



Microsoft Fabricin alkukartoitus ja vertailututkimus

Ammattikorkeakoulututkinnon opinnäytetyö

Tietojenkäsittelyn koulutus

Kevät 2025

Erika Mikonmaa

Koulutus Tietojenkäsittelyn koulutus
Tekijä Erika Mikonmaa
Työn nimi Microsoft Fabricin alkukartoitus ja vertailututkimus
Ohjaaja Tero Keso

Vuosi 2025

Tämän opinnäytetyön tavoitteena oli osallistua Microsoft Fabricin alkukartoitukseen osana kolmivaiheista käyttöönottoa. Työssä tarkasteltiin Microsoft Fabricin ominaisuuksia sekä sen eroavaisuuksia Power BI:hin. Lisäksi tavoitteena oli selvittää, miten Microsoft Fabric voisi ratkaista yrityksen nykyisiä haasteita ja tukea liiketoiminnallisia tavoitteita, kuten Power BI -raportoinnin tehostamista ja datan hallinnan yksinkertaistamista. Opinnäytetyön toimeksiantajana toimi Purso Group Oy, joka on suomalainen alumiiniteollisuuden alalla toimiva yritys.

Opinnäytetyön teoreettisessa osuudessa käsiteltiin liiketoimintatiedon hallintaa ja sen keskeisiä vaiheita. Lisäksi tarkasteltiin vertailtavien järjestelmien ominaisuuksia sekä lyhyesti Microsoft Azurea alustapalveluna. Työssä toteutettiin vertailututkimus, joka aloitettiin nykytilan analyysillä. Nykytilan analyysissä selvitettiin mitä eri järjestelmiä yrityksellä on käytössä ja miten data -ja analytiikka ratkaisut ovat hallinnoitu Power BI ympäristöön. Tutkimuksessa selvitettiin toimeksiantajan tarpeita ja tavoitteita Microsoft Fabriciin siirtymiselle. Tutkimuksessa käytettiin laadullista tutkimusmenetelmää ja aineistoa kerättiin osallistumalla tapaamisiin, joissa valittiin Microsoft Fabricin käyttöönotolle kumppania sekä haastatteleamalla pientä joukkoa ihmisiä. Opinnäytetyö on toiminnallinen.

Työn tuloksena syntyi järjestelmävertailu, jossa havainnollistettiin Microsoft Fabricin ja Power BI:n eroavaisuuksia sekä arvioitiin käyttöönoton vaikutuksia. Tulokset tarjoavat tietoa Microsoft Fabricin ominaisuuksista ja sen käyttöönoton vaikutuksista. Lisäksi ne voivat toimia tukena muille yrityksille, jotka harkitsevat sen käyttöönottoa. Toimeksiantaja on ollut tyytyväinen työn teoreettiseen lähestymistapaan ja saadun palautteen mukaan teoriaosuus tukee hyvin tulevia käytännön toteutuksia. Koska työn aikana ei vielä päästy käytännössä toteuttamaan alustan perustamista, työhön koottua tietoa pidetään hyödyllisenä myös jatkovaiheita ajatellen.

Avainsanat Liiketoimintatiedon hallinta, data, Microsoft Fabric, Power BI

Sivut 49 sivua ja liitteitä 1 sivu

DP Degree Programme in Business Information Technology
Author Erika Mikonmaa Year 2025
Subject Initial Assessment and Comparative Research of Microsoft Fabric
Supervisors Tero Keso

The goal of this thesis was to participate in the initial assessment of the Microsoft Fabric implementation as part of a three-phase project. The thesis examines the features of Microsoft Fabric and its differences from Power BI. The purpose of the thesis was to determine how Microsoft Fabric could address the company's current challenges and support its business goals, such as making Power BI reporting more efficient and simplifying data management. The thesis was commissioned by Purso Group Oy, a Finnish company operating in the aluminum industry.

The theoretical part of this thesis covers the business intelligence and its key phases. It examines the features of the compared systems and provides a brief overview of Microsoft Azure which is the platform-as-a-service for Microsoft Fabric. A comparative research approach was used, beginning with a current state analysis. This analysis identified the different systems the company is using and how data and analytics solutions are managed within the Power BI environment. The study also explored the commissioner's needs and objectives for migrating to Microsoft Fabric. A qualitative research method was applied, with data collected through participation in meetings where a partner was selected for the Microsoft Fabric implementation and by interviewing a small group of people. This thesis is practical in nature.

The research demonstrates that Microsoft Fabric offers significant opportunities for improving reporting and data management processes. The findings provide insight about the features of Microsoft Fabric and its implementation effects. Additionally, they can serve as a useful resource for other companies considering its adoption. The commissioner was pleased with the study's theoretical approach, and feedback indicates that the theoretical section provides strong support for future practical implementations. Since the platform setup was not yet conducted during the study, the gathered information is considered valuable for the next phases.

Keywords Business Intelligence, data, Microsoft Fabric, Power BI
Pages 49 pages and appendices 1 page

Sisällys

1	Johdanto	1
2	Liiketoimintatiedon hallinta	3
2.1	Datan määrittely	3
2.2	Datan hallinta ja varastointi	4
2.2.1	Tietovarasto	5
2.2.2	Data-allas	5
2.2.3	Data Lakehouse	6
2.2.4	Medallion-arkkitehtuuri	7
2.3	Dataintegraatio	9
2.3.1	ETL-prosessi	9
2.3.2	ELT-prosessi	10
2.3.3	CDC	11
2.3.4	Dataputket	12
2.3.5	Datan orkestrointi	13
2.4	Data-analytiikka	14
2.4.1	Prosessin keskeiset vaiheet	14
2.4.2	Pilvianalytiikka	15
2.5	Raportointi ja päätöksenteko	16
3	Vertailussa olevat järjestelmät	18
3.1	Microsoft Azure alustapalveluna	18
3.2	Microsoft Fabric	19
3.2.1	Alustan keskeiset komponentit	19
3.2.2	Datan varastointi	20
3.2.3	Datan hallinta ja käsittely	21
3.2.4	Tekoäly	22
3.2.5	Operatiivinen data	23
3.2.6	Turvallisuus ja hallinta	24
3.3	Power BI	24
3.3.1	Keskeiset osat ja lisenssit	25
3.3.2	Datan varastointi ja hallinta	26
3.3.3	Keskeiset ominaisuudet	27
3.3.4	Tekoäly	28

3.3.5	Turvallisuus ja hallinta	29
4	Työn tavoite ja menetelmät	31
5	Nykytila ja käyttöönoton alkukartoitus.....	33
5.1	Nykytilan analyysi	33
5.2	Tarpeet ja tavoitteet	34
5.3	Keskeiset vaiheet Microsoft Fabricin käyttöönotolle	35
6	Vertailututkimus	37
7	Tulokset	39
8	Johtopäätökset ja pohdinta	41
9	Yhteenveto.....	43
	Lähteet.....	44

Sanasto

ACID	Atomisuus (Atomicity), eheys (Consistency), eristyneisyys (Isolation) ja pysyvyys (Durability) ovat tietokantajärjestelmien periaatteita, jotka varmistavat tietojen eheyden ja luotettavan käsittelyn
Aggregointi	Tietojen keräämistä ja yhdistämistä
Apache Spark	Avoimen lähdekoodin hajautettu tietojenkäsittelyjärjestelmä, joka mahdollistaa suurten tietomäärien käsittelyn ja analysoinnin tehokkaasti
Big data	Suuria ja nopeasti kasvavia tietomassoja, joita analysoidaan kehittyneillä teknologioilla ja menetelmillä
CDC	Change Data Capture, prosessi, joka seuraa ja tallentaa tietokannan muutokset reaaliaikaisesti, kuten lisäykset, päivitykset ja poistot, mahdollistaen ajantasaisen datan käytön
CRM	Customer Relationship Management, asiakkuudenhallintajärjestelmä
Data lake	Data-allas; tietoallas, johon voidaan tallentaa suuria määriä dataa eri muodoissa.
DAX	Power BI:ssä käytetty kieli, joka sisältää funktioita, operaattoreita ja muotoiluja laskennallisten arvojen ja tietomallien analysointiin.
Data Warehouse	Tietovarasto strukturoidun tiedon varastointiin, johon integroidaan dataa useista eri lähteistä analysointia ja raportointia varten
Data mart	Tietovaraston osa, joka on suunniteltu palvelemaan tietyn liiketoiminta-alueen tai käyttäjäryhmän tarpeita.
Dashboard	Visuaalinen käyttöliittymä, joka kokoaa yhteen keskeiset mittarit, raportit ja analyysit, tarjoten reaaliaikaisen yleiskuvan datasta päätöksenteon tueksi
ERP	Enterprise Resource Planning, toiminnanohjausjärjestelmä
ELT	Extract, Load, Transform, data siirretään kohdejärjestelmään heti sen hakemisen jälkeen eri lähteistä ja varsinainen muokkaaminen tapahtuu vasta lopuksi
ETL	Extract, Transform, Load tarkoittaa datan siirtämistä, muokkaamista ja lataamista tietovarastoon
GDPR	General Data Protection Regulation, EU:n yleinen tietosuojasäätös

IoT	Internet of Things, esineiden, koneiden ja laitteiden yhdistämistä internettiin
KPI-mittari	Key Performance Indicator, liiketoiminnan suorituskyvyn mittari
Loki	Tallenne tapahtumista, joka on aikajärjestyksessä kirjattu, esimerkiksi muutokset ja tapahtumat tietojärjestelmissä
MES	Manufacturing Execution System, tuotannonohjausjärjestelmä
MFA	Monivaiheinen tunnistautuminen
NLP	Natural language processing, luonnollisen kielen käsittely
On-premise	Ohjelmistojen ja IT-infrastruktuurin hallintaa organisaation omilla palvelimilla
PaaS	Platform-as-a-service, palvelualusta, joka tarjoaa kehittäjille valmiin ympäristön sovellusten kehittämiseen ja käyttöönottoon ilman tarvetta huolehtia taustalla olevasta infrastruktuurista
PoC	Proof of Concept, on kokeilu- tai testivaihe, jossa arvioidaan, onko jokin idea, teknologia tai ratkaisu toteutuskelpoinen ennen laajempaa käyttöönottoa.
Raakadata	Data alkuperäisessä, jalostumattomassa muodossa
RBAC	Roolipohjainen pääsynvalvonta
Relaatiotietokanta	Tietokanta, jossa data tallennetaan taulukkomuotoon riveinä ja sarakkeina.
RLS	Row level security, rivikohtainen tietosuoja
SaaS	Software as a Service, pilvipohjainen ohjelmistopalvelu
Skeema	Määrittää, kuinka tiedot järjestellään. Sisältää esimerkiksi sarakkeet, taulujen nimet ja datatyypit
SSIS	SQL Server Integration Services, tietojen integrointi työkalu
SQL	Structured Query Language, ohjelmointikieli, jota käytetään relaatiotietokantojen hallintaan ja muokkaamiseen.
SQL Server	Microsoftin kehittämä relaatiotietokannan hallintajärjestelmä
SSO	Single Sign-On, käyttäjän on mahdollista päästä useisiin sovelluksiin yhdellä kirjautumiskerralla

Kuvat

Kuva 1. Liiketoimintatiedon hallinnan keskeiset vaiheet (IBM, 2021).....	3
Kuva 2. Tietovarasto, Data-allas ja Data Lakehouse (Qlik, n.d.a).....	4
Kuva 3. Medallion-arkkitehtuuri (Databricks, 2022)	7
Kuva 4. ETL-prosessin vaiheet (Microsoft, n.d.a).....	10
Kuva 5. ELT-prosessin vaiheet (Microsoft, n.d.a).....	11
Kuva 6. Lokipohjainen CDC-menetelmä (Qlik, n.d.b)	12
Kuva 7. Dataputki (Gulbahar Karasatas, 2024)	13
Kuva 8. Microsoft Fabricin keskeiset komponentit (Microsoft, 2025d)	20
Kuva 9. Power BI:n osat (Microsoft, 2024a)	25
Kuva 10. Keskeiset vaiheet Microsoft Fabricin käyttöönotolle	36

Taulukot

Taulukko 1. Power BI:n ja Microsoft Fabric:in vertailun tulokset	39
--	----

Liitteet

Liite 1. Aineistohallintasuunnitelma	
--------------------------------------	--

1 Johdanto

Tämän opinnäytetyön aiheena on Microsoft Fabricin käyttöönoton alkukartoitus sekä vertailututkimus yrityksen nykyisistä Power BI -ratkaisusta verrattuna Microsoft Fabriciin. Opinnäytetyön toimeksiantajana toimii Purso Group Oy, joka on suomalainen alumiiniteollisuuden alalla toimiva yritys. Purson toiminta alkoi vuonna 1959 Nokian Siurossa. Konserniin kuuluvat Suomen toimipisteiden lisäksi Alankomaissa Nedal Aluminium, Ruotsissa Purso AB ja Saksassa Purso GmbH. Purso kehittää erilaisia alumiinista valmistettuja ratkaisuja teollisuuteen sekä järjestelmäratkaisuja rakennuksiin ja led-valaistukseen. (Purso, n.d.)

Toimeksiantajan siirtymistä Microsoft Fabriciin perustellaan useilla syillä. Raportointia halutaan kehittää Power BI-ympäristössä, koska tällä hetkellä raportit ovat epä johdonmukaisia ja käyttäjät kokevat vaikeuksia löytää tarvitsemiaan tietoja. Yritys on vaihtamassa ERP-toiminnanohjausjärjestelmäänsä ja sen vuoksi on tarve päivittää myös datan hallinta- ja analytiikkaratkaisuja, jotta varmistetaan niiden yhteensopivuus uuden järjestelmän kanssa. Yritys tavoittelee myös parempia lähtökohtia tekoälyn hyödyntämiselle, jonka vuoksi pilveen siirtyminen on tarpeellista. Yrityksellä on useita tehtaita ja toimipisteitä, esimerkiksi Alankomaissa, joten yhtenäisen ja keskitetyn pilvipohjaisen alustan käyttöönotto nähdään tärkeänä askeleena konsernin kehittämisessä.

Työn tavoitteena on osallistua Microsoft Fabricin alkukartoitukseen, joka on osana kolmivaiheista käyttöönottoa. Tavoitteena selvittää, miten se voisi ratkaista yrityksen nykyisiä haasteita ja tukea liiketoiminnallisia tavoitteita, kuten Power BI -raportoinnin tehostamista ja datan hallinnan yksinkertaistamista. Lopputuloksena syntyy järjestelmävertailu, jossa analysoidaan yrityksen nykyiset data- ja analytiikkaratkaisut sekä arvioidaan Microsoft Fabricin ominaisuuksia ja hyötyjä suhteessa nykyisiin Power BI -ratkaisuihin.

Tutkimuskysymykset ovat:

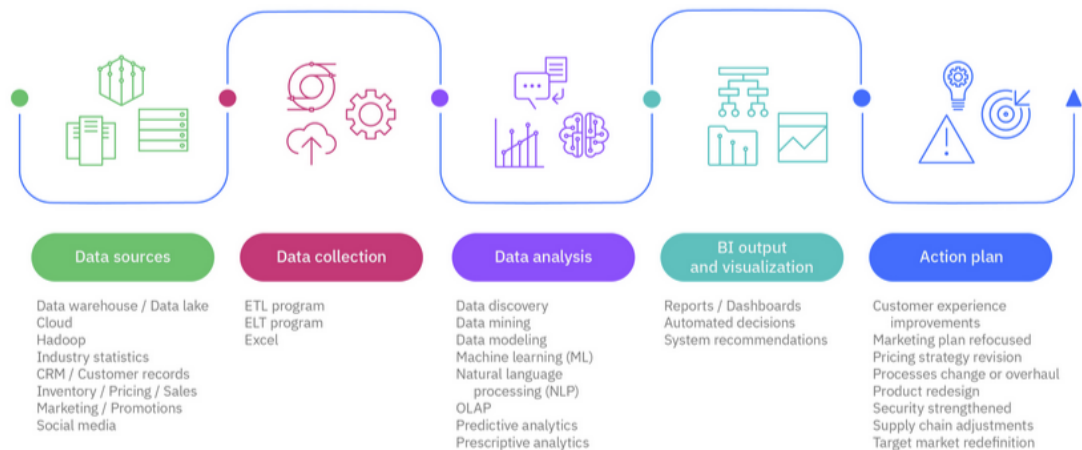
- Miten Microsoft Fabric toimii ja mitä ominaisuuksia se tarjoaa verrattuna Power BI ratkaisuihin?

- Miten Microsoft Fabricin käyttöönotto vaikuttaa yrityksen datan hallintaan ja analytiikkaprosesseihin?
- Mitkä ovat keskeiset vaiheet Microsoft Fabricin käyttöönotossa Azure-alustalle, ottaen huomioon yrityksen tarpeet ja tavoitteet?

2 Liiketoimintatiedon hallinta

Liiketoimintatiedon hallinnalla (Business Intelligence, BI) tarkoitetaan prosessia, jossa organisaatiot keräävät, analysoivat ja hyödyntävät dataa päätöksenteon tueksi (kuva 1.). Liiketoimintatiedon hallinnan avulla organisaatiot voivat muuttaa raakadatan käyttökelpoisempaan muotoon ja saada arvokkaita oivalluksia, joilla parantaa liiketoimintastrategioita ja tehokkuutta. Oivallukset syntyvät yhdistämällä ja analysoimalla informaatiota, jota datasta saadaan. Oivalluksien avulla yritykset voivat tehdä päätöksiä tietoon perusten ja parantaa liiketoimintaansa. (Kumar V K, 2021, ss.7–12)

Kuva 1. Liiketoimintatiedon hallinnan keskeiset vaiheet (IBM, 2021)



2.1 Datan määrittely

Kaikki digitaalisessa muodossa oleva tieto on dataa. Data voi olla monessa eri muodossa, esimerkiksi tietoaaineistona, asiakirjana tai kuvatallenteena. (Erkkilä ym., 2023, s. 17) Data voi olla peräisin monista eri lähteistä ja olla tyypiltään erilaista. Päätyypit ovat strukturoitu, strukturoimaton ja puolistrukturoitu data. Strukturoitu data on järjestetty selkeään, ennalta määriteltyyn muotoon, kuten taulukoihin relaatiotietokannoissa. Strukturoimaton data on vapaamuotoista ja vaikeammin jäseneltävää, kuten sähköpostit, kuvat ja videot. Puolistrukturoitu data on näiden kahden välimuoto ja sellaisia voivat olla esimerkiksi lokitiedot. (Altaiar ym., 2021)

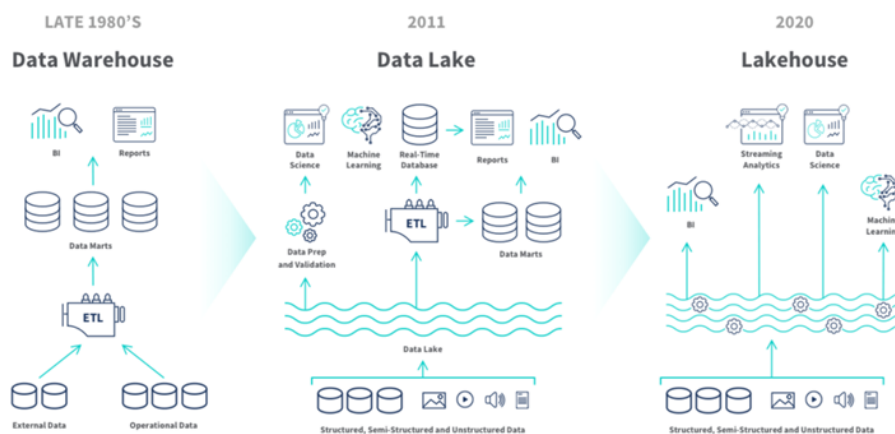
Big data tarkoittaa valtavia tietomassoja, jotka syntyvät eri lähteistä, kuten antureista, sosiaalisista verkostoista ja älypuhelimista ja joita on vaikea hallita perinteisillä työkaluilla. Sen ominaisuuksia kuvataan usein viiden V:n avulla, joita ovat määrä (volume), nopeus (velocity), moninaisuus (variety), arvo (value) ja luotettavuus (veracity). Big datan käsittelyyn tarvitaan erityisiä teknologioita, kuten pilvipalveluja, jotka tarjoavat skaalautuvia ja kustannustehokkaita ratkaisuja tiedon tallentamiseen ja analysointiin. Näiden avulla voidaan yhdistää koneoppimista ja löytää tietomassoista piileviä malleja ja tehdä älykkäitä päätöksiä. (Sandhu, 2022)

Datan kerääminen eri lähteistä, kuten ERP- ja CRM-järjestelmistä, verkkosivustoilta ja IoT-laitteilta vaatii tehokasta integraatiota ja muuntamista käyttökelpoiseen muotoon. Tämän prosessin aikana on tärkeää huomioida datan laatu ja turvallisuus, erityisesti lainsäädännöllisten vaatimusten, kuten GDPR:n näkökulmasta. Pilvianalytiikka tarjoaa keinoja näiden haasteiden ratkaisemiseen ja mahdollistaa joustavan ja tehokkaan analytiikan eri datalähteistä. (Google Cloud, n.d.b)

2.2 Datan hallinta ja varastointi

Kuvassa 2 on esitelty kolme käytetyintä datan varastointiratkaisua. Perinteisesti datan varastointiin on käytetty tietovarastoa (Data Warehouse) ja Data-altaita (Data Lake). Uudempi varastointimalli on nimeltään Data Lakehouse, joka yhdistää näiden kahden edellisen varastointiratkaisun parhaat puolet. (Qlik, n.d.a)

Kuva 2. Tietovarasto, Data-allas ja Data Lakehouse (Qlik, n.d.a)



2.2.1 Tietovarasto

Tietovarasto on keskitetty järjestelmä strukturoidun tiedon varastointiin, johon integroidaan dataa useista eri lähteistä analysointia ja raportointia varten. Tietovarastot hyödyntävät usein moniulotteista tietomallinnusta, jossa liiketoimintatapahtumat kuten myyntitapahtumat mallinnetaan faktoina. Faktoja jäsennetään eri näkökulmista kuten ajan, asiakkaiden tai muiden liiketoimintaan liittyvien ulottuvuuksien perusteella, mikä mahdollistaa monipuolisen ja tehokkaan tiedon analysoinnin. (da Costa Santos, 2020, ss.9–10)

Tietovarastoja voidaan rakentaa kahdella yleisimmällä lähestymistavalla, joita ovat Bill Inmonin top-down-malli ja Ralph Kimballin bottom-up-malli. Inmonin mallissa tietovarasto toimii yrityksen keskitettynä tietovarastona, josta analyysiä varten rakennetaan erillisiä Data marteja eri liiketoiminta-alueille. Kimballin mallissa aloitetaan Data marteista, joita yhdistämällä rakennetaan laajempi tietovarasto. Dimensiomallinnus on keskeinen osa tietovarastojen suunnittelua ja siinä tiedot organisoidaan faktatauluiksi ja dimensiotauluiksi, jotka mahdollistavat tehokkaan tiedonhaun ja analyysin. (da Costa Santos, 2020, ss.10–11)

Perinteiset tietovarastot sijaitsevat yrityksen omissa tiloissa ja vaativat raskasta infrastruktuuria, mikä tekee niiden skaalaamisesta ja ylläpidosta kallista ja rajoitettua. Nykyään monet organisaatiot siirtyvät pilvipohjaisiin tietovarastoihin, jotka tarjoavat joustavuutta, skaalautuvuutta ja edullisempia kustannuksia, koska palveluntarjoaja huolehtii infrastruktuurista. (Google Cloud, n.d.a)

Tietovarastoja käytetään reaaliaikaisten päätösten tekemiseen, eristyneiden datalähteiden yhdistämiseen ja liiketoimintaraporttien tuottamiseen. Ne ovat erityisen hyödyllisiä yrityksille, jotka tarvitsevat suurta datan analysointia ja reaaliaikaista raportointia. Lisäksi ne tukevat koneoppimista ja tekoälyä keräämällä sekä historiallista että reaaliaikaista dataa ennakoivien analyysien tekemiseksi. Se tekee niistä keskeisen ratkaisun organisaatioille, joilla on useita datalähteitä ja big datan analysointitarpeita. (Google Cloud, n.d.a)

2.2.2 Data-allas

Data-allas on alun perin otettu teollisuudessa käyttöön keskitetyksi paikaksi raakadatalle. Data-altaaseen pystytään keräämään raakadataa yrityksen sisäisistä ja ulkoisista lähteistä.

Data-allas arkkitehtuuri tarjoaa joustavuutta ja tehokkuutta suurten datamäärien hallintaan, koska se mahdollistaa erilaisten tietotyyppien, kuten strukturoidun ja strukturoimattoman datan tallentamisen. (Ait Errami ym., 2023)

Data-altaat ovat tärkeitä liiketoiminnassa, sillä ne mahdollistavat raakadatasta johdonmukaisen ja suojatun tallentamisen sekä sen saatavuuden. Esimerkiksi pilvipohjaiset ratkaisut kuten Azure Data Lake tarjoavat skaalautuvia työkaluja, jotka tukevat erilaisten työkuormien käsittelyä, kuten SQL-kyselyitä ja suoratoistoanalytiikkaa. Data-altaita hyödynnetään monilla aloilla innovaatioiden edistämiseksi, kuten suoratoistopalveluissa asiakaskäyttäytymisen analysoimiseksi, rahoitusallalla markkinariskien hallitsemiseksi, terveydenhuollossa potilaspolkujen parantamiseksi ja digitaalisten toimitusketjujen yhdistämiseksi, jolloin varastotieto eri järjestelmistä saadaan keskitetysti yhteen paikkaan. (Microsoft, n.d.b)

2.2.3 Data Lakehouse

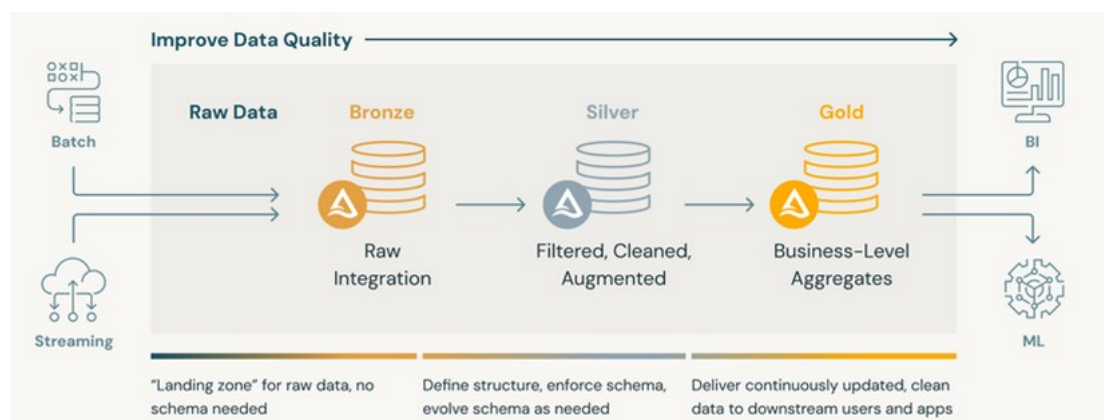
Lakehouse on uusi datanhallinnan arkkitehtuuri, joka yhdistää tietovaraston ja data-altaan parhaat puolet yhdeksi järjestelmäksi. Se hyödyntää pilvipohjaista tallennustekniikkaa ja tarjoaa monia etuja perinteisiin järjestelmiin verrattuna. Lakehouse tukee muun muassa transaktioita, tiedon laadun valvontaa ja analytiikkaa suoraan lähdetiedoista, mikä vähentää viivettä ja parantaa tietojen ajantasaisuutta. Sen avulla voidaan käsitellä monenlaisia tietotyyppisiä ja se tukee useita työkuormia kuten SQL-analytiikkaa, data-analyysiä ja koneoppimista. (Ben Lorica ym., 2020)

Lakehouse arkkitehtuurin ansiosta organisaatiot voivat tallentaa ja käsitellä erilaisia tietotyyppisiä ilman raskaita ETL-prosesseja. Tämä vähentää datan esikäsittelyn tarvetta ja mahdollistaa nopeamman pääsyn tuoreeseen analytiikkaan ja päätöksentekoon. Se tarjoaa myös vahvemman hallinnan ja luotettavuuden kuin perinteiset data-altaat. Sen ACID-yhteensopivuus takaa, että tietojen kirjoittaminen ja päivittäminen tapahtuu turvallisesti ilman tietojen korruptoitumisen riskiä. Tämän lisäksi Lakehouse mahdollistaa tietovaraston ja data-altaan yhdistämisen mikä helpottaa tietojen hallintaa ja sääntöjen noudattamista kuten GDPR:n vaatimuksia. (Ait Errami ym., 2023)

2.2.4 Medallion-arkkitehtuuri

Medallion-arkkitehtuuri on Lakehouse ympäristöön suunniteltu arkkitehtuurimalli. Mallin tavoitteena on parantaa datan laatua ja rakennetta vaiheittain. Se rakentuu kolmesta erillisestä kerroksesta (kuva 3.), joista jokainen kuvaa tallennetun datan laatua. Korkeammat kerrokset edustavat parempaa laatua, mikä mahdollistaa luotettavan ja johdonmukaisen datan hyödyntämisen yrityksen datatuotteissa. (Microsoft, 2023)

Kuva 3. Medallion-arkkitehtuuri (Databricks, 2022)



Kuvassa 3 kuvataan, miten data siirtyy kolmen kerroksen läpi Medallion-arkkitehtuurissa. Pronssikerroksessa tallennetaan raakadata alkuperäisessä muodossaan, hopeakerroksessa data puhdistetaan ja standardisoidaan. Viimeisessä kultakerroksessa dataa rikastetaan vastaamaan liiketoiminnan ja analytiikan tarpeita. Tämä lähestymistapa takaa datan eheydelle, johdonmukaisuuden, eristyksen ja kestävyysominaisuudet (ACID) datan kulkiessa eri kerrosten läpi, mikä optimoi datan analysoinnin tehokkuuden. (Microsoft, 2023)

Pronssikerros sisältää datan alkuperäisessä muodossa ilman että sille on tehty vielä mitään muutoksia. Pronssikerros ei sisällä vielä validoitua tai käsiteltyä tietoa vaan säilyttää sen alkuperäisessä muodossa. Pronssikerrokseen tallennettu data kasvaa ajan myötä, koska uudet tiedot lisätään sinne jatkuvasti ilman, että aiempia tietoja poistetaan. Tämän kerroksen ensisijainen tarkoitus ei ole tarjota suoraa pääsyä analyytikoille tai data-asiantuntijoille, vaan se toimii lähtökohtana myöhemmälle datan rikastamiselle ja siirtämiselle hopeakerrokseen. (Microsoft, 2024b)

Pronssikerros säilyttää kaiken historiallisen datan. Se mahdollistaa tiedon uudelleen käsittelyn ja auditoinnin tarvittaessa. Data voi saapua eri lähteistä, kuten pilvitallennuspalveluista, viestinvälitysjärjestelmistä tai hajautetuista järjestelmistä. Pronssikerroksessa suoritetaan vain vähäinen määrä tietojen puhdistusta tai validointia, jolla vältetään tietojen katoaminen odottamattomien muutosten vuoksi. (Microsoft, 2024b)

Hopeakerroksessa data puhdistetaan ja validoidaan. Pronssikerroksesta tullut data muokataan luotettavampaan ja analysoitavampaan muotoon. Hopeakerroksessa suoritetaan useita keskeisiä toimintoja datan laadun varmistamiseksi. Näihin kuuluu skeeman tarkastaminen, puuttuvien arvojen käsittely, päällekkäisten tietueiden poistaminen sekä myöhässä saapuvan tai väärässä järjestyksessä olevan datan korjaaminen. Lisäksi datan eheyttä parannetaan laadunvalvontatarkistuksilla, skeeman evoluution hallinnalla ja tietotyyppien muunnoksilla. Tämä vaihe varmistaa, että data on eheää ja virheetöntä ennen sen siirtämistä eteenpäin. (Microsoft, 2024b)

Hopeakerros on myös ensimmäinen vaihe, jossa voidaan aloittaa datamallinnus. Tässä vaiheessa voidaan määrittää, miten moniulotteinen tai puolistrukturoitu data esitetään. Data voidaan myös normalisoida useisiin tauluihin tai sen skeemaa voidaan suoristaa analyysin tarpeiden mukaan. Hopeakerros toimii välietappina ennen kultakerrosta ja varmistaa, että jokaisesta tietueesta on vähintään yksi validoitu ja aggregoimaton versio, jotta jatkokäsittely ja analysointi sujuisivat tehokkaasti. (Microsoft, 2024b)

Kultakerros sisältää jalostettua ja analysoitavaksi optimoitua dataa, jota käytetään raportoinnissa, analytiikassa ja liiketoiminnan päätöksenteossa. Tieto on usein aggregoitua ja suodatettua esimerkiksi tiettyjen aikajaksojen tai maantieteellisten alueiden mukaan, jotta se vastaisi paremmin liiketoiminnan tarpeisiin. Kultakerros auttaa varmistamaan, että organisaation käytettävissä on luotettavaa ja selkeästi jäsenneltyä dataa, joka tukee tehokasta analysointia. (Microsoft, 2024b)

Kultakerroksessa data mallinnetaan liiketoimintalogiikan mukaisesti, esimerkiksi dimensiomallinnusta hyödyntäen. Tämä mahdollistaa tietojen väliset yhteydet ja keskeisten mittareiden määrittelyn, jotta analyytikot ja päätöksentekijät voivat löytää nopeasti tarvitsemansa tiedot. Joissain organisaatioissa kultakerroksia voidaan luoda useita eri liiketoimintayksiköille, kuten henkilöstöhallinnolle, talousosastolle tai IT:lle, jotta jokainen yksikkö saa juuri sen tiedon, jota se tarvitsee. (Microsoft, 2024b)

Kultakerros on optimoitu suorituskyvyn kannalta, minkä ansiosta kyselyt ja raportit toimivat tehokkaasti. Kun data on hyvin jäsennelty ja liiketoiminnan tarpeiden mukaisesti käsitelty, voidaan siitä luoda selkeitä ja helposti hyödynnettäviä näkymiä, jotka tukevat organisaation strategista päätöksentekoa ja operatiivista tehokkuutta. (Microsoft, 2024b)

2.3 Dataintegraatio

Datan integrointi on kehittynyt merkittävästi vuosikymmenten aikana ja sen rooli on kriittinen nykyaikaisessa datan hallinnassa. Datalähteiden määrän ja monimuotoisuuden kasvaessa integraatiosta on tullut yhä monimutkaisempaa. Datan yhdistämisen haasteet eivät rajoitu pelkästään teknisiin ratkaisuihin, vaan ne liittyvät myös datan laatuun, synkronointiin ja yhteensopivuuteen eri järjestelmien välillä. (Pierre-Yves BONNEFOY ym., 2024)

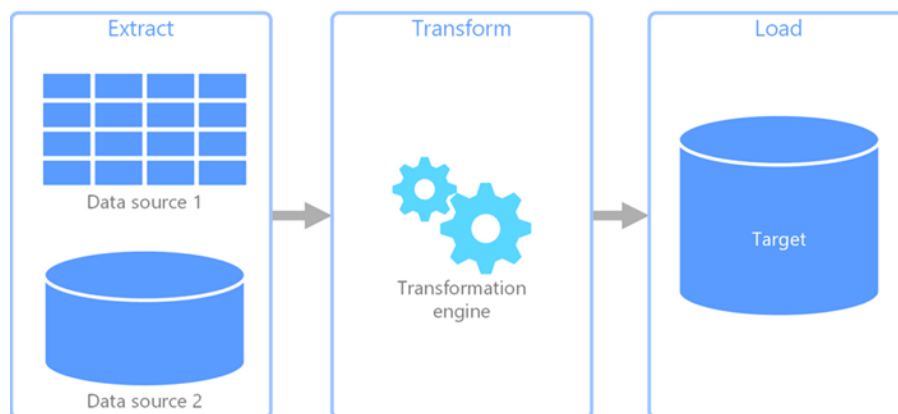
Pilvipalveluiden ja hajautettujen järjestelmien yleistyessä integraation merkitys on kasvanut entisestään. Organisaatiot eivät enää käsittele ainoastaan omia sisäisiä tietokantojaan, vaan ne hyödyntävät myös ulkoisia tietolähteitä, kuten pilvipohjaisia analytiikka-alustoja, IoT-laitteita ja kolmansien osapuolten rajapintoja. Tämä edellyttää tehokkaita mekanismeja tiedon siirtämiseen, muokkaamiseen ja tallentamiseen. (Pierre-Yves BONNEFOY ym., 2024)

2.3.1 ETL-prosessi

ETL eli extract, transform ja load on kolmivaiheinen prosessi datan siirtoon, muokkaamiseen ja lataamiseen. ETL on yleinen prosessi organisaatioissa, kun tarkoituksena on yhdistää tietoja useista eri järjestelmästä yhdeksi tietokannaksi data-altaaseen tai tietovarastoon. Organisaatiot ovat hyödyntäneet ETL-menetelmää jo vuosikymmeniä, mutta viime aikoina tietolähteet ja kohdejärjestelmät ovat siirtyneet pilveen. Lisäksi perinteisten eräajona suoritettavien (batch) ETL-prosessien rinnalle on tullut reaaliaikaisia (streaming) ETL-putkia, joissa data käsitellään jatkuvana virtana. Monet yritykset hyödyntävät jatkuvaa datan käsittelyä yhdistettynä eräajoihin, esimerkiksi historiadatan täydentämiseksi tai tietojen uudelleenprosessointiin. (Google Cloud, n.d.c)

Kuvassa 4 esitellään ETL-prosessin vaiheet. Ensimmäisessä vaiheessa data haetaan (extract) eri lähteistä, kuten vanhoista tietokannoista, on-premise järjestelmistä tai verkkolähteistä. Data jää tämän jälkeen odottamaan jatkokäsittelyä staging-alueelle. Sen jälkeen data muokataan (transform) ja puhdistetaan haluttuun muotoon, jotta se voidaan tallentaa sopivaan järjestelmään, kuten esimerkiksi data-altaaseen tai tietovarastoon. Vaiheeseen kuuluu tyypillisesti eri datalähteiden yhdistäminen, päällekkäisten, virheellisten tai puutteellisten tietojen poistaminen sekä datan standardointi ja normalisointi. Viimeisessä lataus vaiheessa (loading) data tallennetaan kohdejärjestelmään, kuten tietokantaan tai data-altaaseen. Tämän jälkeen dataa voidaan hyödyntää raportoinnissa ja analytiikassa. (Google Cloud, n.d.c)

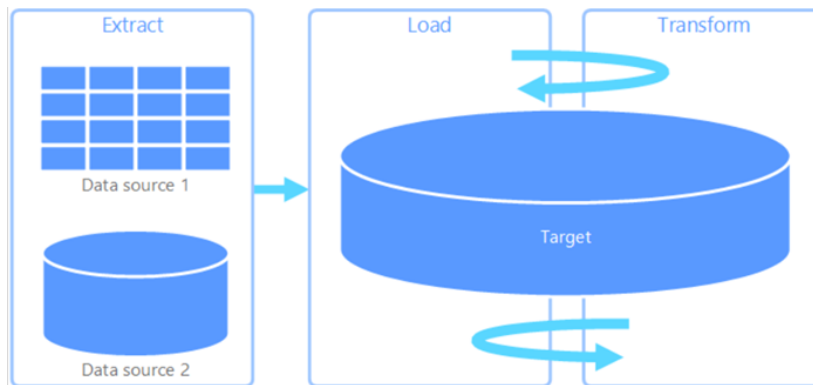
Kuva 4. ETL-prosessin vaiheet (Microsoft, n.d.a)



2.3.2 ELT-prosessi

ELT-prosessissa data siirretään kohdejärjestelmään heti sen hakemisen jälkeen eri lähteistä ja varsinainen muokkaaminen tapahtuu vasta lopuksi (kuva 5.). Tämä lähestymistapa yksinkertaistaa arkkitehtuuria, koska se hyödyntää kohdejärjestelmän laskentatehoa eikä vaadi erillistä muuntamismoduulia. Kohdejärjestelmän on oltava tarpeeksi tehokas ELT-prosessille, jotta sitä voidaan käyttää datan muokkaamista varten. (Microsoft, n.d.a)

Kuva 5. ELT-prosessin vaiheet (Microsoft, n.d.a)



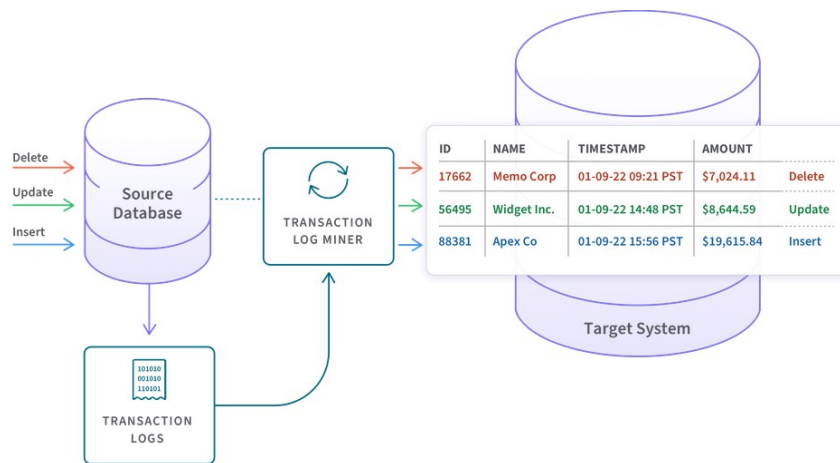
ELT:tä hyödynnetään usein big data -ratkaisuihin, joissa data tallennetaan ensin skaalautuvaan varastoon. Sen jälkeen datan muunnokset suoritetaan suoraan kohdejärjestelmässä. Tämä poistaa tarpeen kopioida dataa eri järjestelmien välillä, mikä nopeuttaa käsittelyä ja vähentää tallennuskustannuksia. Lisäksi ELT tukee joustavampia tallennusratkaisuja, kuten Parquet-muotoa, joka optimoi datan hakemisen ja indeksoinnin. (Microsoft, n.d.a)

2.3.3 CDC

CDC tulee sanoista change data capture ja sillä tarkoitetaan prosessia, joka tunnistaa ja tallentaa tietokantaan tehdyt muutokset reaaliajassa tai lähes reaaliajassa. Sen avulla voidaan siirtää vain muuttunut data analytiikka- tai varastointijärjestelmiin ilman, että koko tietokantaa tarvitsee kopioida uudelleen. Tämä tekee datansiirrosta nopeampaa ja vähentää tuotantojärjestelmien kuormitusta. Pilvipalveluihin siirryttäessä se alentaa myös tiedonsiirtokuluja. CDC tukee erilaisia kohdejärjestelmiä, kuten perinteisiä tietovarastoja, data-altaita sekä reaaliaikaisia analytiikkajärjestelmiä. (Petrie ym., 2018)

CDC voi käyttää eri tekniikoita, kuten tietokantakyselyitä, trigger-pohjaisia ratkaisuja tai lokinlukijoita (kuva 6.), jotka havaitsevat muutokset suoraan tietokannan lokitiedoista. Moderneissa tietoinfrastruktuureissa CDC on keskeinen osa dataputkia, joissa data siirtyy vaiheittain raakadatasta analytiikkavalmiiseen muotoon. CDC:n avulla organisaatiot voivat varmistaa, että analytiikkajärjestelmät ja liiketoimintapäätökset perustuvat aina ajankohtaiseen ja tarkkaan tietoon ilman liian suurta kuormitusta tietokannoille. (Petrie ym., 2018)

Kuva 6. Lokipohjainen CDC-menetelmä (Qlik, n.d.b)



Kuvassa 6 on kuvattu lokipohjainen CDC-menetelmä. Se on tehokkain menetelmä käyttää CDC:tä tietokannan muutosten hallintaan, koska se kuormittaa lähdejärjestelmää vähiten. Tässä lähestymistavassa käytetään tietokannan transaktiologeja, jotka rekisteröivät automaattisesti kaikki lisäykset, päivitykset ja poistot lokitiedostoon ilman, että tämä vaikuttaa lähdejärjestelmän suorituskykyyn. Lokipohjainen menetelmä mahdollistaa tietokannan muutosten tehokkaan ja lähes reaaliaikaisen tunnistamisen ja tallentamisen, mikä parantaa datan ajankohtaisuutta analytiikka- tai varastointijärjestelmissä ja vähentää ylläpitokustannuksia. (Qlik, n.d.b)

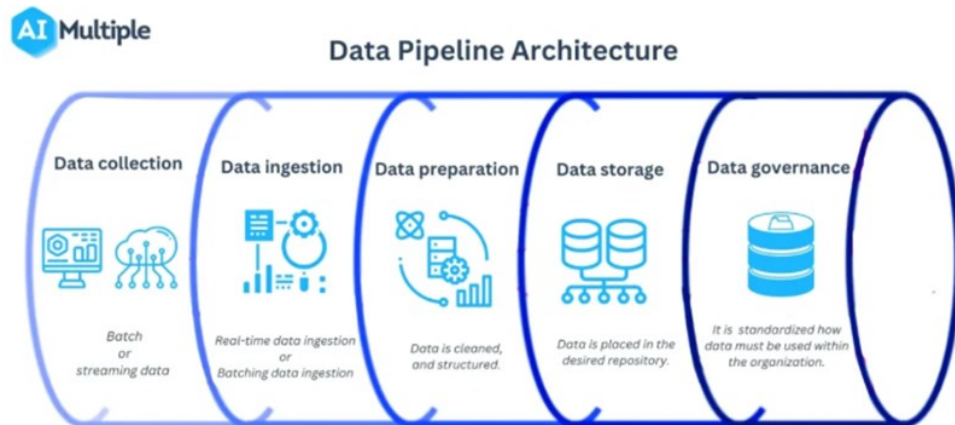
2.3.4 Dataputket

Dataputket keräävät, muokkaavat ja siirtävät dataa analysointia varten esimerkiksi tietovarastoon tai data-altaaseen (kuva 7.). Tämä prosessi varmistaa, että data on yhteensopiva kohdejärjestelmän kanssa, joka voi olla esimerkiksi relaatiotietokanta. Datan käsittelyvaiheessa tietoa suodatetaan, suojataan ja aggregoidaan, jotta se on käyttökelpoista eri tarkoituksiin. (Cole Stryker, 2024)

Yrityksille ja organisaatioille dataputket ovat välttämättömiä, koska ne hallinnoivat suuria tietomääriä eri lähteistä. Raakadata ei kuitenkaan usein ole suoraan käyttökelpoista, vaan se täytyy puhdistaa, yhdistää ja muokata ennen analysointia. Nykyään monet dataputket ovat pilvipohjaisia, mikä mahdollistaa skaalautuvuuden ja tehokkaan hallinnan. Pilvipalvelut

tarjoavat myös valmiita työkaluja datan keräämiseen, käsittelyyn ja varastointiin, mikä helpottaa organisaatioiden tiedonhallintaa ja analytiikkaa. (Cole Stryker, 2024)

Kuva 7. Dataputki (Gulbahar Karasatas, 2024)



Dataputken toiminta koostuu tyypillisesti viidestä päävaiheesta. Vaiheet ovat esitetty kuvassa 7. Ensimmäinen vaihe on datan kerääminen esimerkiksi tietokannoista tai reaaliajasta. Toinen vaihe käsittelee datan syöttämisen, jossa data haetaan eräajoina (batch) tai suoratoistona (streaming). Kolmannessa vaiheessa data puhdistetaan ja valmistellaan sellaiseen muotoon, että sitä voidaan käyttää. Sen jälkeen neljännessä vaiheessa data tallennetaan kohdejärjestelmään. Viimeinen vaihe on datan hallintaan liittyvä vaihe, jossa määritellään, että organisaation eri osastot käyttävät sitä yhtenäisten sääntöjen ja standardien mukaisesti. (Gulbahar Karasatas, 2024)

2.3.5 Datan orkestrointi

Datan orkestrointi on prosessi, jossa eri lähteistä tuleva data kerätään, muunnetaan ja synkronoidaan automaattisesti analyysia varten. Se automatisoi ETL-vaiheet poistamalla tarpeen manuaalisille skripteille ja yhdistämällä tietovarastot ohjelmistojen avulla. Tämä varmistaa, että data on reaaliajassa käytettävissä, tarkkaa ja valmiina analysoitavaksi. Datan orkestrointi ratkaisee keskeisiä haasteita, kuten datasiiilot, datavirtojen pullonkaulat ja tehottomat työnkulut, mikä nopeuttaa liiketoiminnan päätöksentekoa ja parantaa datan hallintaa. (Atlan, 2024)

Yritykset hyödyntävät datan orkestrointia vähentääkseen manuaalista työtä ja virheitä, parantaakseen tietoturvaa ja varmistaakseen datan reaaliaikaisuuden. Automaatio nopeuttaa tietojen yhdistämistä, muuntamista ja lataamista, mahdollistaen nopeammat oivallukset ja skaalautuvuuden. Esimerkiksi yrityksen taloustietojen automaattinen ETL-prosessi voi poistaa päällekkäiset maksut, varmistaa datan puhtauden ja toimittaa sen analytiikkatyökaluille ilman viiveitä. (Atlan, 2024)

2.4 Data-analytiikka

Analytiikka on yleiskäsite ja siitä on monta eri tulkintaa. Yleisesti analytiikalla voidaan viitata data-analytiikkaan tai liiketoiminta-analytiikkaan. Liiketoimintatiedon hallinnan (BI) prosessin näkökulmasta katsottuna analytiikka keskittyy tuottamaan oivalluksia datan ja tiedon analysoinnin ja tutkimisen avulla. Se ei sisällä raportointia tai tiedon louhintaa, vaan on faktapohjaista ja määrällistä analyysiä, joka auttaa vastaamaan liiketoiminnan kysymyksiin ja tekemään ennusteita. Deskriptiivinen -, ennakoiva -ja ohjaava analytiikka ovat sen kolme päätyyppiä. (Kumar V K, 2021, ss. 38–39)

2.4.1 Prosessin keskeiset vaiheet

Data-analyysi on moniosainen prosessi ja työkalu, jota käyttämällä dataa tarkastellaan, puhdistetaan ja muokataan sellaiseen muotoon, jossa sitä voidaan käyttää päätöksenteon tukena ja hyödyllisten tietojen löytämisenä. Data-analyysiprosessi koostuu kuudesta keskeisestä vaiheesta. Ensimmäisessä vaiheessa määritellään tavoitteet ja asetetaan selkeät kysymykset, jotka ohjaavat analyysia. Tämän jälkeen siirrytään datan keräämiseen eri menetelmillä, esimerkiksi kyselyiden ja tietokantojen avulla. Kolmannessa vaiheessa puhdistetaan kerätty data tarkistamalla se virheiden varalta ja korjaamalla mahdolliset epätarkkuudet. (Crabtree & Nehme, 2024)

Neljännessä vaiheessa aloitetaan varsinaisen analyysin tekeminen, jossa sovelletaan tilastollisia menetelmiä kaavojen ja trendien löytämiseksi datasta. Analyysin tulokset tulkitaan ja esitetään visuaalisesti kaavioiden tai grafiikoiden avulla. Lopuksi löydökset kootaan tarinamuotoon, mikä auttaa viestimään tulokset selkeästi ja ymmärrettävästi. Tämä koko prosessi mahdollistaa organisaatioiden tekemään tietoon perustuvia päätöksiä ja parantamaan toimintaansa. (Crabtree & Nehme, 2024)

Data-analyysin tekemiseen on useita erilaisia menetelmiä. Ennakoivat analytiikkamenetelmät hyödyntävät historiallista ja nykyistä dataa tulevien tapahtumien ennustamiseen. Näiden menetelmien avulla voidaan analysoida suuria tietomääriä ja käyttää tilastollisia ja koneoppimismalleja ennustusten laatimiseen. Ennakoiva analytiikka tarjoaa merkittäviä etuja, kuten tuotannon optimointia, kysynnän ennustamista ja asiakaspalvelun parantamista. Esimerkiksi teollisuudessa se voi auttaa ennakoimaan koneiden huoltotarpeita, mikä vähentää odottamattomia tuotantokatkoksia. Vähittäiskaupassa analytiikka voi ennustaa suosittuja tuotteita sesonkiaikoina, parantaen varastonhallintaa ja asiakastytyvyyttä. (Snowflake, n.d.)

Asiakaskäyttäytymisen ymmärtämisessä voidaan käyttää klusterianalyysia menetelmänä. Klusterianalyysi ryhmittelee samankaltaisia objekteja siten, että samassa ryhmässä olevat objektit ovat toisiaan lähempänä kuin eri ryhmiin kuuluvat. Sentimenttianalyysi käyttää menetelmänään tekstianalyysia ja luonnollista kielenkäsittelyä, jonka avulla se pyrkii tunnistamaan subjektiivista tietoa esimerkiksi asiakaspalautteesta. (Crabtree & Nehme, 2024)

Datan analysointiin on käytettävissä monia työkaluja, kuten Python, R, SQL, Power BI, Tableau ja Excel. Python ja R ovat ohjelmointikieliä, jotka tarjoavat laajan valikoiman kirjastoa datan käsittelyyn ja tilastolliseen analyysiin. SQL on olennainen työkalu relaatiotietokantojen hallintaan. Power BI ja Tableau puolestaan tarjoavat interaktiivisia visualisointeja ja raportteja, jotka auttavat muuntamaan raakadataa merkitykselliseksi näkemyksiksi. Excel on laajalti käytetty työkalu, joka soveltuu hyvin niin yksinkertaisiin kuin monimutkaisiin analyysitehtäviin. (Crabtree & Nehme, 2024)

2.4.2 Pilvianalytiikka

Pilvianalytiikka on datan analysoinnin muoto, joka siirtää datankäsittelyn ja tallennuksen pilvipalveluihin. Tämä mahdollistaa suuren ja monimutkaisen liiketoimintadatan analysoinnin tehokkaasti hyödyntämällä algoritmeja ja pilviteknologioita. Pilvianalytiikan avulla organisaatiot voivat tunnistaa trendejä, tehdä ennusteita ja saada liiketoimintatietoa ilman, että niiden tarvitsee huolehtia omasta infrastruktuuristaan. Pilvitoimittaja huolehtii palvelun ylläpidosta ja skaalautuvuudesta. (Google Cloud, n.d.b)

Pilvianalytiikka tarjoaa samanlaisia ominaisuuksia kuin perinteiset paikalliset on-premise datan analytiikkaratkaisut, mutta se mahdollistaa joustavamman ja kustannustehokkaamman lähestymistavan. Organisaatiot voivat käyttää pilvianalytiikkaa monenlaisiin käyttötarkoituksiin, esimerkiksi tietovarastointiin ja reaaliaikaiseen analytiikkaan. Tämän ansiosta data voidaan yhdistää useista lähteistä ja analysoida nopeasti, mikä auttaa liiketoimintapäätöksissä ja strategian kehittämisessä. (Google Cloud, n.d.b)

Pilvianalytiikka jakautuu kolmeen päätyyppiin, joita ovat julkinen pilvi, yksityinen pilvi ja hybridipilvi. Julkinen pilvi tarjoaa analytiikkapalveluja useille organisaatioille samassa ympäristössä, jolloin käyttäjät jakavat resursseja, kuten virtuaalikoneita ja tietovarastoja, pitäen kuitenkin tietonsa eristyksissä. Tämä malli on kustannustehokas, mutta sen tietoturva ei ole yhtä vahva kuin yksityisessä pilvessä. (Google Cloud, n.d.b)

Yksityinen pilvi on varattu vain yhdelle organisaatiolle ja voi sijaita joko sen omassa datakeskuksessa tai kolmannen osapuolen palvelimilla. Se tarjoaa paremman tietoturvan, mutta sen ylläpito ja skaalautuvuus ovat kalliimpia. Hybridipilvi yhdistää julkisen ja yksityisen pilven hyödyt, jolloin organisaatio voi käsitellä ei-herkkää dataa julkisessa pilvessä ja säilyttää arkaluonteisen datan yksityisessä pilvessä. Lisäksi monet organisaatiot hyödyntävät hybridipilvianalytiikkaa, jolloin dataa tallennetaan useissa julkisissa pilvissä, mikä mahdollistaa tietojen analysoinnin ilman siirtämistä tai kopioimista. (Google Cloud, n.d.b)

2.5 Raportointi ja päätöksenteko

Raportoinnilla tarkoitetaan prosessia, jossa liiketoimintatiedosta kehitetään, suunnitellaan ja toimitetaan raportteja päätöksentekijöille. Raportit voivat sisältää esimerkiksi KPI-mittareita, kaavioita ja taulukoita, jotka kuvaavat liiketoiminnan suorituskykyä eri näkökulmista. Liiketoimintatiedon raportointia voidaan tuottaa interaktiivisina dashboardeilla, koontinäytöillä tai staattisilla raporteilla. Raportointia käytetään liiketoiminnan suorituskyvyn hallinnassa, päätöksenteossa ja mahdollisuuksien ja ongelmien tunnistamisessa. (Kumar V K, 2021, ss. 36–38)

KPI-mittarilla (Key Performance Indicator) tarkoitetaan mittareita, jotka osoittavat liiketoiminnan suorituskykyä. KPI-mittareiden mittaaminen on keskeistä liiketoiminnan

kannattavan kasvun ja tavoitteiden saavuttamisen seuraamisessa. KPI-mittareita on yrityksessä yleensä vain muutama, koska liian monen mittarin seuraaminen voi olla haastavaa. KPI-mittareita voi olla monilla eri tasoilla yrityksessä, esimerkiksi eri osastoilla ja tiimeissä. (Kumar V K, 2021, s. 31–32)

Liiketoimintatiedon hallinnan prosessin lopuksi tehdään päätöksiä, jotka perustuvat datasta saatuun tietoon. Tietoon perustuva päätöksenteko on yleistymässä, mutta silti monet päätökset tehdään yrityksissä edelleen vaiston varassa. Organisaatioiden kasvaessa päätöksentekovastuu siirtyy alemmille tasoille, joten on tärkeää, että työntekijät tekevät oikeita päätöksiä liiketoiminnan kehittämiseksi ja tällöin BI:tä voidaan käyttää tukena päätöksenteossa. (Kumar V K, 2021, ss. 10, 38)

3 Vertailussa olevat järjestelmät

Tässä luvussa käsitellään järjestelmiä, jotka toimivat vertailun kohteena sekä esitellään lyhyesti Microsoft Azure alustapalveluna. Vertailtavina järjestelminä ovat Microsoft Fabric ja Power BI. Luvussa tarkastellaan näiden järjestelmien keskeisiä ominaisuuksia ja luku toimii teoreettisena pohjana myöhemmälle vertailulle.

3.1 Microsoft Azure alustapalveluna

Microsoft Azure on pilvipalvelualusta (PaaS), joka tarjoaa laajan valikoiman työkaluja ja palveluja yrityksille eri tarpeisiin. PaaS-palveluiden avulla voidaan kehittää, ajaa ja hallita sovelluksia ilman, että tarvitsee huolehtia taustalla olevasta infrastruktuurista. PaaS tyyppisesti käytönotetaan, kun halutaan kehittää pilvipohjaisia sovelluksia, esimerkiksi verkkosivustoja tai analytiikkaratkaisuja. Azure hoitaa automaattisesti skaalautuvuuden, tietoturvan ja ylläpidon, jolloin kehittäjät voivat keskittyä sovellusten rakentamiseen ja innovointiin. (Microsoft, n.d.c)

Microsoft Azuren etuja ovat muun muassa korkea saatavuus. Sen infrastruktuuri on rakennettu varmistamaan palvelujen jatkuvuus eri alueilla ympäri maailmaa. Lisäksi Azure noudattaa tiukkoja turvallisuusstandardeja ja tarjoaa useita sertifikaatteja, jotka suojaavat käyttäjien tietoja ja varmistavat vaatimustenmukaisuuden. Helppo integrointi muihin Microsoftin tuotteisiin ja kolmannen osapuolen sovelluksiin tekee Azuresta joustavan vaihtoehdon monenlaisille liiketoimintaympäristöille. (Soh ym., 2020)

Microsoft Azuren käytössä olevat haittapuolet kohdistuvat enemmän suurille yrityksille. Joustava hinnoittelumalli sopii pienille ja keskisuurille yrityksille, mutta suurilla organisaatioilla voi olla huomattavia kustannuksia johtuen suuresta tietojenkäsittelyn ja laskentatehon tarpeesta. Tällaisissa tapauksissa omien palvelimien ylläpito voi olla taloudellisesti järkevämpää kuin pilvipalveluun siirtyminen. Azuren palveluiden nopeus voi vaihdella käyttäjän maantieteellisen sijainnin mukaan. Yhdysvalloissa ja Euroopassa, palvelu toimii tehokkaasti, kauempana sijaitsevilla alueilla voi ilmetä viiveongelmia. (Hystax, 2024)

Pilvipohjainen infrastruktuuri vaatii myös asiantuntevaa hallintaa, mikä voi tuoda lisähaasteita yrityksille. Azure vähentää fyysisen laitteiston ja sen ylläpidon tarvetta, mutta pilvipalveluiden hallinta edellyttää osaavia IT-ammattilaisia. Suurten organisaatioiden on investoitava Azure-asiantuntijoihin varmistaakseen sujuvan käyttöönoton, tietoturvan ja järjestelmien ylläpidon. Ilman riittävää asiantuntemusta palveluiden hallinta voi osoittautua monimutkaiseksi ja johtaa tehottomaan resurssien käyttöön. (Hystax, 2024)

3.2 Microsoft Fabric

Microsoft Fabric on pilvipohjainen analytiikka- ja datanhallintapalvelu, joka perustuu SaaS -arkkitehtuuriin. Alustan julkaisu tapahtui vuonna 2023 kun Microsoft esitteli sen Build-tapahtumassaan. Sen tarkoituksena on tarjota yrityksille entistä yhtenäisempi ja tehokkaampi ratkaisu datan käsittelyyn, analysointiin ja hyödyntämiseen. (Ghosh, 2024)

Microsoft Fabric soveltuu eri aloille, kuten terveydenhuoltoon, vähittäiskauppaan ja teollisuuteen. Se tukee reaaliaikaista analytiikkaa, skaalautuvaa tallennusta ja edistyneitä koneoppimisominaisuuksia, mahdollistaen mm. petosten tunnistuksen ja IoT-datan analysoinnin. Alusta tukee monia datalähteitä ja datatyyppejä. Se sisältää vahvan tietoturvan ja integroituu sujuvasti Microsoftin työkaluihin. SaaS-palveluna se on helposti käyttöönotettavissa ja tarjoaa joustavia hinnoittelumalleja, esimerkiksi käytön mukaan maksaminen. (Borra, 2024)

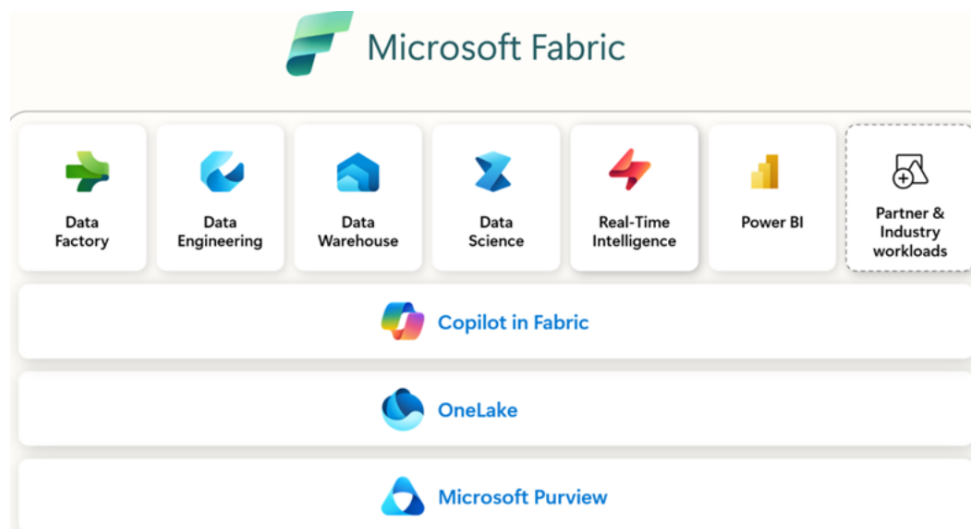
3.2.1 Alustan keskeiset komponentit

Alusta yhdistää uusia ja olemassa olevia komponentteja, joita organisaatiot tarvitsevat data ja analytiikka tarpeisiin. Microsoft Fabric alusta yhdistää Power BI:n, Azure Synapse Analyticsin ja Azure Data Factoryn kaltaiset palvelut, jolloin yritykset voivat käsitellä, analysoida sekä jakaa dataa keskitetysti ilman erillisiä järjestelmiä. (Microsoft, 2025d)

Alustan keskeisiä komponentteja kuvassa 8. Data Factory mahdollistaa datan integroimisen sekä käsittelyn ja Data Engineering suurien datamäärien tallentamisen, käsittelyn, keräämisen ja valmisteluun analysoitavaksi. Data Warehouse toimii keskitettynä tietovarastona, joka mahdollistaa tietojen säilyttämisen ja on tiiviisti integroitu Power BI:hin, mikä helpottaa analysointia ja raportointia. Data Science komponentin avulla voidaan

käyttää ja kehittää koneoppimista. Real-Time Intelligence mahdollistaa reaaliaikaisen datan analysoinnin ja raportoinnin. Power BI tarjoaa työkaluja liiketoimintatiedon visualisointiin ja analysointiin. Lisäksi alusta tarjoaa erilaisia räätälöityjä ratkaisuja yritysten omiin tarpeisiin. (Microsoft, 2025d)

Kuva 8. Microsoft Fabricin keskeiset komponentit (Microsoft, 2025d)



Copilot on yksi Fabricin keskeisistä komponenteista. Copilot tarjoaa tekoälypohjaisia ominaisuuksia käyttäjien avuksi ja työn sujuvoittamiseksi. OneLake toimii keskitettynä tietovarastona koko organisaatiolle. Microsoft Purview on integroitu Fabriciin ja vastaa alustan hallintakäytänteistä, kuten käyttöoikeuksien hallinnasta ja datan hallinnasta. Purview varmistaa, että data säilytetään organisaatiossa turvallisesti. (Microsoft, 2025d)

3.2.2 Datan varastointi

Microsoft Fabricin keskeinen ominaisuus on OneLake. OneLake on yhtenäinen ja keskitetty data-allas, joka on rakennettu Azure Data Lake Storage Gen2:n päälle. Se tarjoaa organisaatioille yhden keskitetyn tallennustilan, joka on helpottaa datan löytämistä, jakamista ja hallintaa. Jokaiseen Microsoft Fabricin asiakkuuteen kuuluu vain yksi OneLake ja sen sisällä on mahdollisuus luoda useita eri työtiloja. OneLaken hierarkkinen rakenne yksinkertaistaa organisaation laajuista hallintaa ja se järjestää datan säilöihin, mikä helpottaa sen käsittelyä. (Microsoft, 2025d)

OneLaken hyötyjä on se, että se poistaa datan käsittelystä syntyvät datasiilot, joita kehittäjät usein luovat hallitessaan omia eristettyjä tallennustilejään. OneLake tarjoaa kaikille kehittäjille yhden yhtenäisen tallennusjärjestelmän, mikä helpottaa datan löytämistä, jakamista sekä tietoturva- ja hallintakäytäntöjen yhdenmukaista noudattamista. (Microsoft, 2025d)

OneLaken keskeinen etu on sen "One Copy" -periaate, joka vähentää tarpeettomia datan kopioita ja mahdollistaa saman datan käytön eri analytiikkatyökaluilla ilman monimutkaisia siirtoja tai muunnoksia. OneLake tukee avoimia tallennusstandardeja, kuten Delta Parquet-muotoa, mikä mahdollistaa suorituskykyiset kyselyt ja yhteensopivuuden useiden analytiikkamootorien kanssa. Lisäksi OneLaken ainutlaatuinen oikopolkuominaisuus (Shortcuts) mahdollistaa datan virtuaalisen yhdistämisen eri tallennuspaikkojen välillä ilman erillistä kopiointia. Oikopolut voivat viitata muihin Fabric-työtiloihin, Azure Data Lake Storage Gen2 -säilöihin, mikä helpottaa hajautetun datan hallintaa. (Kishan, 2024)

3.2.3 Datan hallinta ja käsittely

Microsoft Fabric tarjoaa kattavat datankäsittelyominaisuudet koko analytiikkaprosessille. Se hyödyntää Synapse Analyticsin ominaisuuksia, jotka mahdollistavat sekä erä- että reaaliaikaisen datankäsittelyn. Apache Spark -tekniikan avulla Microsoft Fabric pystyy käsittelemään suuria tietomassoja tehokkaasti ja nopeasti. Se tarjoaa kehittyneet ETL- ja ELT- ominaisuudet Data Factoryn kautta, jonka yksinkertaistaa datan integrointia ja muuntamista. Tämän lisäksi Microsoft Fabric mahdollistaa koneoppimisen ja tekoälymallien integroinnin, mikä avaa uusia mahdollisuuksia datan analysoimiseen ja hyödyntämiseen liiketoiminnassa. (Kanerika, 2025)

Medallion-arkkitehtuurimalli on Microsoft Fabricin suositeltu malli datan käsittelyssä. Arkkitehtuuri jakaa datan kolmeen kerrokseen, joita ovat pronssi (raakadata), hopea (puhdistettu data) ja kulta (valmiiksi analysoitu data). Arkkitehtuurin rakenne parantaa tiedon hallintaa, laatua ja analysoitavuutta. Pronssikerros toimii laajana historiallisten tietojen varastona, hopeakerroksessa dataa siivotaan ja muokataan käyttökelpoisemmaksi ja kultakerros tarjoaa valmiita liiketoimintatason näkemyksiä. (Harrington, 2024)

Medallion-arkkitehtuuri auttaa parantamaan datan skaalautuvuutta ja suorituskykyä. Pronssikerroksessa raakadatamäärät voivat olla suuria, mutta kultakerrokseen päästessä

data on jo puhdistettu ja optimoitu liiketoimintatiedon tarpeisiin. Tämä auttaa tasapainottamaan suorituskykyä ja mahdollistaa nopeammat raporttien ja koontinäyttöjen vastausajat. (Harrington, 2024)

Kolmikerroksinen arkkitehtuurimalli helpottaa myös eri käyttäjäryhmien työskentelyä. Datansinöörit voivat keskittyä tiedon keräämiseen ja muuntamiseen pronssi- ja hopeakerroksissa ja liiketoiminnan päätöksentekijät hyödyntävät esimerkiksi KPI-mittareita, valmiita raportteja ja koontinäyttöjä kultakerroksessa. (Harrington, 2024)

3.2.4 Tekoäly

Microsoft Fabric on tekoälypohjainen analytiikka-alusta, joka helpottaa yritysten datan hallintaa ja hyödyntämistä. Se hyödyntää tekoälyä datan syväanalyysiin, ennusteisiin ja luonnollisen kielen prosessointiin (NLP). Käyttäjät voivat esittää kysymyksiä ja saada ymmärrettäviä vastauksia ilman teknistä osaamista. Integraatio Azuren kanssa mahdollistaa Azure Machine Learningin koneoppimismallien kehittämisen ja seurannan. (Savani, 2024)

Microsoft Fabric käyttää Copilot tekoälyominaisuutta. Copilot on automaattisesti päällä, mutta tarvittaessa järjestelmänvalvojat voivat poistaa sen käytöstä. Organisaation on oltava F64-kapasiteetissa tuetulla alueella, jotta sitä on mahdollista käyttää. Copilotin ominaisuudet Fabricissa noudattavat vastuullisen tekoälyn standardeja ja ne on arvioitu monialaisissa tiimeissä mahdollisten riskien tunnistamiseksi ja hallitsemiseksi. Copilotin toiminnot toimivat parhaiten englannin kielellä. (Microsoft, 2025b)

Copilot tarjoaa monipuolisia työkaluja eri data-alan ammattilaisille. Se automatisoi koodauksen, antaa älykkäitä koodiehdotuksia ja nopeuttaa dataputkien sekä analyysimallien rakentamista. Lisäksi Copilot auttaa Data Factoryssa muokkaamaan dataa sujuvammin luomalla automaattisesti koodia ja selittämällä monimutkaisia prosesseja. (Microsoft, 2025c)

Copilot tarjoaa tekoälyratkaisuja tietovarastointiin, raportointiin ja reaaliaikaiseen analytiikkaan. Se muuntaa luonnollisen kielen SQL:ksi, automatisoi Power BI -raporttien luomista ja analysointia sekä käsittelee suuria tietomääriä reaaliajassa. Copilot hyödyntää

Microsoft Fabricia, Azure OpenAI Servicea ja ja Azure AI Speechia keskusteludatan analysoimiseen ja keskeisten oivallusten löytämiseen. (Microsoft, 2025c)

3.2.5 Operatiivinen data

Teollisuusympäristössä syntyy jatkuvasti valtavia tietomääriä erilaisista lähteistä, kuten automaatiojärjestelmistä ja IoT-laitteista. Microsoft Fabric pystyy yhdistämään dataa eri lähteistä teollisuudessa, kuten MES-tuotannonohjausjärjestelmästä ja ERP-toiminnanohjausjärjestelmästä sekä muista operatiivisen teknologian järjestelmistä. Alustaan kuuluva Real-Time Intelligence (RTI) on erityisesti suunniteltu käsittelemään suuria määriä suoratoistodataa. RTI:n avulla nämä tiedot voidaan kerätä, analysoida ja hyödyntää reaaliaikaisesti. Esimerkiksi tuotantolinjalla tapahtuvat poikkeamat voidaan havaita ja käsitellä välittömästi, mikä voi vähentää kustannuksia. (Nortal, 2024)

Microsoft Fabric on julkaisemassa uusia ominaisuuksia, jotka parantavat teollisuusdatan hallintaa ja analysointia. Nämä ominaisuudet ovat tällä hetkellä esikatseluvaiheessa ja niiden odotetaan olevan käytettävissä vuonna 2025 syyskuuhun mennessä. (Microsoft, 2025a) Microsoft Fabricin teollisuuteen tarkoitettu tietomalli on suunniteltu parantamaan valmistusprosessien hallintaa ja nopeuttamaan oivallusten saamista. Sen avulla tuotantotiimit voivat tehdä tietoon perustuvia päätöksiä ja parantaa tehokkuutta. Malli perustuu alan standardiin (ISA-95) ja sen tavoitteena on standardoida datanhallintaa ja lisäämään operatiivista tehokkuutta teollisuusympäristössä. Tietomalli on vielä tällä hetkellä esikatseluvaiheessa. (Nortal, 2024)

Toinen esikatseluvaiheessa oleva teollisuuteen tarkoitettu ominaisuus Microsoft Fabricissa on Azure AI:n Factory Operations Agent. Tämä työkalu auttaa teollisuusyrityksiä parantamaan toimintatehokkuutta ja nopeuttamaan ongelmanratkaisua hyödyntämällä tekoälyä. (Microsoft, 2024c) Tehdasympäristössä syntyy paljon monimutkaista dataa ja tämä työkalu auttaa etulinjan työntekijöitä hyödyntämään tätä dataa. Mallit mahdollistavat luonnollisen kielen vuorovaikutuksen omien tietojen kanssa. Se helpottaa erilaisten tehtävien, kuten juurisyyntä analyysia, tiedonhakua, koulutusta, ongelmanratkaisua ja omaisuuden hallintaa. (Nortal, 2024)

3.2.6 Turvallisuus ja hallinta

Microsoft Fabricissa käyttäjien todennus tapahtuu Microsoft Entra ID:n kautta. Ulkopuolisille käyttäjille voidaan myöntää pääsy Fabricin sisältöön vain organisaation suostumuksella. Nämä vieraat lisätään Entra ID -hakemistoon Azure-portaalista ja heitä hallitaan kuten muita käyttäjiä, mutta heillä ei ole pääsyä kaikkiin sisältötyyppeihin. (Ali & Schacht, 2024)

OneLake on integroitu Microsoft Entra ID -käyttäjähallinnan kanssa, mikä takaa vahvan tietoturvan ja roolipohjaisen pääsynhallinnan. Fabricin kaikki analytiikkapalvelut käyttävät OneLakea tallennusratkaisunaan, jonka takia datan hallinta, tietoturva ja jakaminen on yksinkertaisempaa verrattuna perinteisiin erillisiin datavarastoihin. (Kishan, 2024)

Microsoft Fabric tarjoaa tehokkaita työkaluja datan hallintaan ja suojaamiseen, mikä varmistaa, että dataa käsitellään asianmukaisesti. Sen ominaisuuksiin kuuluvat datan tunnistamiseen liittyvät suositukset, metatietojen skannaus ja tietoketjun visualisointi, jotka helpottavat datan hallintaa suurissa ympäristöissä. Domaineja käytetään ryhmittämään työtiloja ja datavarastoja, mikä parantaa datan hallintaa. Purview tarjoaa työkaluja datan luokitteluun sekä tietoturvan parantamista ja käyttäjätoimintojen auditointia. (Ali & Schacht, 2024)

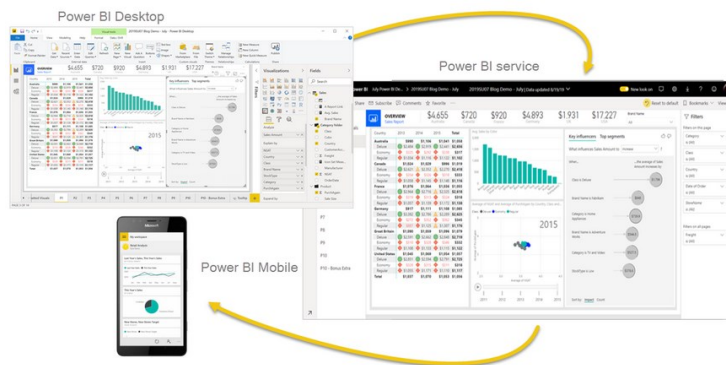
3.3 Power BI

Power BI on Microsoftin kehittämä alusta, joka on suunniteltu liiketoimintatiedon analysointiin ja datan visualisointiin. Sen avulla käyttäjät, pystyvät yhdistämään dataa eri lähteistä, muokata ja analysoida sitä sekä luoda dashboardeja ja interaktiivisia raportteja. Power BI tarjoaa monipuolisia visualisointityökaluja, joiden avulla voidaan tunnistaa trendejä, poikkeamia ja yhteyksiä datasta helposti. Power BI:n käyttö tuo yrityksille merkittäviä etuja, kuten reaaliaikaisen tiedon analysoinnin, paremman päätöksenteon ja tehokkaamman yhteistyön. (Metre ym., 2024)

3.3.1 Keskeiset osat ja lisenssit

Power BI sisältää kolme ympäristöä (kuva 9.). Power BI Desktop on työpöytäsovellus raporttien luomiseen, Power BI Service on verkkopohjainen alusta raporttien jakamiseen ja Power BI Mobile toimii sovelluksena. Power BI tarjoaa myös työkaluja sivutettujen raporttien luomiseen sekä raporttien julkaisemiseen on-premise palvelimille Power BI Report Serverin kautta. Tämä mahdollistaa raporttien hallinnan paikallisesti organisaation palomuurin takana ja käyttäjät voivat tarkastella raportteja eri tavoin, kuten verkkoselaimessa tai mobiililaitteilla. (Microsoft, 2024a)

Kuva 9. Power BI:n osat (Microsoft, 2024a)



Power BI Desktop on ilmainen ja kuuluu kaikkiin Office 365-tilauksiin. Sen käyttäminen edellyttää rekisteröitymisen ja tietolähteiden yhdistämisen, jonka jälkeen on mahdollista luoda raportteja ja visualisointeja. Power BI Service on maksullinen ja sisältää enemmän ominaisuuksia. Sen hallinta tapahtuu Microsoft Azure:n pilvessä. Power BI Servicen käyttämiseksi pitää olla Power BI pro tai Premium lisenssi. Power BI Pro tarjoaa samanlaista toiminnallisuutta ja tallennustilaa kuin Desktop-versio ja lisäominaisuutena mahdollisuuden jakaa sisältöä yksityisesti muiden kanssa tai luoda sovellustyötiloja. Premium lisenssillä saa pääsyn tekoälyn hyödyntämiseen, itsepalveludatan valmistelumahdollisuuksiin ja yksinkertaistettuun tietojen hallintaan. Power BI Mobile mahdollistaa pääsyn kaikkiin Power BI Desktopissa, Power BI Pro/Premiumissa ja Power BI Report Serverissä luotuihin koontinäyttöihin ja raportteihin. (AltexSoft, 2022)

3.3.2 Datan varastointi ja hallinta

Power BI käyttää omaa tietomallin varastointia, mutta se ei tarjoa sisäänrakennettua tallennusmahdollisuutta. Se yhdistyy eri datalähteisiin ja käyttää sarakepohjaista tallennustapaa. Power BI:llä on rajoituksia datakokonaisuuksien kanssa, jotka riippuvat lisenssistä. Suurten tietomäärien varastointiin tarvitaan ulkoista tallennustilaa. Power BI on optimoitu analyttisten tietomallien tallentamiseen. (Kanerika, 2025)

Monista lähteistä tuleva data on yhdistettävä yhteiseen tietomalliin (CDM), ennen kuin dataa voidaan hyödyntää Power BI:ssä. CDM on keskitetty tietomalli, joka toimii yhteisenä paikkana sovellusten ja tietolähteiden väliselle datalle. Se varmistaa tietojen johdonmukaisuuden ja sovellusten yhtenäisyyden, jolloin jokaisen sovelluksen tulee osata käsitellä CDM:ää. CDM sisältää valmiita taulurakenteita, jotka kattavat useimpien liiketoimintaprosessien tarpeet. (Rad, 2019) Power Query, Dataflows ja ETL-työkalut mahdollistavat yhteisen tietomallin rakentamisen. (Keboola, 2023)

Power Query on työkalu, joka auttaa tietojen muuntamisessa ja valmistelussa Power BI desktopissa. Power Query sisältää graafisen käyttöliittymän, jonka avulla käyttäjät voivat yhdistää tietoja eri lähteistä ja soveltaa muunnoksia, kuten suodatusta ja sarakkeiden poistamista. Power Query tukee myös monia datalähteitä ja tarjoaa yli 350 erilaista muunnosmahdollisuutta. Käyttäjät voivat luoda muunnoksia ilman ohjelmointitaitoja ja työkalu on saatavilla eri Microsoft-tuotteissa, kuten Power BI:ssä ja Excelissä. Power Queryn avulla voidaan myös luoda toistettavia prosesseja, mikä helpottaa tietojen päivittämistä ja hallintaa. (Microsoft, 2025e)

Power BI service:ssä toimiva Dataflows mahdollistaa itsenäisemmän tavan datan hallintaan kuin Power Query. Käyttäjät voivat valita tietolähteitä, määrittää sarakkeiden arvotyypit ja otsikot, liittää tauluja yhteen luodakseen skeeman sekä tallentaa dataflown yhteiseen varastoon Azureen. Tämän jälkeen Power BI työpöytä versiossa voidaan valita se tietolähteeksi ja tuoda tiedot työtilaan analysoitavaksi. Haasteena dataflown käyttämisellä on rajoitukset tietolähteiden osalta, joka voi vaikuttaa käyttäjien joustavuuteen ETL-prosessin rakentamisessa. (Keboola, 2023)

Power BI:n datan valmistelussa voidaan käyttää myös ETL tai ELT työkaluja. Näiden avulla käyttäjät voivat automatisoida tiedon keräämisen, muuntamisen ja lataamisen eri vaiheita.

ETL-työkalut mahdollistavat erilaisten tietolähteiden yhdistämisen ja datan puhdistamisen ennen sen siirtämistä tietovarastoihin, kuten SQL-serveriin. ETL-työkalut tarjoavat joustavia vaihtoehtoja tietojen muuntamiseen. ETL etuna on se, että työkalut antavat käyttäjälle omistajuuden datan putkistoista toisin kuin Power BI Service. ETL-työkaluja voi käyttää ilman riippuvuutta Power BI:stä, mikä tarjoaa enemmän joustavuutta ja hallintaa datan käsittelyssä. (Keboola, 2023)

Power BI voi olla hidas käsittelemään dataa ja luomaan raportteja. Suuret tietoaaineistot hidastavat suorituskykyä merkittävästi ja yli 20 000–30 000 rivin käsittely voi aiheuttaa ongelmia. Tämä voi vaikuttaa käyttäjien tehokkuuteen. Suorituskykyongelmaa voi kiertää jakamalla tiedot useisiin kyselyihin ja kategorisoimalla ne esimerkiksi päivämäärien mukaan, mutta tämä vaatii ylimääräistä työtä ja aikaa. Power BI ei myöskään tarjoa sisäänrakennettua työkalua datan puhdistamiseen. Se olettaa, että analysoitava data on jo valmiiksi laadukasta ja puhdistettua. Datat puhdistamista varten tarvitsee käyttää jotain muuta työkalua. (CCS Learning Academy, 2024)

Power BI ei tue muistinsisäistä datankäsittelyä ja tämä hidastaa suurten tietoaaineistojen analysointia ja tekee prosessista aikaa vievän. Monimutkaisten datasettien kanssa työskentely vaatii usein enemmän käsittelyaikaa ja päivittämistä. Power BI:n tietojen päivitys ei aina tapahdu reaaliajassa, koska sen rakenne vaikeuttaa datan hallintaa ja päivitysnopeutta. (CCS Learning Academy, 2024)

3.3.3 Keskeiset ominaisuudet

Power BI on tehokas ja edullinen analytiikkatyökalu, joka tarjoaa hyviä puolia erityisesti pienille ja keskikokoisille yrityksille. Power BI käy hyvin myös yrityksille, joissa on käytössä jokin muu kuin perinteinen tietovarasto. Ohjelman automaattinen tietojen päivitys ja raporttien luonti vähentävät manuaalista työtä, mikä parantaa tehokkuutta ja auttaa yrityksiä pysymään ajan tasalla tärkeistä tiedoista. (Lavanya ym., 2023, ss. 12, 16)

Power BI tarjoaa monipuoliset ja visuaaliset esitystavat, kuten kaaviot, taulukot ja scatter plot -kuvaajat, joiden avulla voidaan tarkastella liiketoiminnan suorituskykyä selkeästi. Se tukee useita tietolähteitä, kuten talous-, asiakas- ja markkinadataa. Power BI:ssa käyttäjät voivat tutkia dataa syvemmin interaktiivisten raporttien avulla ja reagoida nopeasti muuttuviin liiketoiminnan tarpeisiin. (Banerjee ym., 2023)

Power BI ei ole täysin joustava räätälöityjen analyysien suhteen. Monimutkaisten raporttien tai erikoistuneiden analyysien toteuttaminen voi vaatia lisätyökaluja tai koodaustaitoja, kuten DAX- tai Power Query -kieltä, mikä voi olla haaste aloittelijoille. Power BI:n käyttöliittymä on suhteellisen helppokäyttöinen, mutta sen tehokas hyödyntäminen edellyttää perehtymistä ja harjoittelua. (Banerjee ym., 2023)

Power BI:ssa on ominaisuuksia reaaliaikaisen datan hyödyntämiseen ja analysoimiseen. Tässä on kuitenkin lisenssikohtaisia eroja. Power BI Pro lisenssin käyttäjät voivat aikatauluttaa enintään kahdeksan tietojen päivitystä päivässä ja Premium lisenssin käyttäjät voivat tehdä jopa 48 päivitystä päivässä eli 10 minuutin välein. Reaaliaikaisen analyysin tekemiseen on kaksi vaihtoehtoa, joita ovat automaattinen sivupäivitys ja suoratoistodatsetit. Automaattinen sivupäivitys mahdollistaa tietojen päivittämisen jopa sekunnin välein. Se vaatii yhteyden on-premise data gatewayhin sekä raportin julkaisemisen Premium- tai Embedded-kapasiteettiin. Suoratoistodatsetit tarjoavat reaaliaikaista dataa Azuren data streamista tai REST API:sta ja niitä voidaan analysoida välittömästi Power BI Service -palvelussa. Suoratoistodatsetin visulisointitavat ovat rajalliset. (Shapiro, 2024)

KPI-mittarit Power BI:ssa ovat mittareita, jotka auttavat yrityksiä seuraamaan ja arvioimaan liiketoimintatavoitteidensa saavuttamista. Ne visualisoidaan Power BI:ssa erilaisilla grafiikoilla, joita ovat esimerkiksi KPI-visualisointi, mittariviisari tai korttiviisari. KPI-mittarit näyttävät suorituskyvyn, vertailtavat tuloksia asetettuihin tavoitteisiin ja auttavat tunnistamaan trendejä ja mahdollisuuksia liiketoiminnassa. KPI-mittarien reaaliaikainen päivitys Power BI:ssa onnistuu yhdistämällä reaaliaikaisiin tietolähteisiin tai käyttämällä suoraa kysely yhteyttä. (Bothma, 2024)

3.3.4 Tekoäly

Power BI sisältää monia tekoälyominaisuuksia, mutta ne ovat käytössä vain Power BI Premium lisenssin käyttäjille. AI Insights käyttää ennakkoon koulutettuja koneoppimismalleja datan valmisteluun. Premium lisenssin avulla voidaan myös käyttää Text Analytics sekä Vision toimintoja, joita voi suorittaa Power BI Service ympäristössä. Näihin kuuluu esimerkiksi avainfaasien poiminta tekstistä, sentimenttianalyysi, kielen tunnistus ja kuvien tunnistet. Toiminnot käyttävät Azure Cognitive Services algoritmeja. (Microsoft, 2025f)

Power BI Premium -lisenssi mahdollistaa myös syvällisemmän analytiikan hyödyntämisen yhdistämällä Azure Machine Learningin tarjoamat ennustemallit Power BI:hin. Tämän avulla käyttäjät voivat luoda tarkkoja ennusteita ja malleja historiallisten tietojen perusteella, mikä auttaa liiketoiminnan päätöksenteossa. Power BI:n tekoälyominaisuudet, kuten poikkeavuuksien havaitseminen ja luonnollisen kielen tunnistus, tekevät datan analysoinnista entistä helpompaa. Toiminnot mahdollistavat käyttäjille poikkeavuuksien tunnistamisen ja tietojen kysymisen luonnollisella kielellä ilman, että käyttäjältä vaaditaan syvällistä teknistä osaamista. (Marina Pigol, 2024)

3.3.5 Turvallisuus ja hallinta

Power BI:n turvallisuus perustuu Microsoft Entra ID:hen, joka mahdollistaa käyttäjien todennuksen sähköpostiosoitteen ja salasanan avulla. Tämän prosessin kautta varmistetaan, että vain valtuutetuilla käyttäjillä on pääsy palveluun ja sen tarjoamiin tietolähteisiin. (Microsoft, 2025 g)

Microsoft Entra ID:n tärkeä ominaisuus on monivaiheinen tunnistautuminen (MFA). Tämä käytäntö lisää ylimääräisen suojakerroksen käyttäjien kirjautumisprosessiin. MFA vaatii käyttäjiä vahvistamaan henkilöllisyytensä toisen menetelmän, kuten tekstiviestin tai autentikointisovelluksen avulla. Organisaatiot voivat myös määrittää erilaisia ehdollisia pääsykäytäntöjä, esimerkiksi sijaintiin tai käytettävään laitteeseen liittyen. Microsoft Entra ID:n yksittäinen kirjautumisen ominaisuus (SSO) on myös hyödyllinen turvallisuus ominaisuus. Sen avulla käyttäjät voivat kirjautua useisiin sovelluksiin yhdellä kertaa. Tämä parantaa käyttökokemusta ja lisää turvallisuutta, koska se vähentää muistettavien ja hallittavien salasanoiden määrää. (Creative Networks, 2024)

Roolipohjainen käyttöoikeuksien hallinta (RBAC) on keskeinen osa Power BI:n käyttöoikeuksien hallintaa. Työtilojen avulla käyttäjät voivat organisoida sisältöä ja hallita pääsyä määrittämällä rooleja, kuten Admin, Member, Contributor ja Viewer. Tällä varmistetaan, että arkaluontoisia tietoja ei joudu luvattomille käyttäjille. Datan turvallisuuden parantamiseksi voidaan myös käyttää rivikohtaista turvallisuutta (RLS), joka rajoittaa käyttäjien näkyvyyttä vain niihin tietoihin, jotka liittyvät heidän rooliinsa. (Creative Networks, 2024)

Power BI:n tietoturvan varmistamiseksi on tärkeää käyttää salausta, valvoa käyttäjätoimintaa ja suojata tietolähteitä. Tietojen salaus tapahtuu sekä levossa että siirron aikana Azure Storage salauksen ja Transport Layer Security (TLS) avulla. Käyttäjätoiminnan säännöllinen seuranta auttaa havaitsemaan mahdollisia uhkia, kuten luvattomia kirjautumisyriytyksiä ja raporttien käyttöä. Lisäksi on tärkeää suojata tietolähteet, esimerkiksi käyttämällä Power BI -portteja turvallisiin yhteyksiin ja varmistamalla, että vain valtuutetut käyttäjät pääsevät käsiksi tietoon. (Creative Networks, 2024)

4 Työn tavoite ja menetelmät

Työn tavoitteena on osallistua Microsoft Fabricin käyttöönoton alkukartoitukseen ja tehdä vertailututkimus, jossa analysoidaan yrityksen nykyiset data- ja analytiikkaratkaisut sekä arvioidaan Microsoft Fabricin ominaisuuksia ja hyötyjä suhteessa nykyisiin Power BI -ratkaisuihin. Tutkimuksen tuloksena syntyy järjestelmävertailu. Tässä kappaleessa kuvataan tutkimuksen tietoperusta ja aineistonkeruumenetelmän valinta.

Työn tutkimusmenetelmäksi on valittu laadullinen tutkimusmenetelmä. Laadullisessa tutkimuksessa keskitytään tutkittavaan ilmiöön ihmisten kokemusten, merkitysten ja käsitysten kautta. Tavoitteena on tavoittaa ihmisten omat kuvaukset heidän kokemastaan todellisuudestaan ja ymmärtää heille merkitykselliset asiat. Laadullisen tutkimuksen aineistona voidaan käyttää haastattelua, havainnointia ja muuta ihmisen tuottamaa teksti- ja kuvamateriaalia. Tämä lähestymistapa mahdollistaa syvällisen analyysin, joka ei rajoitu pelkästään tilastollisiin tietoihin ja saadaan esiin myös henkilökohtaisia kokemuksia ja käsityksiä. Laadullisen tutkimuksen tavoitteena ei ole täydellisen totuuden löytäminen vaan syventää ymmärrystä tutkittavasta ilmiöstä. (Vilka, 2021)

Tutkimuksen aineisto kerätiin laadullisin tutkimuksen menetelmin pieneltä joukolta ihmisiä. Opinnäytetyön kirjoittamisen aikana tehtiin valintaa toimeksiantajan kumppanista Microsoft Fabricin käyttöönotolle ja tätä prosessia varten käydyt keskustelut toimivat aineistona. Lisäksi aineistoa kerättiin käymällä yksittäisiä keskusteluja henkilöiden kanssa, joilla oli tietoa yrityksen data- ja analytiikkaratkaisusta.

Haastattelu on valittu tutkimuksen aineistonkeruumenetelmäksi, koska se mukautuu monenlaisiin tilanteisiin. Haastattelun etuja on se, että saadaan esiin enemmän yksilöllisempiä kokemuksia. Haastattelun hyötyihin kuuluu myös se, että voidaan pyytää tarkennuksia tai selityksiä aiheesta. Hyviä puolia on myös se, että kysymykset voidaan kohdistaa henkilöille, joilla on kokemusta tutkittavasta aiheesta. (Puusa ym., 2020)

Haastattelun huonoja puolia voivat olla esimerkiksi tulkintavirheet. Tulkintavirheitä voi syntyä esimerkiksi, jos kysymyksiä ei muotoilla selkeästi. Tutkija saattaa myös vaikuttaa haastateltavan vastauksiin johdattelemalla kysymyksiä tietyllä tavalla. Jännitys ja sosiaalisen hyväksynnän vaikutus voivat myös vaikuttaa vastauksiin. Haastattelut voivat tapahtua sähköisesti, puhelimesta tai kasvokkain. (Puusa ym., 2020)

Haastattelulajeja on useita ja tässä opinnäytetyössä on käytetty teemahaastattelua ja avointa haastattelua. Teemahaastattelu on vapaamuotoinen ja etenee etukäteen sovitun teeman mukaan. Tällöin tutkijalla on oltava riittävä ymmärrys tutkimuksen kohteena olevasta ilmiöstä ja keskustelussa on tärkeää, että molemmat osapuolet ymmärtävät kontekstin ja puhuvat samaa kieltä. Avoimessa haastattelussa taas ei ole tarkasti ennalta suunniteltuja kysymyksiä, vaan keskustelu etenee luonnollisesti tietyn aihepiirin ympärillä. Avoin haastattelu on hyödyllinen erityisesti esitutkimusvaiheessa, kun halutaan kerätä alustavaa tietoa. (Puusa ym., 2020)

Tutkimuksessa on käytetty myös havainnointia, joka on yleinen laadullisen tutkimuksen menetelmä. Havainnointi keskittyy aluksi ympäristöön, toimintaan ja ihmisiin ja myöhemmin tutkija voi tarkastella erityisiä ilmiöitä. Havainnointia voi tukea valokuvilla ja muistiinpanoilla. Havainnointi voi tapahtua myös digitaalisilla alustoilla. Tutkijan rooli on nostaa esiin tärkeitä havaintoja, jotka tuovat merkityksellistä tietoa tutkimuksen aiheesta. (Puusa ym., 2020)

Luotettavuuden ja eettisyyden arviointi on keskeinen osa laadullista tutkimusta ja sen tavoitteena on varmistaa tutkimuksen uskottavuus ja tulosten todenmukaisuus. Opinnäytetyön aineistoa on käsitelty huolellisesti ja tutkimuksen luotettavuuden varmistamiseksi on laadittu tutkimuslupa ja aineistonhallintasuunnitelma toimeksiantajan kanssa. Haastatteluaineiston kerääminen ja analyysi on toteutettu systemaattisesti, jotta saatu tieto kuvastaisi tutkittavaa ilmiötä mahdollisimman tarkasti. (Hakala & oy, 2024)

5 Nykytila ja käyttöönoton alkukartoitus

Tässä luvussa tutkitaan toimeksiantajan nykyisiä data- ja analytiikkaratkaisuja. Luku alkaa nykytilan analyysillä, jossa selvitetään, miten data- ja analytiikkaratkaisut on tällä hetkellä toteutettu yrityksessä. Sen jälkeen kuvataan mitkä ovat tarpeet ja tavoitteet siirtymisessä Microsoft Fabriciin. Luvun lopussa selvitetään Microsoft Fabricin käyttöönoton keskeiset vaiheet.

5.1 Nykytilan analyysi

Toimeksiantajan data tulee useista eri järjestelmistä, jotka ovat pääasiassa on-premise pohjaisia. Muutamia pilvipohjaisia järjestelmiä on käytössä, mutta niiden osuus raportoinnista on vähäinen. Suurin osa datasta kulkee SQL Serverin kautta Power BI:n.

Suuri osa datasta tulee kahdesta ERP- järjestelmästä, jotka vastaavat toiminnanohjauksesta. Taloudenhallintaan käytetään Wintime-järjestelmä ja henkilöstöhallinnolla on käytössä Sympa. Tuotannonohjauksessa hyödynnetään useita järjestelmiä, kuten MES (iPes, DieShop), Novi, Galcont, Paintline ja Progrex. Myynnin hallintaan käytetään LimeCRM:ää ja Wapicea. Työajanseuranta hoidetaan Jotbar-järjestelmällä.

Jokaisella järjestelmällä on oma roolinsa liiketoiminnan eri osa-alueilla. Data tulee monessa eri muodossa ja lähteestä, joten sen vuoksi on rakennettu ETL-prosesseja. ETL-prosessit ovat pääosin raakadatan replikointia ilman merkittäviä muunnoksia. Tämä mahdollistaa BI-raportoinnin ilman suoria yhteyksiä lähdejärjestelmiin. SSIS hallinnoi ETL-prosesseja. Reaaliaikaisten muutosten tarkasteluun käytetään CDC:tä, joka on rakennettu ERP-järjestelmän rinnalle toimimaan. CDC:lle on rakennettu erillinen käyttöliittymä, josta voidaan seurata tärkeitä valvonnan alla olevia muutoksia tehtaalla. CDC replikoi kaiken datan SQL serverille.

Eri järjestelmistä saatu data kulkeutuu SQL-tietokantaan, joka toimii replikoituna tietokantana. Datan replikoiminen eri järjestelmistä voi tapahtua osittaisena (partial data replication) tai täydellisenä (full data replication). Tietokannasta data siirtyy Power BI -

ympäristöön, jossa sitä hyödynnetään raportoinnissa. Toimeksiantajalla on käytössä Power BI Service.

Toimeksiantajalla on yli 150 raporttia jakautuneena 20–30 työtilaan. Tärkeäksi nostettu raportti on toimitusvarmuusraportti, joka tarjoaa tärkeää tietoa toimituksien onnistumisista. Raporttien rakennetta ei ole keskitetty eikä semanttisia malleja hyödynnetä uudelleen, vaan jokainen raportti tehdään erikseen. Operatiivisten raporttien päivitysväli on 5–15 minuuttia, kun taas useimmat raportit päivittyvät 1–8 kertaa päivässä.

5.2 Tarpeet ja tavoitteet

Toimeksiantajan kanssa käydyissä keskusteluissa nousi useita tärkeitä syitä, miksi yritys haluaa siirtyä Microsoft Fabriciin. Yrityksen suurin datan lähde on tällä hetkellä on-premise pohjainen ERP-järjestelmä, jota ollaan vaihtamassa pilvipohjaiseksi tulevina vuosina. Tämä siirtyminen tarkoittaa, että yrityksen on myös tehtävä muutoksia data- ja analyytiikkaratkaisuihinsa.

Raportointi nykyisessä Power BI -ympäristössä koetaan hankalaksi ja yritys toivoo parannuksia, jotta päätöksenteko voisi perustua entistä paremmin tietoon. Raportointia tulee yhtenäistää niin että se tapahtuu samalla tavalla konsernin sisällä ja läpinäkyvyys säilyy. Tärkeänä nähdään, että raportointi saadaan helpommin toteutettavaksi. Uusien raporttien tekijöiden perehdytys on työlästä, sillä data on raakamuodossa ilman selkeitä semanttisia malleja, mikä myös lisää virheriskiä ja raporttien kehityksen kuormittavuutta. Datan ja raporttien epäjohdonmukaisuus on mahdollista, koska yhtenäisiä malleja ei hyödynnetä ja raporttien hallinnan puutteet tekevät oikean raportin löytämisestä haastavaa. Lisäksi raporttien välillä on päällekkäisyyksiä, koska selkeä kategorisointi puuttuu. Lähdedataa käytetään suoraan, jolloin pienetkin muutokset vaikuttavat kaikkiin raporteihin erikseen.

Lisäksi yritys haluaa hyödyntää tekoälyn tarjoamia mahdollisuuksia tulevaisuudessa, mutta nykyiset on-premise ratkaisut eivät tue tätä kehitystä. Microsoft Fabricin käyttöönoton avulla voidaan integroida tekoälyratkaisuja analyytiikkaan, joka on tärkeää, kun halutaan ennakoita esimerkiksi myyntiä, asiakaskäyttäytymistä tai varastonhallintaa.

Toimitusvarmuusraportti on yksi yrityksen keskeisimmistä ja seuratuimmista raporteista. Sen kehittäminen nähdään tärkeänä osana Microsoft Fabricin käyttöönottoa, sillä raportin tarkkuuden ja luotettavuuden parantaminen tukee liiketoiminnan päätöksentekoa ja operatiivista tehokkuutta. Toimitusvarmuusraportin kehittämällä halutaan tuoda myös läpinäkyvyyttä konsernin sisällä, että raportointi tapahtuu eri maissa samalla tavalla.

Yrityksessä halutaan kehittää toimintaa konsernitasolla. Alankomaissa sijaitseva Nedal Aluminium on myös samojen muutosten edessä, mikä tarjoaa mahdollisuuden yhtenäisten ratkaisujen luomiseen konsernin sisällä. Tarkoituksena on tehdä Microsoft Fabricin käyttöönotto yhdessä.

5.3 Keskeiset vaiheet Microsoft Fabricin käyttöönotolle

Microsoft Fabricin käyttöönotto toimeksiantajayrityksessä on vasta alkuvaiheessa, joten Microsoftin Azure-alustalle toteutettavan käyttöönoton kaikkia vaiheita ei ole vielä opinnäytetyön kirjoitusvaiheessa saatu toteutettua. Tarkoituksena on kuvata vaiheita, joita käyttöönotolle on suunniteltu yhteistyössä kumppanin kanssa. Työn kirjoitusvaiheessa on aloitettu alkukartoitus Microsoft Fabricin käyttöönottamiselle.

Toimeksiantaja on valinnut Microsoft Fabricin käyttöönottamiseen kumppanin, jonka tehtävänä on vastata teknisestä arkkitehtuurista ja pääasiallisesta toteutuksesta. Kumppanilta odotetaan asiantuntemusta hybridisiirtymän hallinnassa sekä kykyä tukea arkkitehtuurin suunnittelussa ja toteutuksessa. Kumppanin rooli on keskeinen Microsoft Fabricin käyttöönotossa ja toimeksiantaja tavoittelee pitkäjänteistä yhteistyötä. Toimeksiantaja pilvisiirtymä toteutetaan hybridimallilla. Tavoitteena on hyödyntää mahdollisimman paljon olemassa olevaa on-premise-infrastruktuuria.

Kumppanin kanssa käydyissä keskusteluissa on kuvattu, että data-alustan muutos toteutetaan kolmessa vaiheessa (kuva 10.). Ensimmäinen vaihe koskee alkukartoitusta, jonka osana tämä opinnäytetyö toimii. Tämä vaihe kestää noin kuukauden ja sen tarkoituksena on kartoittaa toimeksiantajan nykytila. Nykytilan analyysillä selvitetään lähdejärjestelmät sekä liiketoimintatiedon hallinnan teknologiat. Lisäksi on tärkeää käydä keskusteluja eri sidosryhmien kanssa sekä määritellä tavoitetila. Tavoitellulla tilalla kuvataan, että millaiseen lopputulokseen pyritään. Laaditaan myös suunnitelma (roadmap), jossa kuvataan vaiheet ja aikataulut tavoitteeseen pääsemiseksi.

Kuva 10. Keskeiset vaiheet Microsoft Fabricin käyttöönotolle



Toisessa PoC-vaiheessa testataan ja pilotoidaan alustaa käytännössä, jotta voidaan arvioida sen soveltuvuutta ennen laajempaa käyttöönottoa. ERP-järjestelmät integroidaan datalähteiksi. Päivitetään roadmapia seuraavia kehitysvaiheita varten ja harjoitellaan käyttöönottamista. Data varastoidaan Microsoft Fabricin OneLakeen. Käytetään Medallion-arkkitehtuurimallia ja se mallinnetaan semanttiseksi malliksi, jota käytetään Power BI:ssa.

Viimeinen vaihe pitää sisällään jatkokehityksen ja ylläpidon. Tässä vaiheessa lisätään uusia integraatioita ja datalähteitä. Olemassa olevia ratkaisuja parannellaan ja data-alustaa laajennetaan operatiiviselle datalle.

6 Vertailututkimus

Tässä luvussa vertaillaan Power BI:n ja Microsoft Fabricin ominaisuuksia. Vertailun tavoitteena on ymmärtää, miten nämä kaksi järjestelmää eroavat toisistaan ja millaisia vaikutuksia Microsoft Fabricin käyttöönotolla on.

Power BI ja Microsoft Fabric eivät ole suoraan verrattavissa toisiinsa, koska niillä on erilaiset käyttötarkoitukset. Power BI on tarkoitettu liiketoimintatiedon visualisointiin ja raportointiin. Microsoft Fabric on kattava data-alusta, joka mahdollistaa koko datan elinkaaren hallinnan sekä reaaliaikaisen tiedon hallinnan, tekoälyn ja keskitetyn datan varastoinnin OneLake:ssa. Power BI integroituu myös Fabriciin, joten Power BI:n käyttäminen onnistuu edelleen.

Toimeksiantaja toimii teollisuusalalla, jossa tehokas datan hallinta on keskeinen vaatimus. Alalla kertyy dataa useista eri järjestelmistä ja on tärkeää, että nämä tiedot voidaan yhdistää ja analysoida tehokkaasti. Reaaliaikainen datan hallinta on erityisen tärkeää, koska se mahdollistaa nopeat reagoinnit tuotantoprosessien muutoksiin. Tällä voidaan parantaa operatiivista tehokkuutta ja varmistaa liiketoiminnan sujuvuus. Microsoft Fabric sisältää Real-Time Intelligence (RTI) -komponentin, joka on suunniteltu käsittelemään suuria määriä suoratoistodataa. RTI mahdollistaa tiedon keräämisen, analysoinnin ja hyödyntämisen reaaliajassa. Tämä tarjoaa organisaatioille mahdollisuuden havaita ja käsitellä esimerkiksi tuotantolinjan poikkeamat välittömästi, mikä parantaa operatiivista tehokkuutta ja vähentää häiriöitä.

Power BI:ssa monimutkaisemmat raportoinnit tai analyysit tarvitsevat DAX- tai Power Query -kielen osaamista ja se voi tuottaa haasteita, jos osaavia henkilöitä ei löydy organisaation sisältä. Fabric sisältää työkaluja, kuten Data Factoryn ja Data Engineeringin, jotka helpottavat datan käsittelyä ilman ulkoisia ETL-ratkaisuja. Lisäksi Microsoft Fabricin tekoälyominaisuudet voivat vähentää manuaalisen työn määrää ja auttaa organisaatioita hyödyntämään dataa tehokkaammin ilman syvällistä teknistä osaamista. OneLake toimii keskitettynä tietovarastona Microsoft Fabricissa ja keskitetty tiedonvarastointi helpottaa datan hallintaa ja käyttämistä organisaatiossa.

Puhtaan datan käyttö on teollisuusalan yritykselle tärkeää, jotta vältetään virheellisten tietojen käyttö ja varmistetaan luotettavat päätökset. Microsoft Fabric tarjoaa

ominaisuuksia, jotka parantavat datan laatua ja vähentävät virheriskejä. Medallion-arkkitehtuurimallin avulla data jaetaan kolmeen kerrokseen, mikä mahdollistaa puhtaamman ja skaalautuvamman datan käytön. Microsoft Purview vastaa datan hallintakäytänteistä, auttaen organisaatioita pitämään datan turvallisena ja hallitsemaan sen luokittelua ja jäljitettävyyttä. Power BI:ssä ei ole sisäänrakennettua datan puhdistusta.

Microsoft Fabricissa käyttäjien todennus tapahtuu Microsoft Entra ID:n kautta. Ulkopuolisille käyttäjille voidaan myöntää pääsy vain organisaation suostumuksella. Microsoft Fabric tarjoaa myös työkaluja, kuten metatietojen skannausta ja tietoketjun visualisointia, jotka tukevat suurten ympäristöjen datanhallintaa. Purview tarjoaa työkaluja datan luokitteluun sekä tietoturvan parantamista ja käyttäjätoimintojen auditointia. Power BI:ssa turvallisuus perustuu myös Entra ID:hen ja se hyödyntää monivaiheista tunnistautumista (MFA) sekä roolipohjaista käyttöoikeuksien hallintaa (RBAC). Yhteenvetona voidaan todeta, että molemmat tarjoavat hyviä turvallisuusratkaisuja, mutta Fabricin käyttöönotto tarjoaa laajempia turvallisuusominaisuuksia datan hallintaan, mikä parantaa turvallisuutta entisestään.

Microsoft Fabric tarjoaa monia tekoäly- ja analytiikkatyökaluja, jotka mahdollistavat tarkempien ennusteiden tekemisen ja parantavat päätöksentekoa. Tekoäly auttaa syventämään datan analysointia, ennustamaan tulevia trendejä ja käsittelemään luonnollista kieltä. Käyttäjät voivat esittää kysymyksiä ja saada selkeitä vastauksia ilman syvällistä teknistä osaamista.

Medallion-arkkitehtuurimalli helpottaa eri käyttäjäryhmien työskentelyä datan parissa. Kultakerroksessa liiketoiminnan päätöksentekijät voivat keskittyä valmiisiin raportteihin, kun taas pronssi- ja hopeakerroksessa voidaan paneutua teknisempiin asioihin, kuten esimerkiksi tiedon keräämiseen ja muuntamiseen. Tämä malli vähentää raporttien tekijöiden perehdyttämiseen liittyviä haasteita ja parantaa raporttien johdonmukaisuutta, joka voi johtaa luotettavampiin päätöksiin.

Power BI ja Microsoft Fabric tarjoavat erilaisia ratkaisuja liiketoimintatiedon hallintaan ja analysointiin. Power BI on vahva työkalu visualisointiin ja raportointiin, mutta sen datan hallintamahdollisuudet ovat rajalliset verrattuna Microsoft Fabriciin. Fabricin laajat datan hallinta-, analytiikka- ja tekoälyominaisuudet tekevät siitä hyvän ratkaisun erityisesti organisaatioille, jotka tarvitsevat kattavan data-alustan koko datan elinkaaren hallintaan.

7 Tulokset

Tulokset osoittavat, että Microsoft Fabric tarjoaa laajemmat mahdollisuudet datan hallintaan ja analysointiin, kun taas Power BI keskittyy erityisesti datan visualisointiin ja raportointiin. OneLake toimii keskitettynä tietovarastona koko organisaation käyttöön, mahdollistaen datan tehokkaan jakamisen ja hallinnan.

Microsoft Fabricin tekoälyominaisuudet tarjoavat organisaatioille uusia keinoja hyödyntää dataa, mikä oli yksi toimeksiantajan keskeisistä syistä siirtyä Microsoft Fabricin käyttöön. Tekoälyä voidaan hyödyntää esimerkiksi raportoinnissa, mikä parantaa analyysin tarkkuutta ja päätöksenteon perustaa.

Toimeksiantajalla oli haasteena uusien raporttien tekijöiden perehdyttäminen, koska dataa käytetään suoraan lähdejärjestelmästä, eikä selkeitä semanttisia malleja ole käytössä. Microsoft Fabricin Medallion-arkkitehtuurimalli selkeyttää datan hallintaa ja eri käyttäjäryhmien työskentelyä datan parissa. Lisäksi puhtaampi data vähentää virheriskejä.

Tulosten havainnollistamista varten on luotu taulukko 1 havainnollistamaan eroja järjestelmien välillä ja vaikutuksia Microsoft Fabricin käyttöönotosta. Ominaisuuksien valinnassa on otettu huomioon toimeksiantajayrityksen tarpeet ja tavoitteet, jotka liittyvät datan hallintaan, tekoälyyn ja raportoinnin tehostamiseen sekä muita keskeisiä ominaisuuksia.

Taulukko 1. Power BI:n ja Microsoft Fabric:in vertailun tulokset

Ominaisuus	Power BI	Microsoft Fabric	Vaikutukset
Datan tallennus	Ei sisäänrakennettua tallennusmahdollisuutta.	OneLake toimii keskitettynä tallennuspaikkana kaikille datatyypeille.	Parantaa datan hallintaa, poistaa tarpeen hajautetuille tietovarastoille. Vähentää datasiiloja.
Tekoäly	Tekoälyominaisuudet käytössä vain Power BI Premium lisenssillä.	Copilot, kehittyneet koneoppimismallit ja ennakoiva analytiikka	Tarkemmat liiketoimintaennusteet ja parempi päätöksenteko.
Raportointi ja analytiikka	Monipuoliset ja interaktiiviset visualisoinnit. DAX-kieli monimutaisimpien raporttien luomisessa.	Reaaliaikainen ja ennakoiva analytiikka. Sisältää Power BI:n toiminnot.	Laajempi analytiikkakattavuus ja tekoäly osaksi päätöksentekoa.

		Tekoälyavusteiset toiminnot.	
Datan käsittely	Power Query, Dataflows tai ELT/ETL-työkalut. Hidas käsittelemään suuria tietoaaineistoja.	Tukee sekä eräajoa että reaaliaikaista käsittelyä, Apache Spark suurille tietomassoille, edistyneet ETL/ELT-toiminnot	Parantaa datan käsittelyä. Suurempien tietoaaineistojen käsittely nopeampaa.
Kustannukset	Käyttäjakohtainen lisenssi.	Maksaminen käytön mukaan tai varattu kapasiteetti	Vaatii tarkempaa kustannusseurantaa.
Reaaliaikainen data	Power BI:ssa Pro-lisenssillä enintään 8 ja Premium-lisenssillä jopa 48 päivitystä päivässä.	RTI mahdollistaa reaaliaikaisen datan käsittelyn ja analysoinnin.	Nopeampi reagointi liiketoiminnan tarpeisiin. Ajantasaiset oivallukset liiketoiminnan tiedosta.
Turvallisuus ja hallinta	Microsoft Entra ID:n perustuva. MFA ja RBAC.	Microsoft Entra ID, johon käyttäjät ja OneLake yhdistyvät. Purview datan luokitteluun, tietoturvaan ja auditointiin. Metatietojen skannaus ja domaineja työtilojen ja datavarastojen ryhmittelyyn.	Parempi käyttäjien ja datan hallinta. Lisää tietoturvaa.
Käyttökohteet	Liiketoimintatiedon analysointi ja datan visualisointiin.	Kokonaisvaltainen data-alusta datanhallintaan ja analytiikkaan.	Kaikki data- ja analytiikkaratkaisut yhdessä paikassa.
Datan puhdistus	Ei sisäänrakennettua datan puhdistustyökalua	Medallion-arkkitehtuuri ja Purview	Data puhtaampaa ja virheriski pienenee

8 Johtopäätökset ja pohdinta

Työn teoriaosuudessa käsitellään liiketoimintatiedon hallinnan eri vaiheita ja erityisesti niitä prosessin vaiheita, jotka ovat tämän työn kannalta keskeisiä. Tarkoituksena hahmottaa millaisia vaiheita data käy läpi ennen kuin sitä voidaan hyödyntää raportoinnissa sekä ymmärtää, että miksi nykyisiä datanhallintaratkaisuja halutaan päivittää tai korvata uudella järjestelmällä. Datan hallinta ja sen kulku liiketoimintatiedon hyödyntämiseen on monivaiheinen prosessi ja sen takia myös teoriaosuudessa oli tarpeen käsitellä tätä aihetta laajasti.

Työn käytännön osuudessa kuvataan toimeksiantajan nykytila sekä selvitetään, mitä järjestelmiä yrityksellä on käytössä, miten data- ja analytiikkaprosessit on rakennettu ja millaista raportointia Power BI -ympäristössä tehdään. Tämän jälkeen käsitellään tarpeita ja tavoitteita, jotka lisäävät ymmärrystä siitä, miksi nykyiset Power BI -ratkaisut eivät ole riittäviä ja miksi järjestelmiä ollaan muuttamassa. Lopputuloksena laadittiin järjestelmävertailu, jossa havainnollistetaan taulukon avulla Power BI:n ja Microsoft Fabricin eroja ominaisuuksien perusteella sekä arvioidaan käyttöönoton vaikutuksia.

Tulosten perusteella Microsoft Fabric tarjoaa merkittäviä mahdollisuuksia raportoinnin ja datan hallinnan kehittämiseen. Keskitetty data-alusta tehostaa tiedonhallintaa ja varmistaa, että eri sidosryhmät pääsevät hyödyntämään ajantasaista tietoa. Lisäksi yhteistyö konsernin muiden toimijoiden, kuten Alankomaissa sijaitsevan Nedal Aluminiumin kanssa, tukee yhtenäisten ratkaisujen kehittämistä koko organisaatiossa.

Power BI on suunniteltu liiketoimintatiedon analysointiin ja datan visualisointiin, mutta sen rajoitteena on, että se ei tarjoa kattavia datanhallintaominaisuuksia, kuten datan varastointia tai ennakoivia analytiikkamenetelmiä. Microsoft Fabric puolestaan suuntaa yrityksiä enemmän kohti tietoon perustuvia päätöksentekoprosesseja, mahdollistaen laajemmat analytiikkaratkaisut. Suurena apuna tässä ovat tekoälyominaisuudet, jotka auttavat myös niitä käyttäjiä, joilla ei ole vahvaa teknistä osaamista.

Järjestelmien kustannusten vertailu jäi vähäiseksi, koska työssä keskityttiin enemmän järjestelmien ominaisuuksiin ja niiden vaikutuksiin. Tarkempi kustannusten arviointi olisi kuitenkin ollut tärkeä osa vertailua ja se olisi voinut tarjota lisäarvoa.

Työ on toiminut oppimiskokemuksena, jonka kautta on saatu hyödyllistä tietoa liiketoimintatiedon hallinnan prosessien kehittämiseen ja käyttöönottoprosessin seuraaviin vaiheisiin. Työ voi myös toimia apuna myös muille yrityksille, jotka miettivät Microsoft Fabricin käyttöönottamista. Alkukartoituksen jälkeen seuraavana vaiheena on PoC-vaihe, jossa alusta perustetaan ja siinä hyödynnetään alkukartoitusvaiheessa kerättyä tietoa ja suunnitelmia.

Toimeksiantaja on ollut tyytyväinen työn teoreettiseen lähestymistapaan, sillä Microsoft Fabric on uusi järjestelmä kaikille osapuolille ja teoriaosuus tarjoaa hyödyllisen pohjan sen ymmärtämiseen. Kaikki eivät ole vielä ehtineet perehtyä työhön, mutta saadun palautteen perusteella vaikuttaa siltä, että teoreettinen osuus tukee hyvin tulevia käytännön toteutuksia. Se auttaa hahmottamaan järjestelmän mahdollisuuksia ja vaikutuksia ennen varsinaista käyttöönottoa.

Aikataulu on tuonut omat haasteensa, koska työn kirjoitusvaiheessa oltiin vasta alkukartoitusvaiheessa, eikä Microsoft Fabricia ollut vielä otettu käyttöön. Tästä huolimatta työhön koottua tietoa voidaan hyödyntää seuraavissa vaiheissa, kun käyttöönoton suunnittelua ja toteutusta edistetään.

9 Yhteenveto

Opinnäytetyössä tutkittiin Microsoft Fabricin keskeisiä ominaisuuksia ja verrattiin niitä Power BI:hin. Vertailututkimuksen avulla arvioitiin, miten järjestelmän käyttöönotto voisi vaikuttaa toimeksiantajan nykyisiin datanhallinta- ja analytiikkaprosesseihin. Lisäksi kartoitettiin toimeksiantajan tarpeita ja tavoitteita sekä kuvattiin Microsoft Fabricin käyttöönoton keskeiset vaiheet.

Tutkimuskysymyksiin pystyttiin vastaamaan kattavasti teoriatasolla, mutta koska Microsoft Fabricin käyttöönotto oli vielä alkuvaiheessa, niin työn painopiste oli järjestelmän ominaisuuksien ja vaikutusten tarkastelussa käytännön toteutuksen sijaan. Kaikkiin kolmeen tutkimuskysymykseen saatiin vastaukset, mutta käyttöönotto kuvattiin suunniteltujen vaiheiden pohjalta, koska opinnäytetyö tehtiin ensimmäisessä alkukartoitusvaiheessa, joka oli vielä kesken työn kirjoitusvaiheessa. Käyttöönoton vaikutuksia kuvattiin teoriatasolla vertailututkimuksen avulla, mutta tästä ei vielä saatu käytännön kuvaa siitä, miten käyttöönotto vaikuttaa yrityksen data- ja analytiikkaprosesseihin.

Opinnäytetyön aikana minulle konkretisoitui liiketoimintatiedon hallinnan monivaiheisuus, erityisesti teollisuusalan kontekstissa. Osallistuminen uuden järjestelmän käyttöönoton valmisteluun havainnollisti, kuinka tärkeää on panostaa huolelliseen suunnitteluun ja alkuvaiheen selkeään määrittelyyn. Työ toimi oppimiskokemuksena, jonka kautta pystyin syventämään osaamistani datan hallinnasta yleisesti. Osallistuin useisiin tapaamisiin, joissa valittiin kumppania vetämään Microsoft Fabric -migraatiota. Näissä keskusteluissa sain arvokasta oppia siitä, kuinka monia erilaisia lähestymistapoja on olemassa tämän siirtymän toteuttamiseksi.

Tulevaisuudessa jatkotutkimusaiheina olisi hyödyllistä tutkia Microsoft Fabricin käyttöönoton etenemistä ja sen vaikutuksia käytännössä. Jatkotutkimus voisi keskittyä esimerkiksi siihen, miten hyvin käyttöönotto on sujunut, mitä haasteita on kohdattu ja millaisia hyötyjä yritys on saavuttanut pitkällä aikavälillä.

Lähteet

- Ait Errami, S., Hajji, H., Ait El Kadi, K., & Badir, H. (2023). Spatial big data architecture: From Data Warehouses and Data Lakes to the LakeHouse. *Journal of Parallel and Distributed Computing*, 176, 70–79. <https://doi.org/10.1016/j.jpdc.2023.02.007>
- Ali, A., & Schacht, B. (2024). *Learn Microsoft Fabric: A Practical Guide to Performing Data Analytics in the Era of Artificial Intelligence*. Birmingham: Packt Publishing, Limited.
- Altaiar, H., Lee, J., & Peña, M. (2021). *Cloud Analytics with Microsoft Azure: Transform your business with the power of analytics in Azure*. Packt Publishing, Limited.
- AltexSoft. (2022, lokakuuta 6). *The Good and the Bad of Microsoft Power BI Data Visualizatio*. AltexSoft. <https://www.altexsoft.com/blog/power-bi-pros-cons/>
- Atlan. (2024, joulukuuta 9). *Data Orchestration 101: Benefits, Parts & Example (2025)*. <https://atlan.com/data-orchestration-101/>
- Banerjee, Dr. K., Das, S., & Nath, S. (2023). Data visualization approach for business strategy recommendation using power BI dashboard. *International Journal of Research in Management*, 6(1), 168–175. <https://doi.org/10.33545/26648792.2024.v6.i1b.138>
- Ben Lorica, Michael Armbrust, Reynold Xin, Matei Zaharia, & Ali Ghodsi. (2020, tammikuuta 30). What Is a Lakehouse? *Databricks*. <https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>
- Borra, P. (2024). Microsoft Fabric Review: Exploring Microsoft’s New Data Analytics Platform. *International Journal of Computer Science and Information Technology Research*, 12(2), 34–39. <https://doi.org/10.5281/zenodo.11502585>
- Bothma, J. (2024, maaliskuuta 20). *A Step-By-Step Guide to Visualizing KPIs in Power BI*. <https://www.datacamp.com/tutorial/power-bi-kpi>
- CCS Learning Academy. (2024, lokakuuta 23). Top 10 Power BI Limitations: Must Know List for 2025. *CCSLA Learning Academy*. <https://www.ccslearningacademy.com/top-power-bi-limitations/>

Cole Stryker. (2024, kesäkuuta 14). *What Is a Data Pipeline?* | IBM.

<https://www.ibm.com/think/topics/data-pipeline>

Crabtree, M., & Nehme, A. (2024). *What is Data Analysis? An Expert Guide with Examples*. Datacamp.

<https://www.datacamp.com/blog/what-is-data-analysis-expert-guide>

Creative Networks. (2024, elokuuta 7). *10 Best Practices for Power BI Data Security—Creative*

Networks. <https://www.creative-n.com/blog/best-practices-for-power-bi-data-security/>

da Costa Santos, M. A. M. (2020). *Monitoring Framework for Clinical ETL processes and associated performance resources*. <https://repositorio-aberto.up.pt/bitstream/10216/131915/2/440759.pdf>

Databricks. (2022, syyskuuta 3). *What is a Medallion Architecture?* Databricks.

<https://www.databricks.com/glossary/medallion-architecture>

Erkkilä, M., Kaivola, T., & Still, V. (2023). *Datanhallinta-asetus: Kansallinen täytäntöönpano. Liikenne- ja viestintäministeriö, 2023, s. 17.*

Ghosh, D. (2024). *Mastering Microsoft Fabric: SAASification of Analytics*. Berkeley, CA: Apress L. P.

<https://doi.org/10.1007/979-8-8688-0131-0>

Google Cloud. (n.d.a). *What is a Data Warehouse?* Google Cloud. <https://cloud.google.com/learn/what-is-a-data-warehouse>

Google Cloud. (n.d.b). *What is cloud analytics?* Google Cloud. <https://cloud.google.com/discover/what-is-cloud-analytics>

Google Cloud. (n.d.c). *What is ETL?* Google Cloud. <https://cloud.google.com/learn/what-is-etl>

Gulbahar Karasatas. (2024, joulukuuta 17). *7 Best Data Pipeline Tools With Key Capabilities in 2025.*

AIMultiple. <https://research.aimultiple.com/data-pipeline-tools/>

Hakala, J. T., & oy, G. (2024). *Laadullisen tutkimuksen ABC : menetelmäopas opinnäytteen tekijälle.*

Gaudeamus. <https://www.ellibslibrary.com/fi/hamk/9789523454613>,

<https://hamk.finna.fi/Record/vanaicat.142668>

Harrington, J. (2024). *Importance of the Medallion Architecture in Microsoft Fabric. Omnitech.*

<https://www.omnitech-inc.com/blog/the-importance-of-medallion-architecture-in-microsoft-fabric/>

Hystax. (2024, huhtikuuta 10). *Pros and cons of utilizing Microsoft Azure services* | Hystax.

<https://hystax.com/pros-and-cons-of-utilizing-microsoft-azure-services/>

IBM. (2021, elokuuta 9). *What Is Business Intelligence (BI)?* | IBM.

<https://www.ibm.com/think/topics/business-intelligence>

Kanerika. (2025, helmikuuta 7). Microsoft Fabric vs Power BI: How They Differ and Which One You Need. *Kanerika*. <https://kanerika.com/blogs/microsoft-fabric-vs-power-bi/>

Keboola. (2023, helmikuuta 27). *Power BI Data Preparation in 5 steps*.

<https://www.keboola.com/blog/power-bi-data-preparation>

Kishan. (2024, helmikuuta 21). One lake – one drive of data (a Microsoft fabric product). *Medium*.

<https://medium.com/@kishansingh2411/one-lake-one-drive-of-data-a-microsoft-fabric-product-01d6aa618fd7>

Kumar V K, A. (2021). *Business Intelligence Demystified: Understand and Clear All Your Doubts and Misconceptions about BI (English Edition)*. BPB Publications.

<http://ebookcentral.proquest.com/lib/hamk-ebooks/detail.action?docID=30593786>

Lavanya, A., Sindhuja, S., Gaurav, L., & Ali, W. (2023). A Comprehensive Review of Data Visualization Tools: Features, Strengths, and Weaknesses. *International Journal of Computer Engineering in Research Trends*, 10(1), Article 1. <https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0102>

Marina Pigol. (2024). The Ultimate Power BI AI Features Overview. *Alpha Serve*.

<https://www.alphaservesp.com/blog/the-ultimate-power-bi-ai-features-overview>

Metre, K. V., Mathur, A., Dahake, R. P., Bhapkar, Y., Ghadge, J., Jain, P., & Gore, S. (2024). An Introduction to Power BI for Data Analysis. *International Journal of Intelligent Systems and Applications in Engineering*, 12(1s), Article 1s.

Microsoft. (2023, marraskuuta 15). *Implement medallion lakehouse architecture in Fabric—Microsoft Fabric*. <https://learn.microsoft.com/en-us/fabric/onelake/onelake-medallion-lakehouse-architecture>

- Microsoft. (2024a, maaliskuuta 22). *What is Power BI? - Power BI*. <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>
- Microsoft. (2024b, lokakuuta 30). *What is the medallion lakehouse architecture? - Azure Databricks*. <https://learn.microsoft.com/en-us/azure/databricks/lakehouse/medallion>
- Microsoft. (2024c, marraskuuta 18). *Factory Operations Agent in Azure AI Foundry (preview)—Microsoft Cloud for Manufacturing*. <https://learn.microsoft.com/en-us/industry/manufacturing/manufacturing-data-solutions/overview-copilot-factory-operations>
- Microsoft. (2025a, tammikuuta 23). *New and planned features for Microsoft Cloud for Manufacturing, 2025 release wave 1*. <https://learn.microsoft.com/en-us/industry/release-plan/2025wave1/cloud-manufacturing/planned-features>
- Microsoft. (2025b, tammikuuta 26). *Enable Copilot in Fabric—Microsoft Fabric*. <https://learn.microsoft.com/en-us/fabric/fundamentals/copilot-enable-fabric>
- Microsoft. (2025c, tammikuuta 26). *Overview of Copilot in Fabric—Microsoft Fabric*. <https://learn.microsoft.com/en-us/fabric/fundamentals/copilot-fabric-overview>
- Microsoft. (2025d, tammikuuta 26). *What is Microsoft Fabric—Microsoft Fabric*. <https://learn.microsoft.com/en-us/fabric/fundamentals/microsoft-fabric-overview>
- Microsoft. (2025e, tammikuuta 27). *What is Power Query? - Power Query*. <https://learn.microsoft.com/en-us/power-query/power-query-what-is-power-query>
- Microsoft. (2025f, helmikuuta 26). *Connect to AI Insights in Power BI Desktop—Power BI*. <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-ai-insights>
- Microsoft. (2025g, helmikuuta 28). *Power BI Security—Power BI*. <https://learn.microsoft.com/en-us/power-bi/enterprise/service-admin-power-bi-security>
- Microsoft. (n.d.a). *Extract, transform, load (ETL)—Azure Architecture Center*. <https://learn.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>
- Microsoft. (n.d.b). *What is a Data Lake? Data Lake vs. Warehouse | Microsoft Azure*. <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-lake>

Microsoft. (n.d.c). *What is PaaS? Platform as a Service | Microsoft Azure*.

<https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-paas>

Nortal. (2024, toukokuuta 22). *Unlocking Manufacturing Insights with Microsoft Fabric*. Nortal.

<https://nortal.com/insights/unlocking-manufacturing-insights-with-microsoft-fabric/>

Petrie, K., Ankorion, I., & Potter, D. (2018). *Streaming Change Data Capture*. O'Reilly Media, Inc.

http://sfx.finna.fi/nelli19?url_ver=Z39.88-2004&ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/sfxit.com:opac_856&url_ctx_fmt=info:ofi/fmt:kev:mtx:ctx&sfx.ignore_date_threshold=1&rft.object_id=4100000005043501&svc_val_fmt=info:ofi/fmt:kev:mtx:sch_svc&,
<https://hamk.finna.fi/Record/nelli19.4100000005043501>

Pierre-Yves BONNEFOY, CHAIZE, E., Heckel, S., MANSUY, R., & TAZI, M. (2024). *Definitive Guide to Data Integration: Unlock the Power of Data Integration to Efficiently Manage, Transform, and Analyze Data*. Packt Publishing.

Purso. (n.d.). *Purso*. Purso. <https://purso.fi/purso/>

Puusa, A., Juuti, P., & Aaltio, I. (Toim.). (2020). *Laadullisen tutkimuksen näkökulmat ja menetelmät*. Gaudeamus.

Qlik. (n.d.a). *What is a Data Lakehouse? Definition, Features & Benefits*. Qlik.

<https://www.qlik.com/us/data-lake/data-lakehouse>

Qlik. (n.d.b). *What is Change Data Capture (CDC)? Definition, Best Practices*. Qlik.

<https://www.qlik.com/us/change-data-capture/cdc-change-data-capture>

Rad, R. (2019, tammikuuta 28). What is the Common Data Model and Why Should I Care? Part 3 of Dataflow Series in Power BI. *RADACAD*. <https://radacad.com/what-is-the-common-data-model-and-why-should-i-care-part-3-of-dataflow-series-in-power-bi>

Sandhu, A. K. (2022). Big data with cloud computing: Discussions and challenges. *Big Data Mining and Analytics*, 5(1), 32–40. Big Data Mining and Analytics.

<https://doi.org/10.26599/BDMA.2021.9020016>

Savani, H. (2024, joulukuuta 4). *AI-Powered Insights: Exploring Analytics in Microsoft Fabric*.

<https://www.aegisofttech.com/insights/analytics-in-microsoft-fabric/>

Shapiro, C. (2024). *Real-Time Analytics in Power BI*. Analytics8. <https://www.analytics8.com/blog/real-time-analytics-in-power-bi/>

Snowflake. (n.d.). *Data Analytics for Manufacturing*. Snowflake.

<https://www.snowflake.com/trending/data-analytics-manufacturing/>

Soh, J., Copeland, M., Puca, A., & Harris, M. (2020). *Overview of Azure Platform as a Service* (ss. 43–55). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-5958-0_3

Vilka, H. (2021). *Tutki ja kehitä* (5., päivitetty painos). PS-kustannus.

<https://www.elliblibrary.com/fi/hamk/9789523701731>,

<https://hamk.finna.fi/Record/vanaicat.138708>

Liite 1: Aineistonhallintasuunnitelma

Opinnäytetyön aineiston kuvaus

Aineisto hankitaan haastattelun, havainnoinnin ja kuvallisten aineistojen avulla. Haastatteluaineisto tallennetaan tekstimuodossa kirjoittamalla tekijän omalle koneelle. Havainnointiaineisto koostuu kirjallisista muistiinpanoista sekä mahdollisesti kuvallisista dokumentaatioista. Kuvalliset aineistot voivat sisältää kuvankaappauksia tai muita visuaalisia materiaaleja, esimerkiksi kaavioita.

Aineiston tallennus ja säilytys

Aineisto tallennetaan ja sitä käsitellään opinnäytetyön tekijän omalla salasanalla suojatulla tietokoneella. Aineistosta tallennetaan erilliseen kansioon varmuuskopiot, joita säilytetään erillään analysoitavista tiedostoista. Opinnäytetyön tekijän lisäksi aineistoa käsittelee mahdollisesti myös opinnäytetyön ohjaaja.

Henkilötietojen ja arkaluonteisten tietojen käsittely

Opinnäytetyössä ei käsitellä henkilötietoja.

Aineiston omistajuus

Opinnäytetyön aineiston ja tuloksien omistus- ja käyttöoikeus on opinnäytetyön tekijällä ja toimeksiantajalla.

Aineiston jatkokäyttö työn valmistumisen jälkeen

Tutkimusaineistoa ei jatko käytetä. Opinnäytetyön tekijä säilyttää aineiston tietoturvallisesti vuoden ajan opinnäytetyön hyväksymispäivästä, jotta opinnäytetyön tulokset voidaan tarvittaessa varmistaa ja hävittää tämän jälkeen aineiston tietoturvallisesti.