys kannustaa tutkimaan API-suunnittelutyökalujen ja hallintaratkaisujen käyttöä integraatoratkaisujen helpottamiseksi.

5.6 Yritys E

Yritys E on suuri valtion omistama alansa dataekosysteemin kehittämiseen ja digitalisaatioon keskittyvä toimija. Yrityksen tavoitteena on laajentaa datan ekosysteemiä ja

lisätä datan saatavuutta, laatua ja yhteen toimivuutta, samalla pyrkien alentamaan datan hyödyntämisen kustannuksia. Yrityksen ekosysteemi- ja teknologiajohtaja antoi haastattelussa arvokasta tietoa yrityksen dataintegraatoratkaisuista ja siihen liittyvistä näkökohdista.

Integraatoratkaisut

Yritys E käyttää monipuolisia dataintegraatoratkaisuja dataekosysteemissään. Sisäisesti yritys hyödyntää dataintegraatioissaan MuleSoftia, joka mahdollistaa eri järjestelmien ja sovellusten yhdistämisen API:en kautta. Yrityksen ulkoisissa dataintegraatioissa käytetään AWS-pilvipalveluita, pistemäisiä integraatioita ja API-rajapintoja. AWS-pilvipalvelu mahdollistaa datan vastaanottamisen toimijoiden API-rajapinnoista. Tieto tallennetaan AWS-palvelun dataintegraatioalustalle, ja sitä jaetaan edelleen eri sovellusten ja API-rajapintojen kautta. Yritys hyödyntää myös MQTT-jonoa ja AWS IoT 1-Click -ratkaisuja reaaliaikaisen datan jakamiseen. Vaikka AWS-pilvipalvelu on rahallisesti merkittävä investointi, yritys kokee palvelun tuovan merkittävää positiivista vaikuttavuutta.

Integraatoratkaisujen hyödyt ja haasteet

Yritys E:n integraatoratkaisujen suurin etu on niiden kyky luoda yhteiskunnallista vaikuttavuutta ja mahdollistaa liiketoimintaa muille toimijoille dataekosysteemissä. Dataintegraatioiden avulla datan jakaminen eri toimijoiden kanssa edistää muun muassa turvallisuutta, sujuvuutta, kustannusten optimointia ja uusien liiketoimintamahdollisuuksien luomista.

Yritys E kertoo haasteita datan integroinnissa tuovan se, jos toimijoiden data on vanhoissa järjestelmissä ja vaatii paljon puhdistamista ja muokkaamista. Yritys kokee, että haasteena datan integroinnissa on myös yhteisten tavoitteiden ja vision luominen dataekosysteemille sekä arvon tarjoamisen riittävän nopeasti dataekosysteemin kaikille osallistujille. Koska hyödyt saattavat tulla vasta myöhemmin, pienille yrityksille on vaikea näyttää konkreettisia etuja lyhyellä aikavälillä. Pienet yritykset kaipaavat usein nopeaa tulosta, mikä voi olla haastavaa saavuttaa dataekosysteemin rakentamisen alkuvaiheessa. Haasteena erityisesti pienille toimijoille on yrityksen mukaan datan jakamisen kustannusten ja hyötyjen tasapainottaminen.

Tietoturvan ja tietosuojan huomioiminen datan jakamisessa

Tietosuojan ja tietoturvan varmistaminen on yritys E:lle tärkeää, ja yritys pyrkii suojaamaan datansa asianmukaisilla toimenpiteillä. Yritys suorittaa riskianalyysejä ja yrityksellä on kyvykyys rajoittaa datan jakamista tietyissä tilanteissa, esimerkiksi maantieteellisesti. Yritys itse pyrkii noudattamaan EU-tason standardeja kansallisten sijaan, jos mahdollista.

Dataintegraatioiden vaikutukset yritykselle

Dataintegraatiot ovat mahdollistaneet yhteistyön eri toimijoiden välillä ja tuoneet mukanaan sekä suoria että välillisiä yhteiskunnallisia vaikutuksia. Yritys E:n mukaan sen tavoitteena ei ole ensisijaisesti tehdä voittoa, vaan tuottaa vaikuttavuutta ja mahdollistaa muiden toimijoiden menestymisen dataekosysteemissä. Yrityksen jakaman datan vaikuttavuus Suomessa on merkittävä, ja dataintegraatiot ovat tuoneet mukanaan sekä suoria että välillisiä yhteiskunnallisia vaikutuksia, kuten verotulojen kasvua ja uusien palveluiden syntymistä.

Yrityksen suositukset muille yrityksille

Yritys E suosittelee ensin selvittämään, mitkä ovat dataintegraatioista kiinnostuneen yrityksen arvokkaimmat datasetit muiden näkökulmasta ja ketkä voisivat hyötyä niistä. Yritys korostaa myös datan laadun tärkeyttä. Yritys kehottaa varmistamaan ennen datan jakamista, että jaettava data on korkealaatuista ja relevanttia. Yritys painottaa, että datan jakamista tulisi harkita loppukäyttäjien kannalta.

Yritys suosittelee tekemään ja dokumentoimaan selkeät visiot, tavoitteet ja odotuksen datan jakamiselle, kuuntelemaan tässä sidosryhmiä ja mukauttamaan strategiaa heidän tarpeidensa mukaan. Toisaalta yritys muistuttaa, että kaikkia ei voi koskaan miellyttää.

Yritys painottaa, kuinka tärkeää on pyrkiä minimoimaan datan jakamisen kustannukset. Yritys suosittelee ensin kokeilemaan edullisempaa integraatoratkaisua ja arvioimaan sen hyötyjä ennen suurempaan investointiin sitoutumista.

Yritys E korostaa, että datan jakaminen voi luoda liiketoimintaa ja hyödyttää yritystä, mutta se tulee tehdä harkiten ja kustannustehokkaasti.

5.7 Yritys F

Yritys F on suuri yritys, jolla on pitkä kokemus erilaisista dataintegraatoratkaisuista. Haastattelimme yrityksen data platform -konsulttia. Yritys auttaa muita suuria yrityksiä suunnittelemaan ja rakentamaan erilaisia data-alustoja (engl. data platform).

Integraatoratkaisut

Teknisinä ratkaisuuina yrityksellä on käytössään erilaisia työkaluja, kuten ETL (Extract, Transform, Load) -työkaluja ja API-integraatiotyökaluja, kuten MuleSoft ja Apigee. ETL-työkalut mahdollistavat yksisuuntaiset integraatiot järjestelmien välillä, kun taas API-integraatiot tarjoavat kaksisuuntaisen operatiivisuuden ja mahdollistavat monimutkaisempia toiminnollisuuksia.

Lisäksi yritys käyttää pilvipohjaisia ratkaisuja, kuten Microsoft Azure Cloud, AWS (Amazon Web Services) Cloud ja Google Cloud Platform. Pilvipalvelut tarjoavat yritykselle kustannustehokasta tallennustilaa ja laskentatehoa, mikä on tärkeää datan hallinnassa ja prosessoinnissa.

Integraatoratkaisujen hyödyt ja haasteet

Yritys F:n konsultin mukaan dataintegraatioilla voi olla merkittäviä taloudellisia hyötyjä, kun dataekosysteemi alkaa toimia ja yhteyksiä syntyy. Yritys F:n edustaja näkee dataintegraatioiden suurimana hyötynä rahallisen arvon tuottamisen. Toimivat dataintegraatiot voivat tuoda myös merkittäviä säästöjä yrityksille pidemmällä aikavälillä, vaikka dataintegraatiot ovat usein suuri investointi.

Haasteena yritys mainitsee erityisesti vaikeudet kytkeytyä erilaisiin lähdejärjestelmiin, joissa voi olla vanhaa ja monimutkaista dataa. Lisäksi kulttuurin muutos yrityksessä voi olla hidasta, mutta se vaaditaan, jotta datan jakaminen saataisiin osaksi kaikkea toimintaa. Osaajien löytäminen ja datan kontekstisidonnaisuus tuovat myös omat haasteensa.

Tietoturvan ja tietosuojan huomioiminen datan jakamisessa

Yrityksen edustaja korostaa tietoturvan ja tietosuojan tärkeyttä datan jakamisessa. Edustaja kertoo, että tietoturva- ja tietosuoja-asioita käsitellään huolellisesti tietoturvatiimissä (engl. security team) ja tietosuojatiimissä (engl. data privacy team) ennen kuin dataintegraatioita lähdetään toteuttamaan.

Dataintegraatioiden vaikutukset yritykselle

Dataintegraatiot voivat tuoda yritykselle merkittävää taloudellista hyötyä. Yrityksen konsultti korosti muutosjohtamisen ja kulttuurin muutoksen merkitystä, jotta datan jakaminen saataisiin osaksi yrityksen toimintaa ja yrityskulttuuria.

Yrityksen konsultti nosti datatuotteiden kehittämisen keskeiseksi tekijäksi yrityksen kulttuurin muutoksen yhteydessä, sillä datatuotteet voivat olla tehokkaita välineitä datan arvon optimoinnissa. Datatuotteet ovat käytännön sovelluksia, kuten interaktiivisia raportteja ja visualisointeja, jotka hyödyntävät yrityksen käytössä olevaa dataa ja auttavat siinä, että dataa hyödynnetään parhaalla mahdollisella tavalla liiketoiminnan päätöksen teossa ja kehittämisessä.

Yrityksen suositukset muille yrityksille

Yritys F:n konsultin kokemuksen perusteella dataintegraatioiden onnistumisen avaimet ovat selkeä visio, kulttuurin muutos ja hyvin harkitut investoinnit. Samalla tärkeää on huomioida

tietoturva ja tietosuojat. Konsultti korosti, että datatiimin ja integraatiotiimin tulisi toimia erillisinä yksiköinä, mutta niiden välille tulisi rakentaa tiivis yhteistyö.

Konsultti muistuttaa, että dataintegraatio yrityksen sisällä on ensin saatava toimimaan ennen kuin voidaan miettiä datan jakamista ulospäin. Konsultti suosittelee yrityksiä kartoittamaan ja ymmärtämään, miksi dataa he haluavat jakaa dataa ja miten datan jakamisesta syntyy arvoa. Lisäksi konsultti korostaa, kuinka tärkeää on, että datan tulisi olla helposti hyödynnettävässä muodossa. Konsultti kehottaa yrityksiä tunnistamaan datan tuottajat ja kuluttajat sekä miettimään, miten nämä osapuolet tunnistetaan. Konsultti korostaa riskien minimoimisen tärkeyttä. Konsultin mukaan tulisi lähteä liikkeelle pienistä asioista ja arvon tuotosta, ja vasta sen jälkeen skaalata toimintaa isommaksi, kun hyötyjä on todennettu.

Konsultti suosittelee yrityksille data mesh -ajattelutapaa, koska se tarjoaa tehokkaan tavan hallita ja hyödyntää dataa organisaatiossa. Data meshin perusajatus on hajauttaa datan omistajuus ja hallinta eri liiketoiminta-alueille ja tiimeille organisaatiossa. Tämä hajautettu lähestymistapa mahdollistaa datan tehokkaan hyödyntämisen eri käyttötarkoituksiin.

Data mesh -ajattelutavan mukaisesti konsultti painotti erityisesti yhteistyön tärkeyttä liiketoiminnan ja IT-osaston välillä datatuotteiden kehittämisessä. Datan tuotteistamisessa luodaan rajapintoja ja dokumentaatiota, ja data tarjotaan muiden tiimien käyttöön valmiina datatuotteena. Tällä tavoin dataa voi hyödyntää tehokkaasti eri tarkoituksiin, ja se on helpommin saatavilla päätöksenteon tueksi liiketoiminnassa

6 Yrityksissä esiin tulleet dataintegraatioiden hyödyt ja haasteet

6.1 Yritysten esiin tuomat dataintegraatioiden hyödyt

Yrityshaastatteluissa nostettiin esille useita dataintegraatioiden tuomia hyötyjä. Dataintegraatioilla voi olla merkittäviä taloudellisia hyötyjä ja toimivat dataintegraatiot voivat tuoda merkittäviä säästöjä yrityksille pidemmällä aikavälillä, vaikka alkuinvestoinnit saattavat olla suuria.

Hyvälaatuinen data voi antaa kilpailuedun yritykselle markkinoilla. Oikeanlaisten dataintegraatioiden avulla yritykset voivat hyödyntää datan jakamista tehokkaasti, optimoida toimintaansa, lisätä tehokkuutta ja saavuttaa liiketoiminnallisia etuja. Dataintegraatioiden avulla yritykset voivat parantaa palvelujensa laatua ja tehokkuutta. Integroidun datan avulla yritykset voivat tehdä parempia, tarkempia ja täsmällisempiä liiketoimintapäätöksiä. Dataintegraatioiden avulla yritykset saavat enemmän tietoa asiakkaista ja ne voivat auttaa parantamaan asiakaskokemusta. Parempi asiakaskokemus ja monipuolisemmat palvelut voivat parantaa asiakaspitoa.

Dataintegraatioiden avulla voidaan vähentää manuaalista työtä ja automatisoida prosesseja, mikä johtaa kustannussäästöihin ja lisää tehokkuutta sekä tarkkuutta. Datan jakaminen voi tuoda yritykselle uusia kumppanuuksia, verkostoitumismahdollisuuksia sekä liiketoimintamahdollisuuksia. Lisäksi dataintegraatiot voivat tuoda yritykselle parempaa näkyvyyttä yhteisen markkinoinnin kautta. Dataintegraatiot tuovat myös paljon epäsuoria hyötyjä. Kuten eräs haastatelluista yrityksistä totesi, dataintegraation kannattavuus ei aina ole suoraviivaista, mutta sen epäsuorat hyödyt, kuten ulkopuolisten kehittäjien innovaatiot, voivat olla merkittäviä.

6.2 Yritysten esiin tuomat dataintegraatioiden haasteet

Dataintegraatiot tarjoavat merkittäviä mahdollisuuksia liiketoiminnan kehittämisessä, mutta niihin liittyy myös haasteita, jotka ovat tärkeitä tunnistaa ja hallita. Yhtenä suurimpana haasteena yritykset nostivat esille kulttuurin muutoksen aikaansaamisen hitauden. Kulttuurin muutos yrityksessä on kuitenkin välttämätöntä, jotta datan jakaminen integroituu kaikkeen toimintaan.

Muut keskeisimmät haastatteluissa esille nousseet haasteet ovat lähdejärjestelmien monimuotoisuus ja yhteensopivuus, datan laadun hallinta ja reaaliaikaisuuden ylläpitäminen. Kuten eräässä haastattelussa korostettiin, pienetkin viiveet datan päivityksessä voivat heikentää merkittävästi sen laatua ja sovellusten toimivuutta. Datan keräämisen prosessin ongelmat näkyvät datan laadussa. Huono datan laatu vie helposti

dataintegraation hyödyt ja voi vääristää päätöksentekoa. Jos rajapinnasta tulee huonolaatuista dataa, se vaatii manuaalista läpikäymistä ja saattaa johtaa siihen, että joudutaan rakentamaan koko järjestelmä uudelleen. Testiautomaation ja seurannan avulla yritykset voivat varmistaa datan laadun ja käyttökelpoisuuden sekä automatisoitujen prosessien oikeellisuuden ja luotettavuuden.

Dataintegraation haasteena on myös teknisen osaamisen puute. Dataintegraation toteuttamiseen yritykset tarvitsevat osaavia ammattilaisia. Pienillä yrityksillä voi olla rajoitettu määrä IT-resursseja, mikä voi vaikuttaa dataintegraation toteuttamiseen. Yritykset kokivat useat API-integraatioiden alkuperäiset dokumentaatiot vaikeaselkoisiksi, mikä vaikeuttaa rajapintojen perustamista.

Dataintegraatoratkaisujen valinnat sitovat pitkäksi aikaa ja dataintegraation kustannukset voivat olla merkittäviä, erityisesti monimutkaisissa projekteissa. Tästä syystä dataintegraatoratkaisujen valintaa tulisi harkita ja selvittää huolella ennen dataintegraation investointia, ja tehtyjen päätösten on tuettava yrityksen pitkän aikavälin tavoitteita ja muutoskyvykkyyttä. Yritysten on oltava myös valmiita reagoimaan nopeasti liiketoiminnan vaatimuksiin ja kehittämään dataintegraatioitaan jatkuvasti. On tärkeää, että yritykset miettivät tarkkaan, mitä dataa tallennetaan ja mitä ei. Tulevaisuuden datan käyttötarpeiden ennustaminen on monimutkaista, ja virheelliset päätökset tässä vaiheessa voivat johtaa ongelmiin myöhemmin.

Dataintegraation onnistumisen haasteena nähtiin myös yhteistyön ongelmat toimijoiden välillä. Yritysten on tärkeää luoda yhteiset tavoitteet ja visio dataekosysteemilleen. Sopimuskysymykset ja ehdot voivat aiheuttaa dataintegraatioissa haasteita, ja niistä tulisi sopia huolella. Pienille yrityksille dataekosysteemissä toimiminen voi olla erityisen haastavaa, koska suuret yritykset saattavat hallita suurinta osaa resursseista ja hyödyistä. Pienet yritykset voivat kokea tulevansa sivuutetuksi ja joutuvansa tekemään suhteettoman paljon työtä ekosysteemin ja sen arvon tuoton hyväksi. Lisäksi erityisesti pienet yritykset kaipaavat usein konkreettisia tuloksia nopeasti, ja niiden voi olla haastavaa nähdä hyötyjä pitkällä tähtäimellä.

7 Yrityksissä esiin tulleet parhaat käytänteet dataintegraatiossa

Dataintegraatioiden parhaiden käytänteiden (engl. best practices) on havaittu auttavan yritystä saamaan enemmän arvoa datastaan ja välttämään yleisimpiä ongelmia, kuten datan epäluotettavuutta tai tietoturvariskejä. Suunniteltu ja tehokas dataintegraatio voi olla merkittävä kilpailuetu nykyaikaisessa liiketoimintaympäristössä. Seuraavassa esitetään yrityshaastattelujen pohjalta viisi parasta käytännettä dataintegraatioprosessissa:

1. Selkeä strategia ja visio: Ennen kuin ryhdytään mihinkään datan integraatiotoimiin, on tärkeää luoda selkeä strategia varmistamaan, että dataintegraatiot palvelevat liiketoiminnan periaatteita. Yrityksen tulee selvittää, mitä dataa kerätään, mitä varten sitä kerätään sekä miten se tallennetaan ja jaetaan yrityksen eri osastojen ja järjestelmien kesken tai ulkopuolisille toimijoille.
2. Datan laadun varmistaminen: Dataintegraatioprosessissa on kiinnitettävä erityistä huomiota datan laadun valvontaan ja parantamiseen, jotta datalla voisi luoda arvoa ja datasta olisi hyötyä. Datan laadun parantaminen sisältää datavirheiden tunnistamista, korjaamista ja datan puhdistamista tarvittaessa.
3. Datan standardointi ja yhdenmukaistaminen: Datan monimuotoisuuden hallinta voi olla haastavaa ja lähdejärjestelmät ovat moninaisia. Datan standardointi ja yhdenmukaistaminen ovat kriittisiä, jotta eri järjestelmien ja lähdejärjestelmien data voidaan ymmärtää ja käyttää yhtenäisesti.
4. Tietoturva ja tietosuojat: Yritysten on varmistettava, että dataintegraatioprosessit noudattavat asianmukaisia tietoturvastandardeja ja että arkaluontoisia tietoja käsitellään huolellisesti ja tietosuojalainsäädännön mukaisesti. On tärkeää, että yritykset tunnistavat, millainen tieto on olennaista eri sidosryhmille ja arvioivat sen asianmukaisen suojan tarpeen.
5. Dataintegraatioiden testaus ja seuranta: Uusien dataintegraatioiden käyttöönottoa edeltää kattava testaus varmistaa niiden toimivuuden. Lisäksi on tärkeää seurata dataintegraatioita jatkuvasti, havaita ongelmat ja viiveet ajoissa sekä varmistaa, että ne toimivat odotetusti.

8 Dataintegraatiototeutusten vertailu

8.1 Dataintegraatiomallien ominaisuuksien vertailu

Dataintegraatiomalleilla on erilaisia ominaisuuksia, jotka voivat vaikuttaa niiden soveltuvuuteen eri tilanteissa. Dataintegraatiomallien ominaisuudet ovat keskeisiä tekijöitä dataintegraatiototeutusten valinnassa ja niiden suorituskyvyn arvioinnissa. Taulukko 1 antaa yleiskuvan eri integraatiomallien keskeisistä ominaisuuksista, arvioituna asteikolla 1-5 '+'-merkeillä. Taulukon vertailu on toteutettu kirjallisuuskatsauksen ja haastattelututkimuksen pohjalta.

Integraatiomalli / Ominaisuus	Manuaalinen integraatio	Robottivusteinen ohjelmisto-automaatio	Pistemäinen integraatio	Datan jakeluun perustuva integraatio	Yrityksen laajuinen integraatio (ETL/ESB)	API-integraatiot
Joustavuus	+	++	+++	++++	+++++	++++
Skaalautuvuus	+	++	++	++++	+++++	+++
Datan siirtonopeus	+	+++	++	+++	++++	+++++
Yksinkertaisuus	+++++	+++	+++++	+++	++	++++
Teknologia-orientoituneisuus	+	+++	++	++++	+++++	+++++
Kustannus-tehokkuus	+ .. ++	++++	++ .. +++	++ .. +++	+ ..++++	++++
Tietoturva ja tietosuojaja	+	+++	++	++++	+++++	++++
Monimutkaisuuden hallinta	+	++	++	+++	+++++	+++
Reaaliaikaisuus	+	+++	++++	+++	+++++	+++++
Vikasietoisuus	+	+++	++	++++	+++++	+++ ..++++
Ylläpidettävyys	+	++	+++	+++	++++	+++
Soveltuvuus suurelle datamäärälle	+	+++	+	+++	+++++	++++

Taulukko 1. Dataintegraatiomallien ominaisuudet

Joustavuus

Joustavuus mittaa, kuinka helposti integraatio voi mukautua muutoksiin tai uusiin vaatimuksiin. Dataintegraatioiden joustavuus voi vaihdella paljon riippuen siitä, miten hyvin integraatio on suunniteltu ja toteutettu.

API-integraatiot ja yrityksen laajuiset integraatiot ovat kuitenkin parhaimpia integraatoratkaisuja joustavuudeltaan, kun taas manuaalinen integraatio ja pistemäinen integraatio eivät ole kovin joustavia. Pistemäinen integraatio voi olla joustavaa tietyissä tapauksissa, kuten yksinkertaisissa yhdistämisissä yhden tai kahden järjestelmän kanssa, mutta eivät sovellu hyvin monimutkaisemmille integraatioille. Datan jakeluun perustuva integraatio voi olla joustavaa, kunhan keskitetty keskus on hyvin suunniteltu ja kykenee yhdistämään erilaisia järjestelmiä. Datan jakeluun perustuva integraatio voi olla kuitenkin monimutkaisempi hallita ja ylläpitää.

Skaalautuvuus

Skaalautuvuus kuvastaa integraation kykyä kasvaa käyttäjämäärän tai datamäärän lisääntyessä. API-integraatiot ovat kaikkein skaalautuvimpia integraatoratkaisuja, kun taas manuaalinen integraatio ja pistemäinen integraatio ovat vähiten skaalautuvia.

Datan jakeluun perustuva integraatio voi olla skaalautuva, koska keskitetty navan hallinta helpottaa integraatioiden hallintaa ja ylläpitoa. Lisäksi uusien integraatioiden lisääminen siinä on suhteellisen helppoa. ETL-ratkaisut voivat olla skaalautuvia, koska ne mahdollistavat suuren datan erottamisen, muokkaamisen ja siirtämisen eri järjestelmien välillä. Kuitenkin suurien datamäärien käsittely voi aiheuttaa haasteita ETL-ratkaisun skaalautuvuudessa. ESB-ratkaisut voivat olla skaalautuvia, koska ne tarjoavat keskitetyn tavan hallita ja valvoa eri järjestelmien välistä viestintää. Lisäksi ESB:n avulla uusien integraatioiden lisääminen voi olla helpompaa. API-integraatiot ovat erittäin skaalautuva ratkaisu, koska ne mahdollistavat eri järjestelmien välisen suoran ja reaaliaikaisen viestinnän. API-integraatiot ovat helposti hallittavissa ja skaalattavissa tarpeiden mukaan.

Datan siirtonopeus

Datan siirtonopeus kertoo, kuinka nopeasti data voidaan siirtää järjestelmien välillä. API-integraatiot ovat yleisesti ottaen nopeimpia integraatoratkaisuja, kun taas manuaalinen integraatio on hitain vaihtoehto. Robottiväestöinen integraatio, ETL ja ESB ovat myös tehokkaita ja nopeita tapoja integroida järjestelmiä keskenään. Pistemäinen integraatio ja datan jakeluun perustuva integraatio voivat olla nopeita, mutta niiden nopeus riippuu käytettävistä integraatoratkaisuista.

Yksinkertaisuus

Manuaalinen integraatio on suhteellisen yksinkertainen integraatoratkaisu, mutta se voi olla virhealtista ja aikaa vievää, ja vaatii käyttäjältä paljon manuaalista työtä. Pistemäinen integraatio ja API-integraatiot ovat myös suhteellisen yksinkertaisia ratkaisuja, kun taas yrityksen laajuiset integraatiot ovat kaikista monimutkaisimpia ratkaisuja, jotka vaativat eniten teknistä osaamista ja aikaa.

Teknologiaorientoituneisuus

Teknologiaorientoituneisuus mittaa, kuinka modernia tai edistyksellistä käytetty teknologia on. Manuaalinen ja pistemäinen integraatio ovat vähiten teknologiaorientoituneita, kun taas ETL, ESB ja API-integraatiot ovat korkean teknologian ratkaisuja. Datan jakeluun perustuva integraatio ja robottivusteinen integraatio sijoittuvat näiden välimaastoon teknologiaorientoituneisuuden suhteen.

Kustannustehokkuus

Kustannustehokkuus integraatoratkaisuissa kertoo, kuinka taloudellisesti tehokas integraatio on suhteessa sen tarjoamiin hyötyihin. Se voi riippua useista eri tekijöistä, kuten yrityksen koosta, resurssien saatavuudesta ja henkilöstön osaamisesta. Manuaalinen integraatio voi olla edullinen pienissä mittakaavoissa, mutta se vaatii paljon aikaa ja resursseja, mikä aiheuttaa kustannuksia. Myös ihmisen virheet voivat aiheuttaa kustannuksia virheiden korjaamisessa eikä manuaalinen integraatio ole välttämättä kovin edullinen integraatoratkaisu.

Pistemäinen integraatio voi olla kohtuullisen kustannustehokas pienissä mittakaavoissa, mutta suuremmissa mittakaavoissa se voi vaatia enemmän resursseja ja aikaa eikä ole enää kustannustehokas ratkaisu. Datan jakeluun perustuva integraatio voi olla kohtuullisen kustannustehokas, mutta se vaatii jatkuvaa ylläpitoa ja päivityksiä sekä investointeja toimijaan (engl. hub), joten kovin pienissä integraatioissa se ei ole järkevä integraatoratkaisu. Yrityksen laajuiset integraatoratkaisut voivat aiheuttaa korkeita kustannuksia pienemmissä yrityksissä ja integraatioissa, mutta alkuinvestoinnin jälkeen voivat olla kustannustehokkaita suuremmissa integraatiotarpeissa.

Robottivusteinen integraatio voi olla erittäin kustannustehokas, kunhan robotit on koulutettu ja asennettu oikein. Ne voivat suorittaa toistuvia tehtäviä nopeasti ja virheettömästi. API-integraatiot ovat yleensä kaikista kustannustehokkain ratkaisu eri integraatioiden joukossa, koska ne mahdollistavat eri järjestelmien yhteistyön ilman suuria investointeja tai monimutkaista ylläpitoa.

Tietoturvallisuus

Robottivusteinen integraatio, datan jakeluun perustuva integraatio, ESB-integraatio sekä API-integraatiot voivat olla vahvoja tietoturvallisuuden ja tietosuojan kannalta. ETL-integraatio voi olla suhteellisen tietoturvallinen ja tietosuojaltaan kohtuullinen vaihtoehto, mutta datan siirto prosessin aikana voi altistaa riskille.

Manuaalisessa integraatiossa inhimilliset virheet voivat aiheuttaa tietoturvahaasteita, vaikka toisaalta mahdolliset riskit voivat olla helpommin havaittavissa, koska manuaalinen integraatio tapahtuu käsin. Pistemäinen integraatio voi olla haastava tietoturvallisuuden ja tietosuojan kannalta, sillä järjestelmien välisten yhteyksien hallinta voi olla hankalaa ja altistaa mahdollisille riskeille.

Monimutkaisuuden hallittavuus

Monimutkaisuuden hallittavuus kuvaa kuinka hyvin integraatio pystyy käsittelemään monimutkaisia integraatiotarpeita. API-integraatiot, datan jakeluun perustuva integraatio ja ESB-integraatio ovat hyviä ratkaisuja monimutkaisuuden hallintaan, kun taas manuaalinen integraatio ja pistemäiset integraatiot saattavat olla haastavampia etenkin monimutkaisten kokonaisuuksien integroimiseen. Robottivusteinen integraatio voi auttaa automatisoimaan prosesseja ja vähentämään inhimillisten virheiden riskiä. Kuitenkin monimutkaiset integraatiot voivat robottivusteisessa integraatiossakin edelleen vaatia manuaalista ohjausta ja ylläpitoa. ETL-ratkaisut voivat olla tehokkaita tietojen siirtämisessä ja muokkaamisessa eri järjestelmien välillä. Kuitenkin monimutkaiset transformaatiot voivat aiheuttaa haasteita, ja ETL-prosessien ylläpito voi olla aikaa vievää.

Nopean tiedonsiirron ja reaaliaikaisuuden mahdollistavat parhaiten API-integraatiot, pistemäinen integraatio sekä ESB-integraatiot. Robottivusteisessa ohjelmistoautomaatiossa käytetään ohjelmistorobottia automatisoimaan tietojen siirto eri järjestelmien välillä, ja reaaliaikaisuus voi olla keskitasoa, riippuen ohjelmistorobottin tehokkuudesta ja nopeudesta. ETL-integraatiossa reaaliaikaisuus voi olla haasteellista, sillä prosessi voi viedä aikaa. Datan jakeluun perustuvassa integraatiossa keskitetty toimija voi hidastaa tiedonsiirtoa. Manuaalisessa integraatiossa reaaliaikaisuus voi olla heikko, sillä tietojen siirto vaatii käyttäjän toimenpiteitä ja aikaa.

Vikasietoisuus

Vikasietoisuus (engl. fault tolerance) kertoo, kuinka hyvin integraatio pystyy jatkamaan toimintaa virhetilanteissa tai katkoksissa. API-integraatiot, ESB-integraatio, robottivusteinen ohjelmistoautomaatio ja datan jakeluun perustuva integraatio ovat

yleensä kaikkein vikasietoisimpia integraatoratkaisuja, kun taas manuaalinen integraatio ja pistemäinen integraatio voivat olla haavoittuvaisempia virheille.

Manuaalinen integraatio on altis inhimillisille virheille ja vaatii jatkuvaa valvontaa, joten se on integraatiomalleista kaikista haavoittuvaisin vikojen suhteen. Pistemäinen integraatio voi olla hyvä vikasietoisuuden suhteen, kunhan jokainen integraatiopiste toimii itsenäisesti ja on suojattu vikoja vastaan. Kuitenkin koska pistemäinen integraatio perustuu yhden pisteen integrointiin toisen kanssa, jos yksi osapuoli muuttuu tai vikaantuu, se voi vaikuttaa muihin pisteisiin ja integraatio voi katketa. Robottivusteinen ohjelmistoautomaatio käyttää automaatioita ja ohjelmistoja, jotka kykenevät nopeasti havaitsemaan ja korjaamaan virheitä, kunhan ohjelmistot ovat ajan tasalla ja oikein määritettyjä. Kuitenkin monimutkaisemmat virheet robottivusteisessa ohjelmistoautomaatiossa saattavat vaatia ihmisen väliintuloa.

Datan jakeluun perustuvassa integraatiossa keskitetty toimija voi havaita ja korjata virheitä nopeasti ja reitittää tietoa uudelleen tarpeen mukaan, mutta se voi olla haavoittuvainen, jos toimija kaatuu. Osapuolet (engl. spoke) voivat kuitenkin toimia itsenäisesti ja estää virheiden leviämisen. ETL-prosessit voivat olla vikasietoisia, kunhan tietojen keräys, muokkaus ja siirto on suunniteltu huolellisesti. Kuitenkin virheet voivat vaikuttaa koko prosessin toimintaan. ESB voi havaita ja korjata virheitä nopeasti ja sen keskitetty hallinta voi auttaa vikojen leviämisen estämisessä. Kuitenkin ESB:n kaatuminen voi vaikuttaa koko integraatioprosessiin. API-integraatiot voivat olla erittäin vikasietoisia, kunhan niitä suunnitellaan ja toteutetaan huolellisesti.

Ylläpidettävyys

Ylläpidettävyys kertoo, kuinka helppoa tai vaativaa integraation ylläpito on, mukaan lukien päivitykset, muutokset ja virheidenkorjaukset. API-integraatiot ovat yleensä kaikkein ylläpidettävimpiä, kun taas manuaalinen integraatio ja pistemäinen integraatio voivat olla haastavampia ylläpitää. Manuaalinen integraatio vaatii paljon manuaalista työtä ja päivityksiä jokaisen muutoksen yhteydessä. Tämä tekee ylläpidosta hankalaa ja altista inhimillisille virheille. Pistemäinen integraatio voi olla hankalaa ylläpitää, etenkin jos eri pisteitä tarvitsee päivittää eri aikataululla. Päivitykset voivat aiheuttaa katkoksia ja vaatia paljon manuaalista työtä.

Datan jakeluun perustuva integraatio voi olla melko hyvin ylläpidettävä, kunhan toimija (engl. hub) ja osapuolet (engl. spoke) on suunniteltu ja päivitetty huolellisesti. Päivitykset keskitetään keskuspiisteeseen, mikä helpottaa ylläpitoa. Robottivusteinen integraatio voi olla hyvin ylläpidettävää, kunhan robotit on ohjelmoitu oikein ja päivitykset tehdään huolellisesti. Automaatio helpottaa ylläpitoa ja vähentää virheiden riskiä. ETL- sekä ESB-

integraatoratkaisut voivat olla hyvin ylläpidettäviä. ETL-automaatio auttaa ylläpidossa ESB tarjoaa keskitetyn tavan hallita integraatioita.

Soveltuvuus useille yrityksille

Manuaalinen integraatio ja pistemäinen integraatio soveltuvat ainoastaan pienimuotoisille integraatioille. Robottivasteinen ohjelmistoautomaatio ja datan jakeluun perustuva integraatio soveltuvat useiden järjestelmien välisiin integraatioihin. Yrityksen laajuiset integraatiot ja API-integraatiot taas soveltuvat monimutkaisiin ja mittaviin integraatioihin, jotka vaativat laajaa datan liikuttamista ja monipuolista kommunikointia eri järjestelmien välillä.

8.2 API-tyyppien ominaisuuksien vertailu

Taulukko 2 vertailee yleisimpien eri API-tyyppien ominaisuuksia.

API-tyyppi / Ominaisuus	REST	SOAP	GraphQL	gRPC
Protokolla	HTTP/HTTPS	HTTP/HTTPS, SMTP, TCP	HTTP/HTTPS	HTTP/2
Tiedonmuoto	JSON, XML	XML	JSON	Protocol Buffers
Suorituskyky	Hyvä	Keskimääräinen	Erittäin hyvä	Erinomainen
Tietoturva	Riippuu toteutuksesta	Korkea	Riippuu toteutuksesta	Korkea
Joustavuus	Korkea	Matala	Erittäin korkea	Korkea
Verkkoyhteys	Tilaton	Tilallinen	Tilaton	Tilaton/Tilallinen
Monimutkaisuus	Matala	Korkea	Keskimääräinen	Keskimääräinen
Tietojen suodatus	Rajoitettu	Rajoitettu	Erinomainen	Hyvä
Versiointi	Manuaalinen	Manuaalinen	Manuaalinen	Protobuf-skeemoissa
Selaintuki	Erinomainen	Hyvä	Hyvä	Rajoitettu
Kaksisuuntainen viestintä	Rajoitetusti (WebSockets)	Rajoitetusti (WebSockets)	Rajoitetusti (Subscription)	Kyllä
Virheen käsittely	HTTP-statuskoodit	<Fault>-elementti	Spesifiset virheet	gRPC:n statuskoodit
Alusta-riippumattomuus	Kyllä	Kyllä	Kyllä	Kyllä
Soveltuvuus	Yleiskäyttöinen: verkkopalvelut, mobiilisovellukset	Maksuliikenne, identiteetin hallinta, CRM, finanssipalvelut, telepalvelut, julkishallinto, terveydenhuolto, legacy-järjestelmät ym.	Räätälöity ja tehokas datan haku	Mikropalvelut, reaaliaikaiset sovellukset, tehokas kommunikaatio

Taulukko 2. API-tyyppien vertailutaulukko

Protokolla

REST ja GraphQL käyttävät yleisesti HTTP/HTTPS-protokollia, mikä tekee niistä helposti yhteensopivia useimpien nykyisten verkkoteknologioiden kanssa. SOAP voi käyttää erilaisia protokollia, kuten HTTP, HTTPS, SMTP ja TCP, mikä tarjoaa lisää joustavuutta mutta myös lisää kompleksisuutta. gRPC käyttää HTTP/2:ta, joka on suunniteltu korkeaa suorituskykyä varten ja mahdollistaa kaksisuuntaisen viestinnän.

Tiedonmuoto

REST ja GraphQL tukevat JSON-formaattia, joka on yksinkertainen ja helppo lukea. REST tukee myös XML:ää. SOAP käyttää ainoastaan XML:ää, mikä tekee siitä raskaamman verrattuna JSON:iin. gRPC käyttää Protocol Buffers -binäärimuotoa, joka on tehokas mutta vaatii erityistä käsittelyä.

Suorituskyky

REST on yleensä nopea, mutta ei yhtä tehokas kuin gRPC. SOAP voi olla hidas suurten XML-viestien vuoksi. GraphQL on erittäin joustava ja voi olla hyvin suorituskykyinen riippuen kyselyistä. gRPC on suunniteltu korkeaa suorituskykyä varten, erityisesti palvelinten välisessä viestinnässä.

Tietoturva

Kaikki neljä API-tyyppiä tarjoavat vahvoja suojausmekanismeja, mutta toteutus ja yksityiskohdat vaihtelevat. SOAP API:n tietoturvasäilytys on korkea WS-Security-standardin ansiosta. Muissa API-tyypeissä tietoturva riippuu paljon toteutuksesta ja käytetyistä mekanismeista, kuten OAuth ja JWT.

Joustavuus

REST ja GraphQL ovat erittäin joustavia, helposti laajennettavissa ja tilattomia eli jokainen pyyntö on itsenäinen, mikä tekee sovelluksesta helpommin skaalautuvan ja ylläpidettävän. SOAP on vähemmän joustava, koska sen viestien sisältö on XML-pohjaista ja viestien rakenne on määritelty tarkasti, mikä voi rajoittaa sen soveltuvuutta erilaisiin sovelluksiin. gRPC on joustava, mutta vaatii Protocol Buffers -skeeman.

Verkkoyhteys

REST ja GraphQL ovat tilattomia, mikä helpottaa skaalautuvuutta. SOAP on tilallinen, mikä voi olla hankalaa suurissa järjestelmissä. gRPC voidaan toteuttaa sekä tilallisena että tilattomana.

Monimutkaisuus

REST on yksinkertainen, kun taas SOAP on monimutkaisempi. GraphQL ja gRPC sijoittuvat näiden kahden väliin.

Tietojen suodatus

GraphQL tarjoaa parhaan tietojen suodatuskyvyn, koska siinä voidaan määritellä tarkasti, mitä dataa halutaan. REST ja SOAP ovat rajoitetumpia tässä suhteessa. gRPC tarjoaa hyvän, mutta ei yhtä joustavan suodatuskyvyn kuin GraphQL.

Versiointi

Kaikissa API-tyypeissä versiointi on yleensä manuaalista, mutta gRPC tukee versiointia Protocol Buffers -skeemoissa.

Selaintuki

REST on laajimmin tuettu selaimissa. SOAP ja GraphQL seuraavat perässä. gRPC:n selaintuki on rajoitettua.

Kaksisuuntainen viestintä

Kaksisuuntainen viestintä tarkoittaa sitä, että kommunikointi tapahtuu molempiin suuntiin eli sekä asiakkaalta palvelimelle että palvelimelta asiakkaalle. REST, SOAP ja GraphQL ovat alun perin suunniteltu asiakkaan aloittamaan yksisuuntaiseen viestintään. Reaaliaikainen kommunikointi voidaan kuitenkin mahdollistaa REST- ja SOAP-protokollilla käyttämällä esimerkiksi WebSockets-tekniikkaa, joka mahdollistaa reaaliaikaisen kommunikoinnin REST- tai SOAP-rajapinnan päällä. GraphQL-kyselykielellä kaksisuuntainen viestintä voidaan toteuttaa "subscriptions"-nimisellä toiminnallisuudella. gRPC on sen sijaan suunniteltu alusta alkaen tukemaan kaksisuuntaista kommunikaatiota.

Virheenkäsitely

REST käyttää virheenkäsitelyssä HTTP-statuskoodeja, SOAP käyttää "<Fault>"-elementtiä, GraphQL-kyselykielessä virhetiedot sisällytetään vastauksen body-osan "errors"-kentässä, ja gRPC käyttää omia statuskoodejaan. Lisäksi gRPC mahdollistaa "trailers"-nimellä kulkevien metatietojen lähettämisen, jotka voivat sisältää lisätietoa virheestä.

Alustariippumattomuus

Kaikki API-tyypit ovat periaatteessa riippumattomia alustasta, mutta niiden asentaminen ja käyttö voi vaihdella riippuen kehitysympäristöstä ja ohjelmointikielestä. REST käyttää

standardia HTTP/HTTPS-protokollaa, mikä tekee siitä hyvin alustariippumattoman. REST:ä voidaan käyttää lähes missä tahansa ohjelmointikielessä, joka tukee HTTP-pyyntöjä, ja se tukee monia eri datamuotoja kuten JSON,XML ja HTML. SOAP-pyyntöjä voidaan tehdä mistä tahansa alustasta, joka pystyy käsittelemään XML:ää. GraphQL on myös suhteellisen alustariippumaton ja se tukee monia ohjelmointikieliä, jotka tukevat HTTP-pyyntöjä ja ymmärtävät JSON-muotoista dataa. Kuitenkin sen käyttöönotto ja ylläpito voivat vaatia erityistä osaamista ja kirjastoja.

Vaikka gRPC on suunniteltu olemaan alustariippumaton, se perustuu Protocol Buffers -muotoon ja HTTP/2-protokollaan. Tämä tekee siitä vähemmän sopivan vanhemmille alustoille, jotka eivät tue HTTP/2:tta, ja se saattaa vaatia erityisiä gRPC-kirjastoja. Google tarjoaa virallisia kirjastoja useille ohjelmointikielille, mikä tekee siitä erittäin alustariippumattoman ratkaisun mikropalveluarkkitehtuureille ja järjestelmäintegraatioille.

Soveltuvuus

REST API on yleiskäyttöinen ja sopii hyvin yksinkertaiseen datanvaihtoon. Se soveltuu erityisen hyvin verkkopalveluiden ja mobiilisovellusten kehittämiseen. SOAP API soveltuu erityisesti liiketoiminnallisesti kriittisiin ja turvallisuuden kannalta herkkiin integraatioihin, kuten pankkien, terveydenhuollon ja muiden kriittisten järjestelmien integraatioihin, joissa tarvitaan standardoitua, turvallista datanvaihtoa ja transaktioita. GraphQL API soveltuu hyvin yrityksille, jotka tarvitsevat kohdennettua datan hakua. gRPC API on ihanteellinen suurille yrityksille ja mikropalveluarkkitehtuureja käyttäville yrityksille, reaaliaikaisille järjestelmille sekä korkean suorituskyvyn vaativille sovelluksille.

9 Ohjeita optimaalisen dataintegraatiototeutuksen valintaan

9.1 Yrityksen tarpeiden ja resurssien kartoittaminen

Ennen investointeja dataintegraatoratkaisuihin yrityksen tulee selvittää tarkkaan, kuka datasta hyötyy ja miten. Yrityksen on ymmärrettävä tarkasti, mihin ja mitä varten se dataa tuottaa, miten data voi tuottaa arvoa ja kenelle. Yrityksen tulee myös selvittää tietojärjestelmien ja datan yhteensopivuus, jotta tietoja voidaan vaihtaa ja hyödyntää sujuvasti eri järjestelmien ja sovellusten välillä. Yhteensopivuus varmistaa, että tiedot siirtyvät oikeassa muodossa ja oikeaan aikaan.

Yrityksen on olennaista arvioida omat tarpeensa ja resurssinsa ennen integraation toteutusta sekä valita niiden perusteella sopivin integraatoratkaisu. Erityisesti, kun kyseessä ovat yritysten väliset dataintegraatiot, on tärkeää tunnistaa myös kumppaniyrityksen tarpeet ja odotukset. Tämä vaatii tiivistä yhteistyötä ja kommunikaatiota kumppanin kanssa, jotta integraatio suunnitellaan ja toteutetaan molempien osapuolen tarpeita vastaavaksi. Yritysten välisissä integraatioissa yhteistyö myös kumppaniyrityksen IT-tiimien kanssa on usein välttämätöntä, jotta voidaan varmistaa integraation onnistuminen ja yhteensopivuus.

Dataintegraatiomallin valinnassa on syytä olla huolellinen, koska se määrittelee pitkälti integraation onnistumisen ja sen, kuinka hyvin tietojen yhdistäminen tukee liiketoiminnan tavoitteita. On myös tärkeää varmistaa ratkaisun joustavuus tulevaisuuden muutoksissa sekä sen skaalautuvuus yrityksen kasvun ja muutosten mukana. (Bhadani 2021.)

9.2 Dataintegraatiototeutuksen valinta

Jokaisella integraatiomallilla on omat vahvuutensa ja heikkoutensa, joten kullekin yritykselle sopivan dataintegraatiomallin valinnassa on tärkeää punnita yrityksen tarpeet ja resurssit. Manuaalinen integraatio voi olla helppo ja nopea ratkaisu pienimuotoisiin tarpeisiin, mutta se vaatii työvoimaa, ei yleensä skaalaudu hyvin ja voi johtaa virheisiin ja tehottomuuteen. Pistemäinen dataintegraatio on usein nopea ja helppo toteuttaa, mutta saattaa johtaa monimutkaiseen ja hallitsemattomaan integraatioympäristöön. Robottivasteisen ohjelmistoautomaation etuina ovat manuaalisen työn tarpeen ja inhimillisten virheiden väheneminen, kustannussäästöt sekä nopeus, tehokkuus ja reaaliaikaisuus, mutta se vaatii jatkuvaa ylläpitoa ja valvontaa eikä se sovellu monimutkaisiin, laajoihin integraatioihin.

Datan jakeluun perustuva integraatio tarjoaa keskitetyn ratkaisun, joka voi yksinkertaistaa hallintaa, mutta se saattaa muodostaa pullonkaulan ja olla hankala skaalata. ETL on tehokas suuren tietomäärän siirtämiseen ja muokkaamiseen, mutta sen toteuttaminen voi

9.3 Yritysten kokemukset ja suositukset dataintegraatioon

Haastatellut yritykset suosittelivat, että ennen yrityksen ulkoisiin dataintegraatioihin laajentamista, on ensin tärkeää saada yrityksen sisäiset dataintegraatioprosessit kuntoon ja arvioida huolellisesti yrityksen resurssit ja tarpeet. Yritykset suosittelivat kiinnittämään erityistä huomiota datan laatuun ja sen keräys- ja käsittelytapoihin sekä huolelliseen seurantaan, koska nämä tekijät ovat keskeisiä dataintegraation onnistumisen kannalta. Toisaalta turhan datan keräämisestä voi syntyä ylimääräisiä kustannuksia ja mahdollisia GDPR-tietosuoja-asetukseen liittyviä haasteita. Yritykset suosittelivat suodattamaan tarpeettoman tiedon, koska kaikki tiedot eivät ole kaikille tiimeille tarpeellisia. Yritykset korostivat datan suodattamisessa ja käsittelyssä myös tietoturvan ja tietosuojan huomioimisen tärkeyttä.

Kun dataintegraatiomallia valitaan, on tärkeää määritellä tarkasti integroitavaksi haluttu data ja mistä lähteistä se saadaan. Tämä auttaa hahmottamaan, mitä tietoa yritys tarvitsee ja miten se voidaan parhaiten hyödyntää. Haastatellut yritykset suosittelivat myös varmistamaan, että dataintegraatiota suunnittelevalla yrityksellä on tehokas tapa hallita ja suojata omaa dataansa. Lisäksi yritykset suosittelivat tutkimaan huolellisesti yrityksen olemassa olevia tietojärjestelmiä ja varmistamaan, että ne ovat yhteensopivia integraation kanssa.

Ulkopuolisiin dataintegraatioihin siirryttäessä haastatellut yritykset suosittelivat etsimään kumppaneita, jotka voivat tuoda lisäarvoa yrityksen liiketoiminnalle. Alkuun pitää pohtia, mitä dataa jaetaan, kenelle, ja mitä molemmat osapuolet voisivat siitä hyötyä dataintegraatiosta. Sopimusten vaikutukset datan käyttöön ja jakamiseen tulisi arvioida tarkasti ja neuvotella suotuisat ehdot molemmille osapuolille.

Dataintegraatiomallin valinnassa on arvioitava monia tärkeitä tekijöitä, kuten yrityksen budjetti, teknillinen osaaminen, datan tallennuskapasiteetti ja integroitavien järjestelmien lukumäärä. Yritykset suosittelivat panostamaan henkilöstön koulutukseen ja osaamisen kehittämiseen, erityisesti integraatioihin liittyvissä teknologioissa.

Manuaalisesta integraatiosta yritykset suosittelivat pyrkimään pois kohti automatisoituja integraatoratkaisuja, kuten robottivälineistä ohjelmistoautomaatiota. Haastatellut yritykset olivat yhtä mieltä siitä, että pistemäistä integraatiota voi kokeilla pienimuotoisesti datan arvon toteamiseksi, mutta pitkällä aikavälillä on syytä harkita kattavampaa integraatoratkaisua. Toisaalta yritykset suosittelivat aloittamaan integraatiot ensin pienimuotoisina kokeiluina ja laajentamaan niitä vasta, kun dataintegraation

liiketoimintahyödyt ovat selkeät. Tavoitteena tulisi olla yksinkertaiset ja kustannustehokkaat ratkaisut, joissa ei kuitenkaan tingitä laadusta ja toimivuudesta.

Yritykset kannustivat käyttämään erilaisia integraatiomalleja ja -ratkaisuja tarpeen mukaan eivätkä rajoittumaan vain yhteen vaihtoehtoon. Yritykset suosittelivat harkitsemaan integraatioalustojen käyttöä monimutkaisten integraatioiden hallinnassa, sillä ne mahdollistavat keskitetyn valvonnan ja hallinnan, tarjoavat monipuolisia työkaluja sekä tukea erilaisille integraatiotyypeille. Integraatioalustat voivat säästää aikaa ja resursseja sekä parantaa datan laadunhallintaa. Integraatioalustojen käyttö mahdollistaa yrityksille paremman integraatioiden seurannan, vianetsinnän, päivitykset ja skaalautuvuuden yhden yhtenäisen alustan kautta. Näin ollen integraatioalustat voivat olla tehokas työkalu monimutkaisten integraatioiden toteuttamiseen, mikä edistää liiketoiminnan joustavuutta ja kilpailukykyä nykyaikaisessa liiketoimintaympäristössä.

Yritykset suosittelivat data mesh -ajattelutavan hyödyntämistä, sillä se voi auttaa organisaatioita integroimaan datan tehokkaasti osaksi liiketoimintaprosessejaan. Tämä lähestymistapa edistää datan hyödyntämistä ja päätöksenteon parantamista tarjoamalla käytännöllisen tavan hallita ja jakaa dataa organisaation sisällä.

Dataintegraatioiden käyttöönoton yhteydessä haastatellut yritykset suosittelivat panostamaan riittävään koulutukseen ja tukeen. Yritykset korostivat, että on tärkeää seurata dataintegraatioiden suorituskykyä ja päivittää niitä tarvittaessa, jotta ne vastaavat jatkuvasti yrityksen tarpeita. Samalla yritykset korostivat *"if it ain't broke, don't fix it"* -periaatetta, koska muutokset dataintegraatioissa voivat tuoda mukanaan suuria kustannuksia ja myös riskejä.

Haastatelluista yrityksistä suurimmalla osalla oli käytössään erilaisia API-integraatioita. Myös robottivasteista ohjelmistoautomaatiota käytettiin automatisoimaan prosesseja. Jonkin verran käytettiin vielä myös manuaalisia integraatioita ja pistemäisiä integraatioita. Monet haastatellut yritykset hyödynsivät useita erilaisia integraatiomalleja ja -ratkaisuja sen sijaan, että käyttäisivät vain yhtä ratkaisua. Osalla haastatelluista yrityksistä oli käytössään integraatioalustoja vastauksena monimutkaisiin integraatiotarpeisiinsa.

10 Yhteenveto ja pohdinta

Opinnäytetyössä tarkasteltiin erilaisia integraatiomalleja, analysoiden niiden ominaisuuksia, hyötyjä ja haittoja sekä soveltuvuutta erikokoisille ja -tyyppisille yrityksille. Tutkimuksessa osoitettiin, että dataintegraatiomallit voivat merkittävästi parantaa yritysten operatiivista tehokkuutta, kustannusrakennetta ja kilpailukykyä.

Yrityshaastatteluissa korostettiin datan arvon tunnistamisen merkitystä, ja että dataintegraatioiden onnistuminen ja niiden täyden potentiaalin realisoituminen vaatii koko organisaation asennemuutosta yrityksen johtoa myöten. Haastateltujen yritysten menestystarinat todistivat, että tietoisella lähestymistavalla ja oikeilla ratkaisuilla datan jakaminen ja dataintegraatiot voivat olla merkittävä voimavara liiketoiminnan kasvuun ja kehittämiselle. Ne avaavat myös uusia ovia liiketoiminnalle, missä tiedon hyödyntäminen on yhä keskeisemmässä roolissa. Tutkimustulokset osoittavat, että dataintegraatiomallit voivat parhaimmillaan muuttaa yritysten toimintaa merkittävästi ja niiden hyödyt ulottuvat suorista vaikutuksista myös epäsuoriin tekijöihin, kuten parantuneeseen asiakastyytyväisyyteen, virheiden vähenemiseen ja lisääntyneeseen innovaatiokyvykkyyteen.

Työssä ilmeni, että jokaisella dataintegraatiomallilla on omat vahvuutensa ja heikkoutensa, eikä ole olemassa yhtä integraatiomallia, joka sopisi kaikille yrityksille. Integraatiomallin valinta riippuu monista tekijöistä, kuten yrityksen koosta, toimialasta, erityistarpeista, teknologisesta valmiudesta ja liiketoimintatavoitteista. Integraatiomallin valinta edellyttää huolellista harkintaa ja strategista näkemystä, ottaen huomioon kustannukset, monimutkaisuuden ja toiminnallisuuden. Koska integraatiot ovat taloudellinen investointi, on olennaista huomioida myös yrityksen tulevaisuuden tarpeet. Erityisesti joustavuus ja skaalautuvuus ovat avainasemassa, koska liiketoiminnan kehittyessä integraatiotarpeet saattavat muuttua. Sopivan integraatiomallin valinta on tärkeää, sillä se vaikuttaa pitkällä aikavälillä ja suoraan yrityksen kilpailukykyyn.

Jatkotutkimuksissa voitaisiin laajentaa näkökulmaa useampiin yrityksiin ja toimialoihin, kehittämällä yhä tarkempia suosituksia ja työkaluja dataintegraatoratkaisujen valintaan. Eri toimialojen erityistarpeiden ja niiden vaikutuksen integraatiomallien valintaan ymmärtäminen tarjoaisi uutta tietoa dataintegraatioiden hyödyntämisestä. Tulevaisuuden tutkimuksissa voitaisiin mahdollisesti tutkia myös tekoälyn ja koneoppimisen vaikutusta dataintegraatioprosesseihin. Yrityshaastatteluissa ilmeni, että API-integraatioihin toivottiin parempia ohjeita, joten jatkokehityksenä voitaisiin luoda käytännönläheisiä ja helposti ymmärrettäviä oppaita täydentämään virallista dokumentaatiota.

Jatkotutkimuksissa voitaisiin myös syventää ymmärrystä dataintegraatioiden taloudellisista vaikutuksista. Integraatiomallien taloudellisten vaikutusten, kuten sijoitetun pääoman tuoton ja kokonaiskustannusten, arvioiminen voisi auttaa yrityksiä tekemään tietoisempia valintoja. Tämä voisi edistää yritysten kykyä hyödyntää dataa entistä tehokkaammin - ei vain omaksi edukseen, vaan myös laajemman yhteiskunnallisen hyvän luomiseksi.

Lähteet

AltexSoft. 2019. What is SOAP: Formats, Protocols, Message Structure, and How SOAP is Different from REST. Viitattu 2.9.2023. Saatavissa

<https://www.altexsoft.com/blog/engineering/what-is-soap-formats-protocols-message-structure-and-how-soap-is-different-from-rest/>

AltexSoft. 2020. Comparing API Architectural Styles: SOAP vs REST vs GraphQL vs RPC. Viitattu 7.9.2023. Saatavissa <https://www.altexsoft.com/blog/soap-vs-rest-vs-graphql-vs-rpc/>

Ariyaratne, J. 2019. What do you really need within your solution: ESB or Microservices? Medium. Viitattu 29.7.2023. Saatavissa <https://medium.com/@jagathsisira/what-do-you-really-need-within-your-integration-solution-esb-or-microservices-1552d9803cb4>

Art, A. 2020. What GDPR Means For API Developers. Nordic APIs. Viitattu 4.9.2023. Saatavissa <https://nordicapis.com/what-gdpr-means-for-api-developers/>

Bhadani, P. 2021. Complete Guide to Data Integration - What It Is & Why It Matters. AnyConnector. Viitattu 22.7.2023. Saatavissa <https://anyconnector.com/data-integration/data-driven-strategy.html>

Blueprint. 2020. How to Reduce the Rising Costs of RPA Maintenance and Support. Viitattu 7.7.2023. Saatavissa <https://www.blueprintsys.com/blog/rpa/reduce-rising-costs-rpa-maintenance-and-support>

Blueprint. 2021. What is Intelligent Process Automation (IPA)? All You Need to Know. Viitattu 8.7.2023. Saatavissa <https://www.blueprintsys.com/blog/rpa/what-is-intelligent-process-automation-ipa>

Burk, N. 2017. GraphQL Server Basics: The Network Layer. Prisma Data. Viitattu 2.9.2023. Saatavissa <https://www.prisma.io/blog/graphql-server-basics-the-network-layer-51d97d21861>

Business Technology Standard. 2023. Data, analytiikka ja integraatiot.

Bisnesteknologiamalli, versio 4.5.2. Viitattu 20.6.2023. Saatavissa

<https://btmalli.fi/book/demand/data-analytics-and-integration/> ja

<https://www.managebt.org/content/uploads/Bisnesteknologiamalli-20211102.pdf>

Choi, D., R'bigui, H., Cho, C. 2021. Candidate Digital Tasks Selection Methodology for Automation with Robotic Process Automation. Sustainability 13(16):8980. Viitattu

15.7.2023. Saatavissa <https://doi.org/10.3390/su13168980> (<https://www.mdpi.com/2071-1050/13/16/8980>)

Dearmer, A. 2021. The Five Types of Data Integration. Viitattu 29.6.2023. Saatavissa <https://www.integrate.io/blog/the-five-types-of-data-integration/>

Dilmegani, C. 2023. What is RPA? Robotic Process Automation Tech Guide in '23. Viitattu 6.8.2023. Saatavissa <https://research.aimultiple.com/rpa/>

Gail, M. 2021. The Benefits of using a Third-Party Chat API. QuickBlox. Viitattu 2.9.2023. Saatavissa <https://quickblox.com/blog/5-benefits-of-using-third-party-apis-in-your-online-business/>

Giuliani, S. 2020. Hub-and-spoke vs. point-to-point data synchronization: There's one clear winner. The Official Census Blog. Sutro Labs Inc. Viitattu 29.7.2023. Saatavissa <https://www.getcensus.com/blog/hub-and-spoke-vs-point-to-point-data-synchronization-theres-one-clear-winner>

GreenVulcano Technologies. 2019. Integration evolution: From point-to-point integration to ESB. GreenVulcano ESB. Viitattu 20.7.2023. Saatavissa <https://www.greenvulcano.com/en/integration-evolution-from-point-to-point-integration-to-esb/>

gRPC Authors. 2023a. Supported languages. Viitattu 4.9.2023. Saatavissa <https://grpc.io/docs/languages/>

gRPC Authors. 2023b. Authentication. Viitattu 6.9.2023. Saatavissa <https://grpc.io/docs/guides/auth/>

Gupta, L. 2022. Statelessness in REST APIs. REST API Tutorial. Viitattu 1.9.2023. Saatavissa <https://restfulapi.net/statelessness/>

Guthrie, G. 2023. How microservices and APIs drive scalability and efficiency. Nulab. Viitattu 2.9.2023. Saatavissa <https://nulab.com/learn/software-development/microservices-and-apis/>

Habas, C. 2019. Hub & Spoke Model: Definition, Benefits & Examples. Bizluent. Viitattu 27.7.2023. Saatavissa <https://bizfluent.com/13721308/hub-spoke-model-definition-benefits-examples>

Healthie. 2022. The Advantages & Disadvantages of API Solutions. Viitattu 4.9.2023. Saatavissa <https://www.getthealthie.com/blog/api-advantages-disadvantages>

Heavy AI. 2023. Data Integration. Viitattu 20.6.2023. Saatavissa <https://www.heavy.ai/technical-glossary/data-integration>

HiQ Finland Oy. 2022. Kotimainen Friends integraatioalusta. Viitattu 26.8.2023. Saatavissa <https://hiq.fi/palvelut/integraatio-ja-automaatio/friends/>

Hurja Solutions. 2022. Ohjelmiston integraatio – miksi integroida ja kuinka onnistua? Blogiartikkeli. Julkaistu 12.10.2022. Päivitetty 30.5.2023. Viitattu 22.6.2023. Saatavissa <https://www.hurja.fi/blogi/ohjelmiston-integraatio-miksi-integroida-ja-kuinka-onnistua/>

Inscribe. 2023 Robotic Process Automation (RPA): A Complete Overview. Viitattu 4.7.2023. Saatavissa <https://www.inscribe.ai/robotic-process-automation>

Janani. 2022. SOAP vs. REST: What's the Difference? NamLabs Technologies Pvt Ltd. Viitattu 2.9.2023. Saatavissa <https://www.atatus.com/blog/soap-vs-rest-whats-the-difference/>

Javapoint. 2023. RPA Life Cycle. RPA Tutorial. Viitattu 10.7.2023. Saatavissa <https://www.javapoint.com/rpa-life-cycle>

Keebola. 2022. Complete ETL Process Overview (design, challenges and automation). Viitattu 28.7.2023. Saatavissa <https://www.keebola.com/blog/etl-process-overview>

Khando, K., Shang, G., Sirajul, M. I., Ali, S. 2021. Enhancing employees information security awareness in private and public organisations: A systematic literature review, Computers & Security 106. Viitattu 20.6.2023. Saatavissa <https://doi.org/10.1016/j.cose.2021.102267>

Kirkwood, G. 2019. The Impact of RPA on Employee Engagement, a Forrester Consulting Thought Leadership Paper. UiPath. Viitattu 5.7.2023. Saatavissa <https://www.uipath.com/blog/rpa/impact-of-rpa-on-employee-engagement-forrester>

Kirkwood, G. 2020. UiPath Recognized Highest for Ability to Execute in 2020 Gartner Magic Quadrant for Robotic Process Automation. UiPath. Viitattu 8.7.2023. Saatavissa <https://www.uipath.com/blog/rpa/gartner-magic-quadrant-rpa-report-2020-archives>

Kost, E. A Data Leak Detection Guide for the Tech Industry in 2023. UpGuard. Viitattu 25.6.2023. Saatavissa <https://www.upguard.com/blog/how-tech-companies-can-detect-data-leaks>

Kouzmanov, S. 2015. Analysis and Classification of Business Software Systems Integration Approaches. Viitattu 18.7.2023. Saatavissa https://www.unwe.bg/uploads/Alternatives/06_Kouzmanov.pdf

- Laserfiche. 2023. Guide to Robotic Process Automation. Viitattu 3.7.2023. Saatavissa <https://info.laserfiche.com/resource/guide-to-robotic-process-automation>
- Lawton, G. 2021. Robotic Process Automation (RPA). TechTarget. Viitattu 2.7.2023. Saatavissa <https://www.techtarget.com/searchcio/definition/RPA>
- Malkamäki, S. 2022. Reilut terveystietojärjestelmät Suomen tulevaisuuden valttina. Sitra. Viitattu 21.6.2023. Saatavissa <https://www.sitra.fi/artikkelit/reilut-terveysalan-dataekosysteemit-suomen-tulevaisuuden-valttina/>
- Midpointed. 2021. Ohjelmistorobotiikan tila ja hyödyntäminen suomalaisissa yrityksissä. Viitattu 3.7.2023. Saatavissa <https://www.midpointed.fi/palvelut/ohjelmistorobotiikka/rpa-tutkimus/>
- National Center for Education Statistics. 2023. Safeguarding Your Technology. Chapter 6 Protecting Your System: Information Security. Institute of Education Sciences. Viitattu 22.6.2023. Saatavissa <https://nces.ed.gov/pubs98/safetech/chapter6.asp>
- Neray, G., Scott, S. 2021. Best practices for REST API security: Authentication and authorization. Viitattu 5.9.2023. Saatavissa <https://stackoverflow.blog/2021/10/06/best-practices-for-authentication-and-authorization-for-rest-apis/>
- Nilsen, P. 2018. Pros and Cons of using an Enterprise Service Bus in Digital Transformation. Cetdigit Cloud Services. Viitattu 31.7.2023. Saatavissa <https://www.cetdigit.com/blog/enterprise-integration-pros-and-cons-of-enterprise-service-bus>
- Otieno, B. 2023. From Mess to Meaningful: A Guide to Ensuring Data Quality and Accuracy while Synthesizing Multiple Sources. Technology & Data. Viitattu 14.7.2023. Saatavissa <https://www.linkedin.com/pulse/from-mess-meaningful-guide-ensuring-data-quality-accuracy-otieno/>
- Panhelainen, A. 2008. Security in integration and Enterprise Service Bus (ESB). TietoEnator Corporation (nyk. TietoEvy Corporation). Viitattu 1.8.2023. Saatavissa https://owasp.org/www-pdf-archive/Security_in_integration_and_ESB-OWASP_20091020.pdf
- Pearson Education 2022. Web Services Part 6: Models of Integration. Informit. Viitattu 20.7.2022. Saatavissa <https://www.informit.com/articles/article.aspx?p=28713&seqNum=2>
- Performance Software. 2023. ESB vs. API Gateway: What's the Difference? Akana by Performance. Viitattu 2.9.2023. Saatavissa <https://www.akana.com/blog/esb-vs-api-gateway>

Pratt, M. 2022. The importance of data security in the enterprise. TechTarget. Viitattu 28.6.2023. Saatavissa <https://www.techtarget.com/searchsecurity/feature/The-importance-of-data-security-in-the-enterprise>

Prevalova, I. 2023. 7 Best Practices to Move Away from Manual Data Integration. Viitattu 29.6.2023. Saatavissa <https://www.adverity.com/blog/7-best-practices-to-move-away-from-manual-data-integration>

PRP Business Solutions. 2023. What are the Benefits of Robotic Process Automation (RPA)? PRP Business Solutions. Viitattu 10.7.2023. Saatavissa <https://prpbs.com/en/articles/benefits-of-rpa.html>

Raju, A. 2021. How to explain Robotic Process Automation (RPA) in plain English. American Technology Blog. Viitattu 16.7.2023. Saatavissa <https://blog.american-technology.net/explain-robotic-process-automation/>

Ram, P. 2021. Difference between Session Cookies vs. JWT (JSON Web Tokens), for session management. Viitattu 1.9.2023. Saatavissa <https://medium.com/@prashantramnyc/difference-between-session-cookies-vs-jwt-json-web-tokens-for-session-management-4be67d2f066e>

Reselman, B. 2020. An architect's guide to APIs: SOAP, REST, GraphQL, and gRPC. Viitattu 2.9.2023. Saatavissa <https://www.redhat.com/architect/apis-soap-rest-graphql-grpc>

Richards, M. 2006. The Role of the Enterprise Service Bus. InfoQ. Viitattu 30.7.2023. Saatavissa <https://www.infoq.com/presentations/Enterprise-Service-Bus/>

Rosencrance, L. 2023. 10 types of security incidents and how to handle them. TechTarget. Viitattu 24.6.2023. Saatavissa rajoitetusti <https://www.techtarget.com/searchsecurity/feature/10-types-of-security-incidents-and-how-to-handle-them>

Sakpal, M. 2021. Four Steps to Ensure Robotic Process Automation Security. Gartner. Viitattu 17.7.2023. Saatavissa <https://www.gartner.com/smarterwithgartner/4-steps-to-ensure-robotic-process-automation-security>

Santos, R. 2016. Hub-and-Spoke Architecture. Polarising. Viitattu 25.7.2023. Saatavissa <https://polarising.com/hub-spoke-architecture/>

ServiceNow. 2023. What is robotic process automation (RPA)? Viitattu 3.7.2023. Saatavissa <https://www.servicenow.com/uk/workflows/creator-workflows/what-is-robotic-process-automation.html>

Sitra. 2023a. Dataekosysteemi. Tulevaisuussanasto. Saatavissa

<https://www.sitra.fi/tulevaisuussanasto/dataekosysteemi/>

Sitra. 2023b. Principles-Based Frameworks and Tools for Fair Data Economy. Viitattu

21.6.2023. Saatavissa <https://www.sitra.fi/app/uploads/2021/10/2-pager-principles-based-frameworks-and-tools-nov-2021.pdf>

Smallcombe, M. 2020. ETL vs. ESB:

Key Differences Explained, Pros, Cons & Data Integration Guide. Integrate.io. Viitattu

30.7.2023. Saatavissa <https://www.integrate.io/blog/etl-vs-esb/>

SnapLogic. 2023. Point-to-Point Integration: Advantages, Disadvantages, and Uses.

Viitattu 18.7.2023. Saatavissa <https://www.snaplogic.com/blog/point-to-point-integration-advantages-disadvantages-and-uses>

Soleymanifar, R. 2018. An LP based approximate dynamic programming model to address airline overbooking under cancellation, refund and no-show. Yugoslav Journal of Operations Research. 29: 26-26. DOI: 10.2298/YJOR170918026S. Viitattu 29.8.2023.

Saatavissa rajoitetusti

https://www.researchgate.net/publication/328256621_An_LP_based_approximate_dynamic_programming_model_to_address_airline_overbooking_under_cancellation_refund_and_no-show

Som, S. 2018. SOAP V/S REST. Webkul Software. Viitattu 2.9.2023. Saatavissa

<https://webkul.com/blog/soap-v-s-rest/>

SS&C Blue Prism. 2023. What is RPA? Robotic Process Automation Explained. Viitattu

6.7.2023. Saatavissa <https://www.blueprism.com/guides/robotic-process-automation-rpa/>

Stemmler, K. 2021. 9 Ways To Secure your GraphQL API — GraphQL Security Checklist.

Viitattu 6.9.2023. Saatavissa <https://www.apollographql.com/blog/graphql/security/9-ways-to-secure-your-graphql-api-security-checklist/>

Sythoff, F. 2022. Point-to-Point integrations: 6 Reasons Why They Don't Work for Today's Utilities. Greenbird. Viitattu 21.7.2023. Saatavissa

<https://www.greenbird.com/resources/point-to-point-integrations-6-reasons-why-they-dont-work-for-todays-utilities>

Talend. 2023a. What is Data Integration? Viitattu 20.6.2023. Saatavissa

<https://www.talend.com/resources/what-is-data-integration/>

Talend. 2023b. Data Integration Methods and Strategies. Viitattu 28.6.2023. Saatavissa

<https://www.talend.com/resources/data-integration-methods/>

The GraphQL Foundation. 2023a. A query language for your API. Viitattu 3.9.2023. Saatavissa <https://graphql.org/>

The GraphQL Foundation. 2023b. Schemas and Types. Viitattu 3.9.2023. Saatavissa <https://graphql.org/learn/schema/>

The GraphQL Foundation. 2023c. Queries and Mutations. Viitattu 3.9.2023. Saatavissa <https://graphql.org/learn/queries/>

The GraphQL Foundation. 2023d. Code using GraphQL. Viitattu 3.9.2023. Saatavissa <https://graphql.org/code/#generic-tools>

Tobin, D. 2020. Data Mart vs Data Warehouse: 5 Critical Differences. Integrate.io. Viitattu 30.7.2023. Saatavissa <https://www.integrate.io/blog/data-mart-vs-data-warehouse/>

Tuli, N. 2021. REST API Introduction. Qa Tech Hub. Viitattu 1.9.2023. Saatavissa <https://qatechhub.com/rest-api-introduction/>

Tutorials Point. 2023a. RESTful Web Services – Introduction. Viitattu 1.9.2023. Saatavissa https://www.tutorialspoint.com/restful/restful_introduction.htm

Tutorials Point. 2023b. GraphQL Architecture. Viitattu 3.9.2023. Saatavissa https://www.tutorialspoint.com/graphql/graphql_architecture.htm

Tutorials Point. 2023c. GraphQL Type System. Viitattu 3.9.2023. Saatavissa https://www.tutorialspoint.com/graphql/graphql_type_system.htm

Tutorials Point. 2023d. RESTful Web Services – Security. Viitattu 6.9.2023. Saatavissa https://www.tutorialspoint.com/restful/restful_security.htm#

Tähtinen, S. 2005. Järjestelmäintegraatio. Tarve, vaihtoehdot, toteutus. Helsinki: Talentum. s. 143

Walker, A. 2023. Web Service(Ws) Security Tutorial with SOAP Example. Viitattu 5.9.2023. Saatavissa <https://www.guru99.com/security-web-services.html>

Wallarm. 2023. What is gRPC? Meaning, Architecture, Advantages. Viitattu 4.9.2023. Saatavissa <https://www.wallarm.com/what/the-concept-of-grpc>

Liite 1. Haastattelukirje

Hei!

Olemme kaksi opinnäytetyötä tekevää LAB ammattikorkeakoulun tieto- ja viestintätekniikan opiskelijaa. Teemme opinnäytetöitä datakosysteemeihin liittyen. Datakosysteemi on Sitran tulevaisuussanaston mukaan useista dataverkostoista koostuva verkosto, jossa toimijat tekevät yhteistyötä tavoitteenaan jakaa ja käyttää dataa verkoston sisällä sekä edistää innovointia ja uutta liiketoimintaa. Opinnäytetöidemme aiheet sivuavat toisiaan ja liittyvät samaan hankkeeseen Kyberturvalliset ekosysteemit Etelä-Karjala. Hankkeen tavoitteena on edistää datan jakamiseen perustuvien dataverkostojen kehkeytymistä ja auttaa verkoston yrityksiä rakentamaan toimintaansa kohti reilun datan jakamiseen perustuvia liiketoimintaekosysteemejä. Hanke auttaa pk-yrityksiä tunnistamaan dataan liittyviä liiketoimintamahdollisuuksia ja auttaa yrityksiä luomaan käytäntöjä, joilla ne voivat hyötyä yhteistyössä omistamastaan datasta.

Tavoitteenamme on löytää erilaisia Suomessa toimivia datakosysteemejä ja selvittää, mitkä ovat toimivan datakosysteemiin onnistumisen edellytykset sekä suurimmat haasteet. Marja tekee selvitystä Suomessa toimivista datakosysteemeistä. Heidin työ käsittelee integraatiototeutuksia datan jakamisessa datakosysteemeissä. Töiden tavoitteena on lisätä tietoa ja ymmärrystä datakosysteemeistä.

Haastateltavat henkilöt ja yritykset/yhteisöt anonymisoidaan, jotta niitä ei tunnista valmiista työstä. Tallennamme haastattelun ja lähetämme kysymykset etukäteen, jotta haastatteluun voi valmistautua.

Olisiko sinulla kiinnostusta osallistua Teamsin välityksellä haastatteluun aiheeseen liittyen? Haluaisimme kuulla, millaisia datakosysteemejä teidän yrityksellänne on käytössä ja saada niistä näkemyksiä ja kokemuksia.

Ystävällisin terveisin,

Marja Airaksinen ja Heidi Paananen

tieto- ja viestintätekniikan opiskelijat

LAB ammattikorkeakoulu

marja.airaksinen@student.lab.fi

heidi.paananen@student.lab.fi

Teemme opinnäytetöitä datakosysteemeihin liittyen. Datakosysteemi on Sitran tulevaisuussanaston mukaan useista dataverkostoista koostuva verkosto, jossa toimijat tekevät yhteistyötä tavoitteenaan jakaa ja käyttää dataa verkoston sisällä sekä edistää innovointia ja uutta liiketoimintaa.

Liite 2. LAB_Opinnäytetyö_tietosuojailmoitus

**OPINNÄYTETYÖTÄ KOSKEVA
TIETOSUOJAILMOITUS
EU:n yleinen tietosuoja-asetus (2016/679)
artiklat 13 ja 14**

Laatimispäivämäärä: 22.5.2023

Mitä tarkoitusta varten henkilötietoja kerätään? / Henkilötietojen käsittelyn tarkoitus

Teemme opinnäytetyönä selvitystä Suomessa toimivista dataekosysteemeistä ja niiden integraatiototeutuksista. Työtä varten tarvitsemme tietoa mahdollisimman erilaisista dataekosysteemeistä, niiden hyödyistä liiketoiminnalle ja suurimmista haasteista, sekä teknisestä toteutuksesta.

Mitä tietoja keräämme? / Tutkimusrekisterin tietosisältö

Keräämme yrityksiltä ja yhteisöiltä haastattelemalla henkilöitä. Haastattelun yhteydessä henkilön nimi ja yritys tallentuvat.

Millä perusteella keräämme tietoja? / Henkilötietojen käsittelyn oikeusperuste

Tietojen käsittelyn peruste on yrityksen/yhteisön suostumus sekä henkilön suostumus osallistua haastatteluun.

Mistä kaikkialta henkilötietoja keräämme / Tietolähteet

Keräämme tietoa myönnetyn luvan mukaisesti yrityksestä/yhteisöstä henkilöiltä itseltään haastattelulla sekä mahdollisesti muista julkisesti saatavilla olevista lähteistä.

Kenelle tietoja siirretään? / Tietojen siirto tai luovuttaminen ulkopuolelle

Haastattelutietoja ei luovuteta ulkopuolisille.

Minne tietoja siirretään? / Tietojen siirto tai luovuttaminen EU:n tai Euroopan talousalueen ulkopuolelle

Kerättyjä tietoja ei siirretä EU:n tai Euroopan talousalueen ulkopuolelle, mutta tietoja säilytetään korkeakoulun OneDrive-palvelimilla, jotka sijaitsevat Euroopan talousalueen ulkopuolella.

Kerättyjen tietojen turvallinen säilyttäminen / Rekisterin suojauksen periaatteet

Tietoja säilytetään korkeakoulun tietoturvaisilla palvelimilla ja niihin on pääsy ainoastaan työn tekijöillä Marja Airaksisella ja Heidi Paanasella. Tiedot anonymisoidaan ennen tietojen analysointia.

Kuinka kauan kerättyä aineistoa säilytetään? / Tutkimusaineiston käsittely tutkimuksen päättymisen jälkeen

Tutkimusaineisto poistetaan tutkimuksen päättymisen jälkeen. Korkeakoulun OneDrive-palvelimilla oleva aineisto määritetään vanhenemaan 31.12.2023, jonka jälkeen se säilyy vielä 90 päivää ennen palvelimelta häviämistä.

Millaista päätöksentekoa? / Automatisoitu päätöksenteko

Aineistoa käsiteltäessä ei tapahdu automaattista päätöksentekoa.

Oikeutesi / Rekisteröidyn oikeudet

Rekisteröidyllä on oikeus peruuttaa antamansa suostumus, milloin henkilötietojen käsittely perustuu suostumukseen.

Rekisteröidyllä on oikeus tehdä valitus Tietosuojavaltuutetun toimistoon, mikäli rekisteröity katsoo, että häntä koskevien henkilötietojen käsittelyssä on rikottu voimassa olevaa tietosuojalainsäädäntöä.

Rekisteröidyllä on seuraavat EU:n yleisen tietosuoja-asetuksen mukaiset oikeudet:

- a) Rekisteröidyn oikeus tarkistaa itseään koskevat tiedot.
- b) Rekisteröidyn oikeus tietojensa oikaisemiseen.
- c) Rekisteröidyn oikeus tietojensa poistamiseen. Oikeutta henkilötietojen poistamiseen ei sovelleta, jos tietojen käsittely on tarpeen yleisen edun mukaisia arkistointitarkoituksia taikka tieteellisiä tai historiallisia tutkimustarkoituksia tai tilastollisia tarkoituksia varten, jos oikeus tietojen poistamiseen estää tai suuresti vaikeuttaa henkilötietojen käsittelyä.
- d) Rekisteröidyn oikeus tietojen rajoittamiseen.
- e) Rekisteröidyn oikeus siirtää tiedot toiselle rekisterinpitäjälle.
- f) Rekisteröidyn oikeus vastustaa tietojensa käsittelyä, kun käsittely perustuu yleistä etua koskevaan tehtävään, rekisterinpitäjälle kuuluvaan julkiseen valtaan tai rekisterinpitäjän tai kolmannen osapuolen oikeutettuun etuun.

EU:n yleisen tietosuoja-asetuksen mukaiset rekisteröidyn oikeudet eivät ole automaattisia kaikessa henkilötietojen käsittelyssä.

Tutkimusrekisterin tiedot

Rekisterin nimi on Haastattelut. Kyseessä on kertatutkimus. Tutkimukset kestävät pisimmillään vuoden 2023 loppuun (kyseessä on kaksi tutkimusta samasta aineistosta). Henkilötietoja säilytetään tutkimuksen keston ajan.
--

Rekisterinpitäjän ja yhteyshenkilön tiedot

Marja Airaksinen marja.airaksinen@student.lab.fi

Tutkimuksen suorittajat

Marja Airaksinen Heidi Paananen

Liite 3. Haastattelukysymykset

Haastattelukysymykset

Marjan aihe:

1. Mikä on dataekosysteemin tavoite, mihin tarkoitukseen se on tehty?
2. Mitä dataa jaetaan?
3. Onko datan jakamisessa rajoituksia? (esim. jotain tietoa ei jaeta ollenkaan tai jaetaan vain osa tiedosta)
4. Millaisia tahoja dataekosysteemissä on mukana ja miten laaja se on?
5. Ovatko toimijat toisistaan riippumattomia?
6. Miten eri tahot saatiin hankkeeseen mukaan?
7. Millaisia ongelmia oli vai oliko?
8. Miten mahdolliset ongelmat ratkaistiin?
9. Millaisia teknologioita datan jakamiseen käytetään? (Tähän tulee vastaus Heidin kysymyksissä)
10. Mitä dataekosysteemin avulla on saavutettu?
11. Onko tuloksia mitattu? Jos on, niin miltä tulokset näyttävät?
12. Onko yhteistyö ollut onnistunutta?
13. Mitkä ovat olleet tämän dataekosysteemin menestystekijöitä?

Heidin aihe:

1. Mikä on yrityksenne koko, toimiala ja datan määrä?
2. Millaisia dataintegraatoratkaisuja yrityksessänne käytetään tällä hetkellä datan jakamisessa yrityksenne sisällä? Millaisia dataintegraatoratkaisuja yrityksessänne käytetään datan jakamisessa yrityksenne ulkopuolisten toimijoiden kanssa? Millä kriteereillä päätökset on tehty? Harkitsetteko hyödyntävänne tulevaisuudessa joitakin muita integraatoratkaisuja datan jakamiseen ulkopuolisten toimijoiden kanssa?
3. Mitkä ovat olleet dataintegraation mahdollistaman datan jakamisen suurimmat hyödyt ja mahdollisuudet liiketoiminnallenne? Onko integraatiolla ollut vaikutusta liiketoiminnan kasvuun tai uusien mahdollisuuksien avaamiseen? Minkälaisia muita hyötyjä olette saavuttaneet datan jakamisessa yrityksen sisällä ja muiden yritysten kanssa?
4. Kuinka hyvin dataintegraatoratkaisunne on tukenut yhteistyötä muiden yritysten kanssa datan jakamisen kautta?
5. Minkälaisia haasteita olette kohdanneet datan integroinnissa eri järjestelmien välillä?
6. Miten olette hallinneet ja ylläpitäneet integraatiota datan jakamisen näkökulmasta?
7. Miten tietoturva ja tietosuoja-asiat otetaan huomioon datan jakamisessa ja integroinnissa?
8. Miten olette kouluttaneet ja tukeneet henkilöstöänne integraation käytössä ja hallinnassa?
9. Onko teillä ollut tarvetta sopeuttaa tai muuttaa integraatiota ajan mittaan ja millaisia mahdollisia haasteita tämä on tuonut?
10. Millaisia mahdollisuuksia näette dataintegraatoratkaisujen tarjoavan yrityksille datan jakamisessa ja yhteistoiminnassa?
11. Minkälaisia parhaita käytäntöjä olette kehittäneet integraation hyödyntämisessä datan jakamisessa?
12. Mitä suosituksia antaisitte muille yrityksille, jotka harkitsevat dataintegraation toteuttamista ja hyödyntämistä?

Onko jotain muuta, mitä meidän olisi hyvä tietää?

Liite 4. Opinnäytetyön_suostumuslomake



Suostumus koskien henkilötietojen käsittelyä ja opinnäytetyöhön osallistumista

Minua on pyydetty osallistujaksi opinnäytetyöhön "Dataintegraatoratkaisut yritysten dataekosysteemeissä" liittyvään aineiston keräämiseen. Opinnäytetyön tekijä on Heidi Paananen ja hän tekee opinnäytetyötään LAB-ammattikorkeakoulussa. Tällä lomakkeella ilmoitan halukkuudestani osallistua opinnäytetyön aineiston keräämiseen.

Osallistumiseni edellä mainittuun opinnäytetyöhön on täysin vapaaehtoista. Minulla on oikeus milloin tahansa peruuttaa suostumukseni tai keskeyttää opinnäytetyöhön osallistuminen, eikä minun tarvitse kertoa syytä päätökselleni. Suostumuksen peruuttamisesta tai osallistumisen keskeyttämisestä ei aiheudu minulle mitään kielteisiä seurauksia. Jos keskeytän osallistumiseni, minusta siihen mennessä kerättyjä tietoja voidaan kuitenkin yhä käyttää opinnäytetyössä.

Olen saanut riittävästi tietoa opinnäytetyöstä ja minulla on ollut myös mahdollisuus esittää opinnäytetyötä ja aineiston keräämistä koskevia kysymyksiä. Minulle on annettu aineiston keräämistä koskeva tiedote. Lisäksi minua on informoitu aineiston keräämiseen liittyvästä henkilötietojen käsittelystä tietosuojailmoituksella.

Allekirjoittamalla tämän lomakkeen annan suostumukseni osallistua edellä mainitun opinnäytetyön aineiston keräämiseen sekä suostumukseni henkilötietojeni käsittelyyn tietosuojailmoituksessa kerrotulla tavalla.

Osallistujan allekirjoitus

[Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.](#)

Osallistujan nimenselvennys

Valitse [päivämäärä](#) [Paikka](#)
Päivämäärä ja paikka