

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Ats Raigla 210781IAIB

**Andmearhitektuuri analüüs ärianalüütika
kasutamiseks tootmisettevõttes**

Bakalaureusetöö

Juhendaja: Tarvo Treier
MSc

Tallinn 2024

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Ats Raigla

26.12.2024

Annotatsioon

Käesolev bakalaureusetöö keskendub andmearhitektuuri lahenduse analüüsile ja pakkumisele tootmisettevõttes Wolf Group OÜ, mis on üleminekul Microsoft Dynamics Axapta 2009 tarkvaralt pilvepõhisele ERP-süsteemile Microsoft Dynamics 365 Finance and Operations (D365). Töö peamine eesmärk on läbi viia ettevõtte ärianalüütika andmearhitektuuri analüüs, pakkudes välja realistlik ja ettevõtte äriliste ning tehniliste eesmärkidega kooskõlas olev lahendus.

Töös analüüsitakse probleeme, mis tulenevad uue ERP-süsteemi liidestamisest ettevõtte ärianalüütika platvormiga MS Power BI. Kuna Microsofti tööriistade hulka on lisandunud uusi andmearhitektuuri platvorme, nagu MS Fabric, muutub nende tööriistade võimekuse ja piirangute hindamine keerukaks ülesandeks. Töö analüüsib neid tööriistu ning nende sobivust tootmisettevõtte andmeanalüütika vajaduste rahuldamiseks.

Töö on jaotatud 4 põhiosasse. Esmalt selgitatakse ärianalüütika rolli tootmisettevõttes. Seejärel analüüsitakse ettevõtte andmearhitektuuri nõudeid, funktsionaalsete ja mittefunktsionaalsete kriteeriumite kehtestamisega. Töö võrdleb erinevaid andmearhitektuuri lahendusi, kasutades Saaty analüütilise hierarhia meetodit alternatiivide hindamiseks. Töö lõpposas esitatakse analüüsi tulemused ja praktilised soovitused efektiivse ja skaleeritava andmearhitektuuri rakendamiseks.

Töö tulemused pakuvad lisaks ettevõttele Wolf Group OÜ sobivale lahendusele ka põhjaliku juhendi sarnaste väljakutsete ees seisvatele ettevõtetele. Analüüs tõestab, kuidas kaasaegsed andmearhitektuuri lahendused suudavad täiustada äriotsuste protsesse, parandada analüütikavõimekust ja toetada strateegiliste ärieesmärkide saavutamist.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 47 leheküljel, 6 peatükki, 19 joonist, 18 tabelit.

Abstract

Data Architecture Analysis for Business Analytics in a Manufacturing Company

This bachelor's thesis focuses on the analysis and proposal of a data architecture solution for the manufacturing company Wolf Group OÜ, which is transitioning from Microsoft Dynamics Axapta 2009 to the cloud-based ERP system Microsoft Dynamics 365 Finance and Operations (D365). The primary objective of this thesis is to conduct an analysis of the company's business analytics data architecture and propose a realistic solution aligned with the company's business and technical goals.

The thesis addresses challenges arising from integrating the new ERP system with the company's business analytics platform, MS Power BI. Since Microsoft has introduced new data architecture platforms such as MS Fabric, evaluating the capabilities and limitations of these tools has become a complex task. The thesis analyzes these tools and assesses their suitability for meeting the data analytics needs of a manufacturing company.

The thesis is divided into four main sections. First, it explains the role of business analytics in manufacturing companies. Next, it discusses the company's data architecture requirements, establishing functional and non-functional criteria. The thesis then analyzes and compares various data architecture solutions, using Saaty's Analytical Hierarchy Process (AHP) to evaluate alternatives. The final section presents the analysis results and practical recommendations for implementing an effective and scalable data architecture.

The results of this thesis provide Wolf Group OÜ with a suitable solution and offer a comprehensive guide for other companies facing similar challenges. The research demonstrates how modern data architectural solutions can enhance decision-making processes, improve analytics capabilities, and support the achievement of strategic business objectives.

The thesis is in Estonian and contains 47 pages of text, 6 chapters, 19 figures, and 18 tables.

Lühendite ja mõistete sõnastik

ACID	<i>Atomcity, consistency, isolation and durability</i> , omadused on seotud andmebaasi transaktsioonide töökindluse tagamisega
AHP	<i>Analytical Hierarcheis Process</i> , analüütilise hierarhia protsess ehk Saaty meetod mida kasutatakse muutujate analüüsiks ja võrdluseks
AI	<i>Artificial Intelligence</i> , tehisintellekt, modelleeritud ajuprotsessidest tulenev arvuti suutlikkus jäljendada inimese vaimset tegevust
API	<i>Application Programming Interface</i> , rakenduse masinliides, Ühenduslülili kahe tarkvara omavaheliseks andmete vahetamiseks
Axapta	Microsoft <i>Dynamics Axapta</i> , Microsofti lokaalne majandustarkvara
Azure Active Directory	Microsofti poolt pakutav organisatsiooni kasutajahalduse platvorm
Azure Synapse	Microsofti andmeplatvorm, mis tähendab andmete integreerimise, andmebaasi ja andmeanalüüsi, pakkudes terviklikku lahendust suurandmete töötlemiseks ja analüüsiks.
BI	<i>Business Intelligence</i> , Ärianalüütika
Pronks tase	Medaljon andmetalletuse arhitektuuri algtase, kuhu andmed laaditakse nende toorversioonis ilma töötlemiseta
CRM	<i>Customer Relationship Management</i> , kliendisuhete- ja müügihaldustarkvara
CSV	<i>Comma-separated values</i> , komaga eraldatud väärtused, on faililiik, mis sisaldab tabelikujulisi andmeid tavalises tekstifailis
D365	Microsoft <i>Dynamics 365 Finance and Operations</i> , Microsofti pilvepõhine majandustarkvara

DAX	<i>Data Analysis Expressions</i> , programmeerimiskeel arvutuste ja päringute tegemiseks
Databricks	Pilvepõhine platvorm suurandmete töötlemiseks ja masinõppe arendamiseks, mis toetab Apache Spark raamistikku
Data Activator	Microsoft Fabricu tööriist reaalajas sündmuste avastamiseks ja reageerimiseks
ERP	<i>Enterprise Resource Planning</i> , ettevõtte ressursside planeerimise tarkvara
ETL	<i>Extract, Transform, Load</i> , Andmetöötluses kasutatav protsess, kus laaditakse andmed välisest andmeallikast (<i>Extract</i>), siis muundatakse selliseks, et neil oleks kindel struktuur (<i>Transform</i>) ja siis laaditakse andmebaasi (<i>Load</i>)
GB	<i>Gigabyte</i> , Gigabait
Kuld tase	Medaljon andmetalletuse arhitektuuri tase, kus andmed on täielikult töödeldud, puhastatud ja analüüsiks valmis.
IT	<i>Infotechnology</i> , Infotehnoloogia
Logshipping	Andmebaasi kopeerimise ja sünkroniseerimise meetod, mis kasutab transaktsioonilogsid
Medaljon andmetalletus	Andmearhitektuuri mudel, mis jagab andmed erinevatesse etappidesse: Pronks, Hõbe ja Kuld
MES	<i>Manufacturing Execution System</i> , Tootmisjuhtimise tarkvara
MS	<i>Microsoft Corporation Inc</i> , Maaailma juhtiv tarkvara- ja pilvelahenduste pakkuja
MS Azure Data Lakehouse	<i>Microsoft Azure'i</i> platvorm, mis ühendab andmehoidla ja andmejärve omadused analüütikavõimaluste suurendamiseks
MS Fabric	Microsofti andmeplatvorm, mis integreerib erinevaid andmehalduse, analüütika ja koostöö tööriistu
MS Onelake	<i>Microsoft Fabric</i> platvormi keskne andmehalduskiht, mis toimib andmeallikana kogu süsteemi jaoks
OLAP	<i>Online Analytical Processing</i> , andmeanalüütika mudel, mis toetab keerulisi päringuid ja suurandmete analüüsi

Power BI	Microsofti andmeanalüütika ja visualiseerimise tööriist, mis võimaldab luua aruandeid ja juhtimislaudasid
Power BI Desktop	Microsoft Power BI komponent, mida kasutatakse aruannete koostamiseks ja semantiliste andmemudelite loomiseks
Power Query	<i>Power Query</i> , Andmete hankimise, puhastamise ja transformeerimise tööriist
RS	<i>Reporting Service</i> , staatilised SQL andmetel baseeruvad aruanded
Hõbe tase	Medaljon andmetalletuse arhitektuuri vahetase, kus andmed on puhastatud ja valmis täiendavaks analüüsiks
SQL	<i>Structured Query Language</i> , struktureeritud andmebaasikeel
SQL Live Server	SQL andmebaas, mis võimaldab reaalajas andmetöötlust ja andmete haldamist
SSIS	<i>SQL Server Integration Services</i> , Microsofti tööriist andmete integreerimiseks ja ETL protsesside läbiviimiseks

Sisukord

1 Sissejuhatus	13
1.1 Lahendatav probleem	14
1.2 Töö struktuur ja mõisted.....	14
2 Ärianalüütika tootmisettevõttes	16
2.1 Ärianalüütika liidestus andmeallikatega.....	17
2.2 Ärianalüütika mõju äriprotsessidele tootmisettevõttes.....	18
2.3 Andmearhitektuuri eesmärgid	18
2.3.1 Funktsionaalsed ja mittefunktsionaalsed nõuded	18
2.4 ETL (ELT) protseduurid.....	20
2.4.1 Medaljon andmetalletus.....	21
2.5 Andmemudelid	22
2.5.1 Andmemudelid MS Fabric platvormil.....	23
3 Võimalikud andmearhitektuuri lahendused.....	24
3.1 Lähteülesande kitsendused	24
3.2 Vaadeldavad andmearhitektuuri lahendused	25
3.2.1 Synapse Link + SQL andmebaas (lokaalses serveris).....	25
3.2.2 Synapse Link + Synapse SQL Pool (Serverless).....	28
3.2.3 Synapse Link + Synapse Spark Pool.....	31
3.2.4 MS Fabric platvorm.....	35
4 Vaadeldavate lahenduste võrdlus ja analüüs	38
4.1 Saaty meetod.....	38
4.2 AHP samm sammult.....	38
4.2.1 Probleemi hierarhia koostamine	38
4.2.2 Andmete kogumine	39
4.2.3 Võrdlustest maatriksi loomine.....	39
4.2.4 Omavektor ja kaalu määramine.....	39
4.2.5 Konsistentsuse analüüs.....	39
4.2.6 Alternatiivide hindamine	40
4.3 AHP analüüs	40

4.3.1 Kriteeriumid ja nende suhteline olulisus	40
4.3.2 Alternatiivide võrdlus vaadeldavate kriteeriumite suhtes.	42
4.3.3 Hindamismaatriksi konsistentsuse kontroll	49
4.3.4 Tundlikkuse analüüs	51
4.3.5 Valikute analüüs kriteeriumite üleselt kokku	54
5 Analüüsi tulemused ja järeldused	56
5.1 Saadud tulemus	56
5.2 Autori poolset järeldused	57
6 Kokkuvõte	59
Kasutatud kirjandus	60
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	63

Jooniste loetelu

Joonis 1. Lähteülesande skeem.....	25
Joonis 2. Synapse Link + SQL andmebaas (lokaalses serveris) skeem.	25
Joonis 3. Synapse Link + Synapse SQL Pool Serverless skeem.	29
Joonis 4. Synapse Link + Synapse Spark Pool skeem.	32
Joonis 5. MS Fabric andmehõive skeem.	35
Joonis 6. Kriteeriumite kaalud tulpdiaagrammis.....	42
Joonis 7. Valikute võrdlus täpsuse kriteeriumi suhtes.....	43
Joonis 8. Valikute võrdlus tulevikukindluse kriteeriumi suhtes.....	44
Joonis 9. Valikute võrdlus maksumuse kriteeriumi suhtes.....	46
Joonis 10. Valikute võrdlus turvalisuse kriteeriumi suhtes.	47
Joonis 11. Valikute võrdlus mugavus kriteeriumi suhtes.....	48
Joonis 12. Valikute võrdlus kiirus kriteeriumi suhtes.	49
Joonis 13. Täpsuse kriteeriumi tundlikkuse graafik.	51
Joonis 14. Tulevikukindluse kriteeriumi tundlikkuse graafik.	52
Joonis 15. Maksumuse kriteeriumi tundlikkuse graafik.....	52
Joonis 16. Turvalisuse kriteeriumi tundlikkuse graafik.....	53
Joonis 17. Mugavuse kriteeriumi tundlikkuse graafik.....	53
Joonis 18. Kiiruse kriteeriumi tundlikkuse graafik.	54
Joonis 19. Alternatiivide arvatud sobivusjärjestus graafikul.....	55

Tabelite loetelu

Tabel 1. Synapse Link + SQL andmebaas (lokaalses serveris) hinnastus.....	27
Tabel 2. Synapse Link + Synapse SQL Pool (Serverless) hinnastus.	31
Tabel 3. Synapse Link + Synapse Spark Pool hinnastus.	34
Tabel 4. MS Fabric platvorm hinnastus.	36
Tabel 5. Saaty fundamentaalskaala [13].....	39
Tabel 6. Kriteeriumite paariviisiline võrdlus.....	41
Tabel 7. Kriteeriumite kaalud.	41
Tabel 8. Täpsuse kriteeriumi hinnangud.	42
Tabel 9. Tulevikukindluse kriteeriumi hinnangud.	43
Tabel 10. Maksumuse kriteeriumi hinnangud.	45
Tabel 11. Turvalisuse kriteeriumi hinnangud.....	46
Tabel 12. Kasutusmugavuse kriteeriumi hinnangud.	47
Tabel 13. Kiiruse kriteeriumi hinnangud.....	48
Tabel 14. Hindamismaatriks.....	49
Tabel 15. Kaaluvektor(w).....	50
Tabel 16. Maatriksi ja vektori korrutis ($A \times w$).....	50
Tabel 17. Iga rea suhete tulemus.	50
Tabel 18. Alternatiivide arvutatud sobivusjärjestus.	55

1 Sissejuhatus

Käesolevas bakalaureusetöös vaadeldav tootmisettevõtte Wolf Group OÜ on juhtiv regionaalne ehituskeemia tootmisettevõtte, mis töötab aktiivselt 5 riigis. Olles eesrindlik ettevõtte on Wolf Group OÜ endale aastate jooksul jõudnud luua üsna mastaapse ärianalüütika ökosüsteemi sisaldades mitmeid andmeallikaid ja erinevaid rakenduslike kui ka analüütilisi väljundeid. Kokku on kasutusel 84 aruannet koos alamaaruannetega. Kasutusel on automatiseeritud eelarve ja prognoosimise funktsionaalsus, finantskulude arvestamiseks kulude jagamise tööriist kui ka tootmise planeerimise ning optimeerimise rakendusid.

Ettevõtetes, eriti tootmisettevõtetes nagu Wolf Group OÜ, on ERP süsteemidel (Enterprise Resource Planning) keskne roll, kuna need integreerivad ja ühendavad erinevad ärifunktsioonid – alates tootmise planeerimisest ja varude haldamisest kuni raamatupidamise ja logistika juhtimiseni. Ärianalüütika (Business Analytics) omakorda aitab nendest süsteemidest kogutud andmeid analüüsida ja visualiseerida, andes juhtidele ülevaate operatiivsetest tulemustest ning strateegiliste otsuste toetamiseks vajaliku informatsiooni. Käesolevalt on Wolf Group kaalumas ja sooritamas esimesi samme, enda olemasoleva ERP süsteemi väljavahetamiseks uuema vastu.

Lisaks ERP süsteemidele on ka ärianalüütika muutunud hädavajalikuks tööriistaks ettevõtte juhtimisel, kus igapäevased otsused võivad märkimisväärselt mõjutada protsesside ärilisi resultate. Artiklis “Analytics 3.0” tõdeb Thomas H. Davenport: „Analüütika pole enam ainult aruandluse vahend, vaid reaajas tegutsemiseks vajalik tööriist, mis optimeerib protsesse ja suurendab otsuste täpsust[1].” Tootmisettevõtte, nagu Wolf Group OÜ, vajavad töökindlat ja paindlikku andmearhitektuuri, et tagada ärianalüütika platvormide sujuv ja järjepidev toimimine.

Käesoleva töö aluseks on Wolf Group OÜ praktiline väljakutse – olemasoleva ERP süsteemi Microsoft Dynamics Axapta 2009 vahetamine välja valitud uuema põlvkonna pilvepõhise Microsoft Dynamics 365 Finance and Operations (edaspidi D365) tarkvara vastu. ERP vahetusest tulenev vanade andmeühenduste tõstmine uute peale on peamine

uurimise ja analüüsi subjekt. Töö väljundtulemused ja -järelused ei ole ainult vaadeldavava ettevõtte põhised vaid võiksid olla enamike analoogsete kaasuste analüüsidel ja lahendamistel abiks.

1.1 Lahendatav probleem

Wolf Group OÜ seisab lähiaastatel grupis silmitsi väljakutsega viia läbi põhjalik ja strateegiline üleminek vananenud ERP süsteemilt (Microsoft Dynamics Axapta 2009) uue generatsiooni pilvepõhisele ERP süsteemile Microsoft Dynamics 365 Finance and Operations. See üleminek toob kaasa väljakutseid seoses ärikriitiliste protsesside käimas hoidmisega ülemineku perioodil ja pärast uue ERP juurutamist.

Probleem seisneb peamiselt selles, et uue D365 süsteemi kasutuselevõtt eeldab senise andmevoogude arhitektuuri täielikku ümberkujundamist. Microsofti soovitus on kasutada andmehõive esimeseks etapiks MS Dataverse'i platvormi, kuid see tähendab täiendavate lahenduste analüüsi, mis tagaksid sujuva andmete liikumise ERPist ärianalüütika platvormini.

Töö peamine eesmärk on läbi viia ärianalüütika andmearhitektuuri analüüs, pakkudes välja realistlikud ja ettevõtte vajadustega sobituvad lahendused. Kuigi analüüsi käigus käsitletakse ühe ettevõtte näidet, siis võiks lahendus anda üldisemad juhised mida saaks rakendada ka teistes tööstusettevõtetes. Töös analüüsitakse ja võrreldakse neli kõige realistlikumat andmearhitektuuri lahendust mis tulenevad ettevõtte ärinõuetest. Funktsionaalsed ja mittefunktsionaalsed nõuded on kirjeldatud edaspidiselt (vt peatükk 2.3.1.)

Töö lisa eesmärk on valideerida ja analüüsida uuemaid Microsofti andmetööriistu nagu MS Fabric platvorm.

Hüpoteesiks võib seada väite, et uuem andmearhitektuuri platvorm on vaadeldavates aspektides parem kui olemasolevad või vanemad tehnoloogiad.

1.2 Töö struktuur ja mõisted

Töö on jaotatud neljaks põhiosaks mis omakorda koosnevad töö tulemuste saavutamiseks vajalikest alampeatükkidest ja teemadest. Ärianalüütika tootmisettevõttes (vt peatükk 2)

selgitatakse tootmisettevõtte vajadusi ja üldisemat andmeliikumise tööpõhimõtteid ja protsesse. Täpsemalt analüüsitakse ärianalüütika platvormi liidestust andmeallikatega, andmetöötluse levinud etappe ja ning kehtestatakse funktsionaalsed ja mitte-funktsionaalsed nõuded. Võimalikud andmearhitektuuri lahendused (vt peatükk 3) keskenduvad erinevate vaadeldavate lahenduste analüüsile, siin on autor valinud 4 võimalikult reaalselt lahendust, mida võiks ettevõtte ka päriselt lõpplahendusena kaaluda. Vaadeldavate lahenduste võrdlus ja analüüs (vt peatükk 4) esitab põhjaliku analüüsi, kasutades Saaty meetodit, et hinnata lahenduste tugevusi ja nõrkusi vastavalt määratud kriteeriumitele. Analüüsi tulemused ja järeldused (vt peatükk 5) võtavad kokku saadud tulemused ja järeldused.

Oluline on selgitada töös keskselt kasutatud mõistet andmearhitektuur. Kuna see võib erinevate andmetöö protsesside ja tarkvaradega seonduvate toimingutega tähistada erinevat tähendust. Käesoleva bakalaureusetöö raames peab autor mõiste andmearhitektuuri all silmas omavahel ühenduses olevaid tarkvara (ja riistvara) komponente mis tagavad andmeliikumise andmeallikast kuni andmetarbijani, antud juhul ärianalüütika platvormini. Andmearhitektuuri protsess ja selle ellu viimiseks kasutatavad tarkvarakomponendid võimaldavad läbi viia nii andmete hoiustamise, töötlemise kui ka laadimise toiminguid.

2 Ärianalüütika tootmisettevõttes

Ärianalüütika on tootmisettevõttes palju enam kui pelgalt toetav funktsioon – see on sageli strateegiline ja operatsioonide tõhusust määrav tööriist. Kui traditsioonilistes ettevõtetes võib analüütika piirduda aruandluse või finantsprognosidega, siis tootmisettevõttes on sellel kriitiline roll kogu väärtusahela optimeerimisel ja juhtimisel.

Thomas H. Davenport toob artiklis „Analytics 3.0” esile, et kaasaegses analüütikas pole rõhk ainult andmete kirjeldamisel ja mineviku mõistmisel, vaid reaalses otsuste tegemisel, kus andmed muutuvad otseseks äriväärtuseks: „Analüütika tootmises on suunatud andmete muutmisele tegevuskõlblikeks teadmiseks, mis juhivad operatsioonide efektiivsust ja optimeerivad otsustusprotsesse[1].”

Tootmisettevõttes kasutatakse ärianalüütikat järgmistes põhivaldkondades:

- Tootmisprotsesside jälgimine ja optimeerimine: Reaalses andmepõhised otsused aitavad vähendada seisakuid, optimeerida seadmete kasutust ja parandada toodete kvaliteeti.
- Varude haldamine: Läbi andmepõhise analüüsi on võimalik hoida minimaalseid varude tasemeid, vältides ülevarustatust ja vähendades sellega kaasnevaid kulusid.
- Protsesside prognoosimine: Masinõppe ja tehisintellekti abil saab prognoosida seadmete hooldusvajadusi, hinnata nõudluse muutusi ja optimeerida tarneaahelat.
- Kulude ja efektiivsuse analüüs: Võimalus tuvastada raiskamist, mõõta tootlikkust ja parandada protsesside ja tööjõu kuluefektiivsust.

Viimastel aastatel on masinõppe ja tehisintellekt tootmisettevõttes muutunud lahutamatuks osaks ärianalüütikast. Need tehnoloogiad võimaldavad:

- Täpsemat prognoosimist: Näiteks saab algoritmide abil analüüsida ajaloolisi andmeid ja prognoosida hooldusvajadusi või võimalikke tõrkeid seadmetes enne nende tekkimist.

- Otsustusprotsesside toetamist: Masinõpe võib soovitada alternatiivseid lahendusi keeruliste probleemide korral, pakkudes otsustajatele valikuid, mis põhinevad suuremahulisel andmeanalüüsil.

McAfee ja Brynjolfsson rõhutavad oma artiklis „Big Data: The Management Revolution”, kuidas masinõpe ja suurandmete analüütika võimaldavad mitte ainult paremaid otsuseid, vaid ka kiiremat reageerimist reaalajas muutustele: „Suured andmed ja tehisintellekt võimaldavad ettevõtetel teha kiiremaid ja täpsemaid otsuseid, mis mõjutavad otseselt operatiivset efektiivsust ja reageerimisvõimet turul [2].”

2.1 Ärianalüütika liidestus andmeallikatega

Ärianalüütika efektiivsus ja tulemuslikkus sõltub otseselt andmete kvaliteedist ja kättesaadavusest. Tootmisettevõtte mitmekesised andmeallikad – ERP (*Enterprise Resource Planning*), CRM (*Customer Relationship Management*), MES (*Manufacturing Execution System*) ja muud süsteemid – tuleb liidestada ühtseks ja sujuvalt toimivaks andmevoogude võrgustikuks. See on tehniliselt keerukas ja mitmetahuline protsess, mis hõlmab:

- Andmeallikate kaardistamist: Millised süsteemid sisaldavad olulisi andmeid? Mis on nende täpsus, ajakohasus ja äriiline väärtus?
- Tehnilise integreerimise väljakutseid: Erinevate tarkvarade andmestruktuurid ja liidesed ehk APId (*Application Programming Interface*) nõuavad spetsialiseeritud teadmisi ja ressursse.
- Automaatika ja ajastamise kriitilisust: Andmete reaalaja lähedane või kindlalt seadistatud ja õiges järjekorras uuendamine on vajalik, et vältida andmete vananemist ja nende põhjal tehtavaid valesid otsuseid.

Käesolevast bakalaureusetöös keskendutakse ERP süsteemi liidestamisele ning vaadeldakse erinevate andmearhitektuuride valikut ühekaupa peatükis 3 ning võrreldakse ja analüüsitakse peatükis 4.

2.2 Ärianalüütika mõju äriprotsessidele tootmisettevõttes

Ärianalüütika integreerimine tootmisettevõtte igapäevasesse töösse ei ole ainult tehnoloogiline uuendus, vaid ka ärikultuuriline muutus. Tootmisettevõtted, mis omavad tugevat andmeanalüüsi võimekust, saavad kiiremini kohanduda turu ja nõudluse muutustega. Artiklis "Strength in Numbers" selgitavad E Brynjolfsson, L. Hitt, H. Kim, kuidas andmepõhiste otsuste kasutamine on seotud ettevõtete kõrgema tootlikkuse ja kasumlikkusega: „Ettevõtted, mis kasutavad andmepõhist otsuste tegemist, on keskmiselt 5% tootlikumad ja 6% kasumlikumad kui nende konkurendid[3].”

Kokkuvõttes aitab ärianalüütika tootmisettevõtetel:

- Tõsta efektiivsust ja vähendada kulusid.
- Luua läbipaistvust protsessides ja juhtimises.
- Parandada reageerimisvõimet ja konkurentsivõimet.

2.3 Andmearhitektuuri eesmärgid

Hea andmearhitektuuri eesmärk on lõppkasutajale tagada, et ettevõtte andmed on kergesti kättesaadavad, integreeritud ja usaldusväärsed. Kui andmeallikad on paika pandud, keskendub protsess sellele, kuidas need andmed optimaalselt liiguvad ja milline on vajadus neid töödelda. Protsess tegeleb andmete hõivega andmeallikatest ning andmete puhastamise ja töötlemisega, et valmistada need ette edasiseks kasutamiseks või analüüsiks, kasutades efektiivseid ja turvalisi andmehaldusprotseduure. Peamine eesmärk on luua selline arhitektuur, mis toetab andmemahtude võimalikku kasvu tulevikus ja dünaamilist analüüsikeskkonda, võimaldades kiiret ligipääsu andmetele ning vähendades käsitsi tehtava töö vajadust. Tänapäeval on järjest olulisem tagada ka turvaline ja kontrollitud ligipääs ettevõtte andmetele. Need aspektid puudutavad nii ettevõtte sisemisi protsesse, kui ka võimalike väliseid (sh küberturvalisuse) riske.

2.3.1 Funktsionaalsed ja mittefunktsionaalsed nõuded

Funktsionaalsed ja mittefunktsionaalsed nõuded on kriitilise tähtsusega andmearhitektuuri loomisel, nende eesmärk on tagada, et andmevood töötavad nõuetele vastavalt, tõhusalt ja usaldusväärselt. Funktsionaalsed nõuded keskenduvad süsteemi

võimele teostada kindlaid funktsioone ja protsesse. Mittefunktsionaalsed nõuded määravad lahenduse omadused, mis ei ole otseselt seotud kindla funktsionaalsusega, vaid mõjutavad selle tõhusust ja kasutatavust.

Funktsionaalsed nõuded määravad, kuidas lahendus peaks käituma andmeallikate andmete hõivamisel ja töötlemisel. Nõuete sõnastamine kasutusjuhtude kaudu aitab selgemalt mõista, millistele nõuetele peab lahendus vastama. Kasutusjuhud on kirjeldatud lühidalt, kõrgtaseme formaadis ja nende hulk ei ole lõplik, aga on piisavad antud töö nõuete analüüsiks. Käesoleva töö eesmärkide saavutamiseks fokuseeritakse ja keskendutakse järgmistele funktsionaalsetele nõuetele:

- **Kasutusjuht:** Kogu andmed andmeallikast

Kirjeldus: Kasutaja saab lahenduse kaudu teostada andmehõivet ühest või mitmest andmeallikast.

- **Kasutusjuht:** Jälgi andmete liikumist ja töötlemist

Kirjeldus: Kasutaja saab vajadusel kontrollida protsessi logisid ja andmete liikumist ning töötlemist andmehõive struktuuri erinevate komponentide lõikes.

- **Kasutusjuht:** Seadista andmete töötamise protsesse

Kirjeldus: Kasutaja saab andmetöötluse protsesse ja reegleid vajadusel muuta ning seadistada.

- **Kasutusjuht:** Ajasta automatiseeritud andmetöötluse tsükkel

Kirjeldus: Kasutaja saab seadistada andmete laadimise ja töötlemise tsükli ja ajakava (nt kindel aeg või ajad ööpäevas või reaalaja lähedaseks) ärinõuete täitmiseks.

- **Kasutusjuht:** Loo täiendavaid andmemudeleid

Kirjeldus: Kasutaja saab kogutud, puhastatud, töödeldud ja terviklike andmetega luua (semantilisi) andmemudeleid sisulise väljundi realiseerimiseks ja kasutamiseks ärianalüütika platvormidel.

Käesoleva töö eesmärkide saavutamiseks fokuseeritakse ja keskendutakse järgmistele mittefunktsionaalsetele nõuetele:

- Andmete täpsus ja ajakohasus: Lahendus peab tagama andmete ajakohasuse ning õigsuse.
- Kiirus ja sagedus: Andmete kättesaamine ja laadimine andmeallikast peab toimuma vajadusel sagedaselt ning kiiresti. Käesolevas töös vaadeldavatele lahendustele ei rakendata reaajajas uuenevate andmete nõuet, piisab kui andmed uuenevad kuni 60 minuti jooksul.
- Tulevikukindlus ja skaleeritavus: Lahendus peab olema skaleeritav, et toetada tulevikus suurenevat andmemahutu ja arvutusvõimsuse vajadust ning olema paindlik tehnoloogiliste uuenduste ja muutuste suhtes.
- Maksumus: Lahendus peab olema rahaliselt mõistlik ja optimeeritud. Peab vastama ettevõtte ärilikele eesmärkidele ja võimalustele.
- Andmete turvalisus: Tundlikud ärikriitilised andmed peavad olema kaitstud juurdepääsukontrolliga. Andmed peavad olema hävimisel taastavad.
- Mugavus ja kasutatavus: Lahendus peab olema intuitiivne ja lihtne võimalike muudatuste või seadistuste ellu viimiseks. Lahendus peab soovi korral võimaldama täpset seadistamist.

2.4 ETL (ELT) protseduurid

Käesoleva töö eesmärkide saavutamiseks on oluline mõista ja analüüsida andmearhitektuuris laialdaselt praktiseeritava ETL (*extract, transform, load*) ja ELT protsesside etappe ja olemust.

ETL protseduurid on andmearhitektuuri protsessi tuumik, mille eesmärk on algallikatest pärit andmed töödelda sobivale kujule näiteks andmelattu salvestamiseks, rakendusele või lõpptarbijale edasise kasutamise jaoks.

- *Extract*: Siin etapis toimub andmete kogumine kõigist asjakohastest allikatest, nagu ERP, CRM, finantssüsteemid ja muud rakendused. Siin tuleb kindlaks teha,

millised andmeväljad ja tabelid on vajalikud, ning nende struktuur tuleb võimalusel dokumenteerida.

- *Transform*: Kogutud andmed transformeeritakse, et need vastaksid ühtsele struktuurile ning formaadile. Transformatsiooni käigus tehakse andmete puhastamine (vigade eemaldamine, dubleeritud andmete kustutamine), normaliseerimine ja ühildamine, tagamaks, et andmed on standardiseeritud. Teatud lahenduste puhul realiseeritakse siin etapis ka optimeeritud andmemudel vastavalt ettevõtte äriprotsessidele. ELT protsessi puhul jäetakse etapp *Transform* vahele ja andmete teisendamine toimub lõplikku sihtkohta laadimisel.
- *Load*: Siin etapis toimub andmete laadimine lõplikku sihtkohta, olgu selleks siis andmeladu, analüütika rakendus või mõni muu tarbija. Siin määratakse ka laadimissagedus ja automaatse uuenduse meetodid, et andmeid saaks regulaarselt värskendada.

2.4.1 Medaljon andmetalletus

Käesolevas töös vaadeldakse täpsemini Microsofti pilvepõhiseid andmetöötluse arhitektuure mis talletavad andmeid suuresti samal põhimõttel ja seda nimetatakse medaljon andmetalletuse arhitektuuriks. Samad tööpõhimõtted toimuvad nii MS Azure kui MS Fabric andmeanalüütika platvormidel. ELT protsessid ja juhtimine on lahendatud Data Lakehouse (Azure) või OneLake (MS Fabric) kihis. Rakendatakse ELT järjekorda kuna on täielikult pilvepõhine andmehaldus platvorm ja selliselt on andmetöötlus optimaalsem – ei pea andmeid mitu korda salvestama ja erinevates asukohtades töötleva. Andmeid hoiustatakse medaljon tüüpi struktuuris ning seal sisalduvad andmed kategoriseeritakse kui andmete Pronks, Hõbe, Kuld tasemed.

- **Pronks tase**: Tase hõlmab andmete toomist algallikatest, mille käigus andmed salvestatakse toorsalvestusena. Andmete kvaliteet võib varieeruda ning need võivad sisaldada erinevaid vigu või puudusi. Pronks taseme peamine eesmärk on säilitada andmete originaalversioon, et hiljem oleks võimalik neid vajadusel uuesti töödelda. Data Lake andmebaasstruktuur on põhiline Pronks andmete talletamiseks mõeldud komponent pilvelahendustest [4].

- Hõbe tase: Siin toimub andmete töötlemine ja kvaliteedi tõstmine. Pronks tasemelt pärit andmed puhastatakse, normaliseeritakse ja transformeeritakse.
- Kuld tase: Kuld tasemel asuvad lõpptarbijale suunatud analüüsivõimelised andmed, mis on optimeeritud, agregeeritud ja koondatud spetsiaalsesse andmemudelitesse. Need andmed on täielikult puhastatud, rikastatud ja valmis edasiseks kasutamiseks järgmiste tarbijate sisendina. Reeglina talletatakse Kuld andmed SQL Andmebaasis [4].

2.5 Andmemudelid

Kui soovitakse käesolevas bakalaureusetöös võrrelda eeskätt erinevaid ETL (ELT) protsesside komponente on mõistlik hoida aruandevajadusi täitvad andmemudelid nendest eraldi. See annab võimaluse andmehõive allikate poolset otsa muuta nii, et ei peaks tegema suuri muudatusi ärianalüütikaks kasutatavate andmemudelite sisus, kui vahetame ETL protseduure läbiviivaid andmearhitektuuri komponente.

Tootmisettevõtte andmemudeli realiseerimisel on eelöeldu põhjal kaks võimalikku lähenemist: SQL andmebaas või Power BI Desktopis realiseeritud andmemudelid. Praktikas üks variant ei välista teist ja tavapäraselt kasutataksegi neid koos.

Arvestades tootmisettevõtte vajaduste ärikriitilist olemust ja keskmisest suuremaid andmemahte, on eelistatud valik andmemudeli ja seoste kirjeldamine tähtskeemina SQL andmebaasis. SQL andmebaasi kasutamine võimaldab paindlikkust andmete salvestamisel (sh mitmest allikast), keerukaid ja kohandatavaid päringuid indekseerimise abil ning jõudluse skaleerimist, mis on oluline ajalooliste andmete ja suurte mahtude töötlemisel. See lähenemine tagab suurema organiseerituse andmete täpsuse ja kättesaadavuse üle, mis on tootmisettevõttele kriitilise tähtsusega. Miinusena on siin kindlasti keskmisest keerulisem ja seetõttu kallim andmemudeli haldus ja seadistamine. Tõenäoliselt pole see kompetents ettevõttes olemas ja tuleb teenusepakkujalt sisse osta.

Alternatiivina või osaliselt saab andmemudeli antud kaasuses realiseerida ka Power BI Desktopi sees. Power BI Desktop pakub lihtsat ja kiiret viisi andme hõiveks kasutades Power Query funktsionaalsust. See sobib hästi väiksemate andmekogumite ja konkreetsete aruannete koostamisel. Miinusena ainult Power Query ja Power BI

Desktopiga realiseeritud andmemudel võib jääda aeglaseks suuremate andmemahtude puhul ja pole antud tervikkaasuses parim lahendus.

2.5.1 Andmemudelid MS Fabric platvormil

Antud lõputöös vaadeldakse tehnilise lahendusena ka MS Fabric rakendamist andmearhitektuuri ja andmemudeli realiseerimiseks. Seetõttu peame mõistma ka MS Fabric toimimise eripära andmemudeli realiseerimisel.

MS Fabric platvormil on andmemudeli realisatsioon korraldatud nii, et see toetab otsest ja integreeritud ühendust andmeallikatega, pakkudes paindlikku ja terviklikku andmehalduse lahendust. MS Fabric ühendab andmeallikad, töötlemise ja analüüsi üheks sujuvaks andmevooks. Platvorm kasutab selleks peatükk 2.4.1. kirjeldatud Pronks, Hõbe, Kuld tasemeid. Andmehõive, töötlemine ja laadimine on juhitud läbivalt OneLake kihis, mis on MS Fabric platvormi keskne komponent andmete halduseks ning sisemiste ja välimiste tarbijate liidestamiseks.

MS Fabric võimaldab otseühendusi algallikatega, mis tähendab, et andmed saab pidevalt ja automaatselt uuendada ja välja kutsuda. Lisaks on MS Fabric platvorm integreeritud Power BI-ga, mis tähendab, et Power BI saab kasutada MS Fabric andmemudelit ilma andmeid eraldi laadimata või salvestamata. MS Fabric platvorm püüab realiseerida järgmise generatsiooni andmehõive, töötlemise ja kuvamise põhimõtteid, kus terve andmevoog on terviklikult hallatav ühes keskkonnas mitte kokku pandud eraldi komponentidest.

3 Võimalikud andmearhitektuuri lahendused

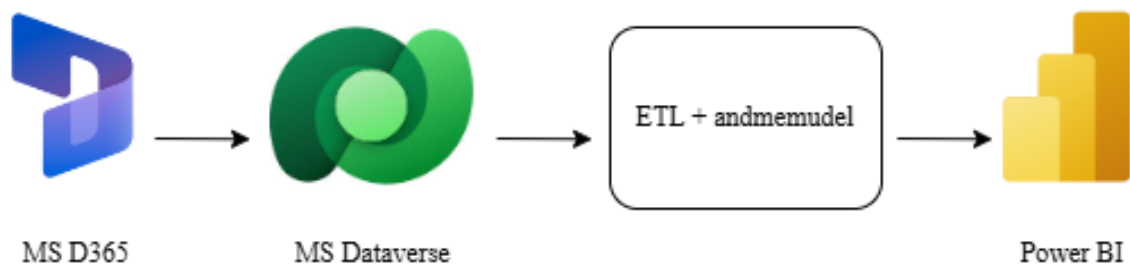
Nagu probleemipüstituses võis lugeda, siis on käesoleva bakalaureusetöö peamine eesmärk analüüsida erinevaid võimalusi kuidas lahenda ära andmete liikumine ja töötlemine ettevõtte juurutatavast majandustarkvarast (ERP) kuni andmetarbijateni milleks on antud juhul Power BI ärianalüütika platvorm.

Lisaeesmärk on analüüsida ja tutvuda Microsofti poolt hiljuti turule tulnud nõ järgmise põlvkonna andmehaldusplatvormiga MS Fabric, mis peaks paremini vastama tekkivatele tarbijate ärinõuetele ja tehnilistele vajadustele.

3.1 Lähteülesande kitsendused

Tulenevalt asjaolust, et analüüsime toimiva tootmisettevõtte äriprotsesse, siis on mõistlik rakendada käesoleva töö käsitlusele samad piirangud mis rakenduvad äriprotsessile ka tegelikkuses:

- Vaadeldav tootmisettevõtte kasutab enda igapäevaste äriprotsesside juhtimiseks peamiselt Microsofti tarkvarasid ja ettevõtte infotehnoloogia infrastruktuur koosneb suuresti Microsofti tarkvara komponentidest ja toodetest. Kuigi teoreetiliselt on võimalik lähteülesande lahenduseks kasutada ka alternatiivsete tootjate tarkvara komponente ja tooteid või luua neid ise, siis ei ole need käesoleva bakalaureusetöö skoop. Lisaks on äriline eesmärk kasutada võimalikult palju olemasolevat ja juba töötavat ärianalüütika infrastruktuuri.
- Spetsiifilisemalt on lähteülesandes defineeritud akuutne ja ärikriitiline probleem olemasoleva lokaalse ERP süsteemi Microsoft Dynamics Axapta 2009 välja vahetamine uuema põlvkonna Microsoft Dynamics 365 Finance and Operations (edaspidi: D365) ERPi vastu mis on pilveteenus. Microsoft pakub optimaalse jõudluse, integratsioonide ja turvalisuse põhjustel D365 andmehõive esimeseks komponendiks andmeplatvormi toodet MS Dataverse. Käesolev analüüs hõlmab seega andmehõivet alates Dataverse andmebaasist kuni andmete lõpptarbijani milleks on Power BI ärianalüütikaplatform (vt Joonis 1).



Joonis 1. Lähteülesande skeem.

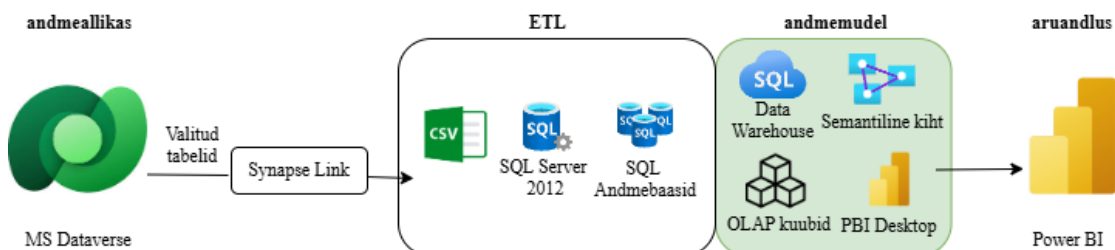
- Bakalaureusetöö lisaeesmärk on analüüsida ja uurida MS Fabric andmehaldus ja ärianalüütika platvormi võimalikku juurutamist ja kasutusele võtmist antud lähteülesande lahendamiseks. Seega võiks kõik või enamus vaadeldavaid näidisarhitektuure sisaldada osaliselt või täielikult MS Fabric platvormi komponente.

3.2 Vaadeldavad andmearhitektuuri lahendused

Tulenevalt funktsionaalsetest ja mittefunktsionaalsetest nõuetest (vt peatükk 2.3.1), lähteülesande kitsendustest (peatükk 3.1) ning võimalusest katta võimalikud erinevad tehnoloogilised kombinatsioonid ette antud raamides on valitud analüüsimiseks ja võrdlemiseks 4 võimalikku andmearhitektuuri voogu.

3.2.1 Synapse Link + SQL andmebaas (lokaalses serveris)

Esimeseks võimalikuks andmearhitektuurseks lahenduseks on luua andmevoog kasutades Synapse Linki ning olemasolevaid SQL andmebaase lokaalses SQL serveris (vt Joonis 2). Tegu on võimalikult sarnase andmearhitektuuriga mis on täna ettevõttes kasutusel.



Joonis 2. Synapse Link + SQL andmebaas (lokaalses serveris) skeem.

- MS Dataverse – Dataverse on Microsofti platvorm andmete salvestamiseks ja haldamiseks, mis võimaldab erinevatel rakendustel, nagu Dynamics 365, Power Apps ja muud Microsofti teenused, andmeid turvaliselt jagada ja hallata. Täna ei ole sisulisi alternatiive D365st andmete kätte saamiseks. MS Dataverse toetab kahte tüüpi andmeühendust edasiste andmetarbijatega. Need on Synapse Link (käesolevalt turul oleva „Export to Data Lake“ mooduli edasiarendus) ning Fabric Link.
- Synapse Link –Synapse Link võimaldab luua Dataverse'is olevad andmed CSV failidena. Lokaalne SQL Server 2012 võimaldab CSV faile laadida, kuid eeldab selleks SSIS (SQL Server Integration Services) protsesside kasutamist.
- SQL Server 2012 – Hetkel kasutusel olev lokaalne Server mis duplikeerib andmed SQL Live Serverist Logshipping tehnoloogia kaudu ja mis on liidestatud Axapta 2009 ERP süsteemiga. Axapta 2009 ei toeta uuemaid kui 2012 SQL versioone.
- SQL andmebaasid – SQL Serverisse on majutatud erinevaid andmebaase milles töödeldakse andmeid mida kasutatakse erinevate andmetarbijate jaoks andmete ettevalmistamisel, sealjuures näiteks aruandluse kui ka andmelao tarbeks. Andmeid (dimensioone ja fakte) uuendatakse öösel toimuvate andmetöötamise protsessidega. Andmed valmistatakse ette järgmistele tarbijatele: OLAP kuubid, SQL andmebaas, kui RS (*Reporting Service*) aruanded.
- OLAP kuubid – Tegu on mitmemõõtmeliste andmestruktuuridega, mis optimeerivad keerulisi andmepäringuid ja aruandlust. Praktikas enam palju ei kasutata, kuid on jäänud varasemate andmetöötamise arhitektuuride tähtsaks osaks.
- Azure SQL Data Warehouse – Relatsiooniline SQL andmebaas, optimaalne on siin luua vajalik (indekseeritud) andmestruktuur, et järgmised tarbijad saaksid siia talletatud andmed kiiresti ja väikse kuluga kätte.
- Power BI Desktop – Power BI Desktop platvormil valmistatakse tavaliselt ette semantiline andmemudel mis tähendab, et luuakse ärikasutajale kasutuskõlbulik ja arusaadav andmestruktuur. Siin korrigeeritakse ja koondatakse väljade nimetusi, luuakse andmete vahelisi mõõdikuid kasutades DAX (*Data Analysis*

Expression) valemeid, luuakse perioodi granuaalsuse moodsikuid lõppkasutamiseks aruannetes jms.

- Power BI – Power BI on ettevõttes kasutusel olev ärianalüütika tööriist, mis võimaldab andmeid visualiseerida ja luua interaktiivseid aruandeid.

Lahendus rahuldab funktsionaalsed nõuded.

Hindame vaadeldavat arhitektuuri vastavust mittefunktsionaalsetele nõuetele:

- Andmete täpsus ja ajakohasus: On tagatud ootuspärasel tasemel.
- Kiirus ja sagedus: Lahendus eeldab CSV faili töötlemist ja laadimist lokaalsesse serverisse. Samuti on kasutusel erinevad lokaalsed andmebaasid ETL protseduuride läbimiseks. Kokkuvõtvalt ei ole tegu kõige sirgjoonelisema ja kiirema lahendusega, kuid kui ärivajadus on andmete uuendamine 60 minuti jooksul, siis see katab vajadused.
- Tulevikukindlus ja skaleeritavus: Vaadeldav lahendus on kõikidest vaadeldavatest kõige vanem. Samuti on siin suureks erisuseks valdavalt lokaalsete serverite ja vanemate tehnoloogiate kasutamine mis ei ole kõige dünaamilisem ja tulevikukindlam. Kui andmemahu või andmemahude töötlemise kiirust on vaja tõsta, siis on vaja teha põhimõttelisi protsessi muudatusi ja näiteks soetada uut riistvara või litsentse. Vaadeldavatest lahendustest on tegu kõige vähem tulevikukindla ja skaleeritava arhitektuuriga.
- Maksumus: Arvestades, et on tehtud investeeringud riistvarasse ja terve ärianalüütika toimivalt üles ehitatud, siis on siin kindlasti maksumuse eelis ülejäänud lahenduste ees. Teiselt poolt on seadistamise ja haldamise osas kindlasti väljakutseid mis eeldavad valdkonna tippspetsialisti kaasatust ja (tasulist) täiendavat ajakulu. SQL Server 2012 omab ka tarkvara litsentsile baseeruvat kulu. Samuti peab arvestama kulusid elektrile ja jahutusele.

Tabel 1. Synapse Link + SQL andmebaas (lokaalses serveris) hinnastus.

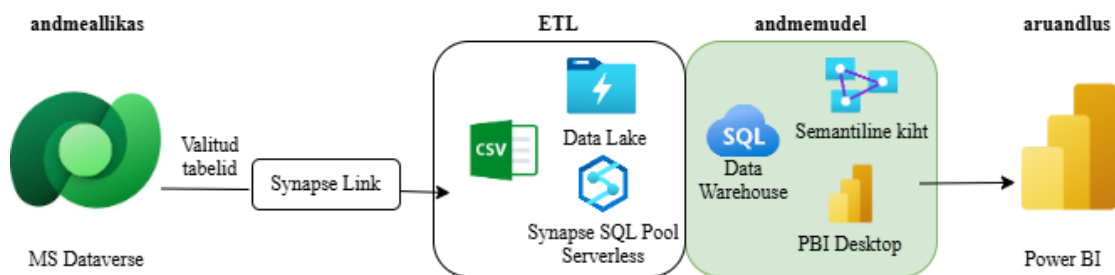
Üksus	Kuluühik	Tarbimine	Igakuine kulu (€)
SSIS või analoog[5]	Arvutusvõimsus	400h	320,0

	0,8€/h		
Elekter/Jahutus	KWh	100 x 0,2€	50,0
Hooldus	40€/h	0,5h	20,0
SQL Server 2012[6]	Oleneb litsentsist	Oleneb litsentsist	200,0
PBI litsentsid	PBI Pro litsents / 9,5€	Kasutaja kohta (5)	47,5
Lahenduse juurutamine	80 spetsialisti töötundi	80h x 60€ / 60 kuud	80,0
Kokku:			717,5

- Andmete turvalisus: Serveritele on võimalik kehtestada heal tasemel ligipääsuõigused ja kasutajapiirangud. Siiski on see tehnilisest kompetentsist ja oskustest. Hea tasemega turvalisuse protseduuri ülesehitamine on kallim ja võib kahjustada kasutusmugavust. Andmete säilimine on samuti suur risk, kuigi andmed on duplikeeritud mitmes serveris ja andmebaasis, siis ei ole need kaitstud füüsiliste kahjustuste (nt tulekahju) eest võrreldes pilvelahendustega.
- Mugavus ja kasutatavus: Kui palgal on oskuslikud spetsialistid ei ole antud lahenduse kasutamine ja haldus otseselt ebamugav, aga on siiski keerulisem võrreldes uuemate tsentraliseeritud keskkondadega süsteemidega. Aruandluse kasutajad PBI poolal ei oma suurt kokkupuudet andmehõivega ja saavad nautida kaasaegset kasutuskogemust analüütika platvormi kasutamisel. Üldjoontes on tegu küllaltki vana (üle 10 aasta vanuse) lahendusega, mis ei vasta enam tänapäevastele kasutusmugavuse ootustele.

3.2.2 Synapse Link + Synapse SQL Pool (Serverless)

Teiseks võimalikuks andmearhitektuurseks lahenduseks on luua andmevoog kasutades Synapse Linki ning Synapse SQL Pool (Serverless) pilvepõhist andmetöötlus komponenti (vt Joonis 3). Tegemine on pilvelahendusega, mis kasutab mõningaid samasid ETL komponente MS Fabric platvormiga, kuid mitte põhilist funktsionaalsust. Kasutuses on peamiselt Azure Synapse komponendid.



Joonis 3. Synapse Link + Synapse SQL Pool Serverless skeem.

Joonisel 3 kujutatud andmevoo komponentide detailsem kirjeldus on järgmine:

- Microsoft Dataverse – Dataverse on Microsofti platvorm andmete salvestamiseks ja haldamiseks, mis võimaldab erinevatel rakendustel, nagu Dynamics 365, Power Apps ja muud Microsofti teenused, andmeid turvaliselt jagada ja hallata.
- Synapse Link – Synapse Link võimaldab reaalajas sünkroonida Dataverse'is olevad andmed CSV failidena otse Data Lake andmebaasiga, sealjuures on oluline, et Synapse Link moodulis on võimalik seadistada ka ainult vajaminevate tabelite ületoomine. See on oluline andmearhitektuuri kogumaksumust ja efektiivsust silmas pidades.
- Azure Data Lake – Andmejärv, siia talletatakse CSVdes olevad andmed edasiseks töötlemiseks. See toimib sageli ülejäänud andmevoo komponentide alusstruktuurina. Azure Data Lake on üks populaarsemaid Microsofti pilvepõhiseid andmebaase struktureerimata või vähe struktureeritud andmete hoiustamiseks. Sellest tulenevalt on ka jooksvad kasutuspõhised kulud optimaalsed.
- Azure Synapse SQL Pool Serverless [7] – Siin toimub põhiline osa ETL protseduuride juhtimisest. Microsofti pilvepõhine andmebaas, mis võimaldab hallata ja töödelda suuremahulisi andmekogumeid. Tema eelis on suhteline odavus ja efektiivsus toorandmete töötlemisel, maksta tuleb kasutuspõhiselt.

- Azure SQL Data Warehouse – Relatsiooniline SQL andmebaas, optimaalne on siin luua vajalik (indekseeritud) andmestruktuur, et järgmised tarbijad saaksid siia talletatud andmed kiiresti ja väikse kuluga kätte.
- Power BI Desktop – Power BI Desktop platvormil valmistatakse tavaliselt ette semantiline andmemudel mis tähendab, et luuakse ärikasutajale kasutuskõlbulik ja arusaadav andmestruktuur. Siin korrigeeritakse ja koondatakse väljade nimetusi, luuakse andmete vahelisi moodsikuid kasutades DAX (*Data Analysis Expression*) valemeid, luuakse perioodi granaalsuse moodsikuid lõppkasutamiseks aruannetes jms.
- Power BI – Power BI on ettevõttes kasutusel olev ärianalüütika tööriist, mis võimaldab andmeid visualiseerida ja luua interaktiivseid aruandeid.

Lahendus rahuldab funktsionaalsed nõuded.

Hindame vaadeldavat arhitektuuri vastavust mittefunktsionaalsetele nõuetele:

- Andmete täpsus ja ajakohasus: On tagatud ootuspärasel tasemel. Synapse Linkiga on võimalik hõivata ainult vajaminevaid andmetabeleid mis on optimaalne ja täpne lahendus.
- Kiirus ja sagedus: Lahendusel on vaadeldutest üks kõige kiirema uuendamisesagedusega tsüklid, sest keerulisi andmetransformatsioone ei läbita. Terve tsükli läbikäimiseks kulub eelduslikult 10 minutit.
- Tulevikukindlus ja skaleeritavus: Vaadeldav andmearhitektuur ei kasuta mõningaid samasid komponente mis on kasutusel MS Fabric puhul, kuid puudu on keskne MS Onelake funktsionaalsus. Lahendus on pilvepõhine ja sellegipoolest võrdlemisi kaasaegne. Struktuur toetab ka andmekoguste suurenemist.
- Maksumus: Avalike allikate põhjal saame tuletada vaadeldava andmearhitektuuri käimas hoidmisele kuluva jooksva maksumuse vastavalt

Tabel 2. Synapse Link + Synapse SQL Pool (Serverless) hinnastus.

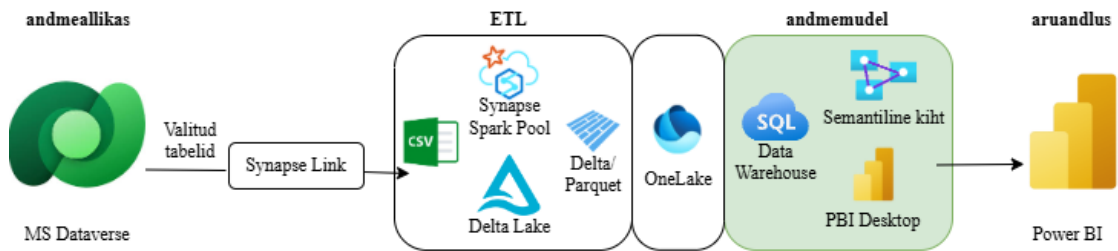
Üksus	Kuluühik	Tarbimine	Igakuine kulu (€)
Data Lake	0.15€ / GB kuus	200 GB salvestusruumi	30,0
Azure Synapse Serverless[8]	5€ / 1 TB protsessitud	20 korda 200 GB = 4 TB	20,0
PBI litsentsid	PBI Pro litsents / 9,5€	Kasutaja kohta (5)	47,5
Lahenduse juurutamine	160 spetsialisti töötundi	160 x 60€/ 60 kuud	160,0
Kokku			257,5

Negatiivse poole pealt koosneb lahendus väga mitmest komponendist mis on vaja alguses juurutada, seadistada ja testida, see on valdkonna tippspetsialisti töö ja tänaste parimate teadmiste pealt kulub selle juurutamiseks vähemalt 160 arendaja töötundi.

- Andmete turvalisus: Terve vaadeldav arhitektuur on Microsofti andmeinfrastruktuuri pilvepõhistest komponentidest ja heal kui mitte väga heal tasemel turvaline. Kasutajaõigused ning nende ligipääsud on hallatud domeenipõhises Azure Active Directory keskkonnas. Siiski koosneb vaadeldav lahendus väga mitmest erinevast komponendist mis pole turvalisuse seisukohast optimaalne. Andmete tagavara koopiade olemasolu on tagatud MS teenuses.
- Mugavus ja kasutatavus: Pärast lahenduse juurutamist on see lõppkasutajale üsna mugav, eeldab siiski piisavat IT (infotehnoloogia) kompetentsi, et võimalikele probleemidele jälile saada ja neid parandada. Kui tekib vajadus andmeid lisada, tuleb nende laadimine käsitsi seadistada (uued tabelid ja väljad ERP süsteemist).

3.2.3 Synapse Link + Synapse Spark Pool

Kolmandaks võimalikuks andmearhitektuurseks lahenduseks on luua andmevoog kasutades Synapse Linki ning CSV töötlemiseks kasutada Spark töövihiid, et luua Delta/Parquet failid (Joonis 4). See lahendus hõlmab nii Azure Synapse kui MS Fabric komponente.



Joonis 4. Synapse Link + Synapse Spark Pool skeem.

Joonisel 4 kujutatud andmevoo komponentide detailsem kirjeldus on järgmine:

- Microsoft Dataverse – Dataverse on Microsofti platvorm andmete salvestamiseks ja haldamiseks, mis võimaldab erinevatel rakendustel, nagu Dynamics 365, Power Apps ja muud Microsofti teenused, andmeid turvaliselt jagada ja hallata.
- Synapse Link –Synapse Link võimaldab reaalajas sünkroonida Dataverse'is olevad andmed CSV failidena otse Data Lake andmebaasiga, sealjuures on oluline, et Synapse Link moodulis on võimalik seadistada ka ainult vajaminevate tabelite ületoomine. See on oluline andmearhitektuuri kogumaksumust ja efektiivsust silmas pidades.
- Azure Data Lake – Andmejärv, siia talletatakse CSVdes olevad andmed edasiseks töötlemiseks.
- Synapse Spark Pool – Spark või Databricks tööviikutega konverteeritakse CSV failid Delta/Parquet formaati.
- Azure Delta Lake [9] – On Azure Data Lake peale loodud andmebaasi kiht mis kõrvaldab mitmed Data Lake puudused. Azure Delta Lake toetab ACID (*atomicity, consistency, isolation and durability*) transaktsioone, päringute küsimist valitud ajahetkel (ajas rändamine) ja muid optimeeritud andmetöötamise funktsioone. *Delta Lake* kasutab andmete hoiustamiseks Parquet andmefaile ja logifaile andmehalduse metainfo ja muudatuste ajaloo jälgimiseks (Delta).
- MS Onelake – OneLake on keskne osa MS Fabric funktsionaalsusest. OneLake on integratsioonikiht andmete liidestamiseks väliste ja sisemiste tarbijate jaoks. OneLake kasutamine antud lahenduses on võimalik ka ära jätta, kuid siis ei ole ka

vajadust tekitada Delta/Parquet faile, ning asjalikum oleks kasutada eelmist andmearhitektuuri lahendust. See lahendus on siiski mõistlik kui on soov tulevikus täielikult MS Fabric andmeplatvormile üle minna.

- Azure SQL Data Warehouse – Relatsiooniline SQL-andmebaas, optimaalne on siin luua vajalik (indekseeritud) andmestruktuur, et järgmised tarbijad saaksid siia talletatud andmed kiiresti ja väikse kuluga kätte.
- Power BI Desktop – Power BI Desktop platvormil valmistatakse tavaliselt ette semantiline andmemudel aruandluse kasutamiseks.
- Power BI – Power BI on ettevõttes kasutusel olev ärianalüütika tööriist, mis võimaldab andmeid visualiseerida ja luua interaktiivseid aruandeid.

Lahendus rahuldab funktsionaalsed nõuded.

Hindame vaadeldavat arhitektuuri vastavust mittefunktsionaalsetele nõutele:

- Andmete täpsus ja ajakohasus: On tagatud ootuspärasel tasemel. Synapse Linkiga on võimalik hõivata ainult vajaminevaid andmetabeleid mis on optimaalne ja täpne lahendus.
- Kiirus ja sagedus: Lahendus on hea kiirusega, kuid natuke aeglasem kui variant 2. Nimelt peab lahendus läbi tegema Spark töövihikutega CSV failide töötluste mis on lisasamm. Andmete uuendamise kiirus D365st PBIsse on hinnanguliselt 60 minutilise intervalliga.
- Tulevikukindlus ja skaleeritavus: Vaadeldav andmearhitektuur on sobiv kui soovitakse tulevikukindlat ja skaleeritavat lahendust, kuid puudub veel äriplane vajadus täielikult MS Fabricu juurutamiseks.
- Maksumus: Võttes aluseks eelmise lahenduse arvutus tuleb juurde lisada ka ressursid Spark Pooli maksumus ja MS Fabric litsentsi soetamine.

Tabel 3. Synapse Link + Synapse Spark Pool hinnastus.

Üksus	Kuluühik	Tarbimine	Igakuine kulu (€)
Data Lake	0.15€ / GB kuus	200 GB salvestusruumi	30,0
Azure Synapse	5€ / 1 TB	20 korda 200 GB = 4 TB	20,0
Spark Pool	Arvutusvõimsus keskmiselt 48€/h (48 tuuma)	Oleneb sagedusest	300,0
MS Fabric[10]	Arvutusvõimsus Alguses F4 0,9 €/h	Oleneb sagedusest	300,0
PBI litsentsid	PBI Pro litsents / 9,5€	Kasutaja kohta (5)	47,5
Lahenduse juurutamine	160 spetsialisti töötundi	160 x 60 €/ 60 kuud	160,0
Kokku			857,0

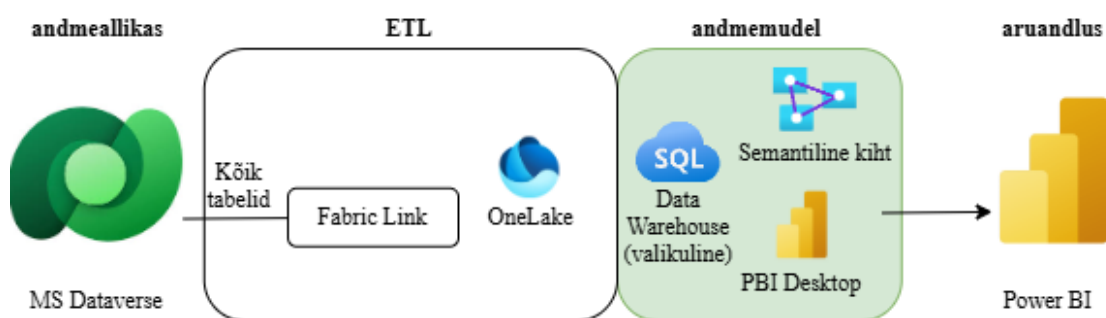
Lisaks koosneb lahendus väga mitmest komponendist mis on vaja alguses juurutada, seadistada ja testida, see on valdkonna tippspetsialisti töö ja tänaste parimate teadmiste pealt kulub selle juurutamiseks vähemalt 160 arendaja töötundi.

- Andmete turvalisus: Terve vaadeldav arhitektuur on Microsofti andmeinfrastruktuuri pilvepõhistest komponentidest ja heal kui mitte väga heal tasemel turvaline. Kasutajaõigused ning nende ligipääsud on hallatud domeenipõhises Azure Active Directory keskkonnas. Siiski koosneb vaadeldav lahendus väga mitmest erinevast komponendist mis pole turvalisuse seisukohast optimaalne. Andmete tagavara koopiade olemasolu on tagatud MS teenuses.
- Mugavus ja kasutatavus: Pärast lahenduse juurutamist on see lõppkasutajale üsna mugav, eeldab siiski piisavat IT (infotehnoloogia) kompetentsi, et võimalikele probleemidele jälile saada ja neid parandada. Kui tekib vajadus andmeid lisada, tuleb nende laadimine käsitsi seadistada (uued tabelid ja väljad ERP süsteemist). Lahenduses on kasutusel võetud OneLake andmehalduskiht mis funktsioneerib

kui SQL andmeühenduspunkt e. *Endpoint* ja tagab tulevikus hõlpsad integratsioonid võimalike lisanduvate andmetarbijatega.

3.2.4 MS Fabric platvorm

Neljandaks võimalikuks andmearhitektuurseks lahenduseks on luua andmevoog kasutades terve MS Fabric platvormi funktsionaalsust.



Joonis 5. MS Fabric andmehõive skeem.

Joonis 5 kujutatud andmevoo komponentide detailsem kirjeldus on järgmine:

- Microsoft Dataverse – Dataverse on Microsofti platvorm andmete salvestamiseks ja haldamiseks. Antud andmearhitektuuri lahenduses kasutatakse Dataverse andmebaasi ka Parquet failide hoiustamiseks.
- Fabric Link [11] – Toob üle kõik D365 andmetabelid ja teeb vajalikud konversioonid ning talletab andmed Dual-write tehnoloogia abil Dataverse andmebaasi. Protsess on täielikult Microsofti poolt hallatud ja kasutajale nähtamatu.
- MS Onelake – OneLake on keskne integratsioonikiht andmete liidestamiseks välise ja sisemiste tarbijate jaoks.
- Azure SQL Data Warehouse – Relatsiooniline SQL-andmebaas, optimaalne on siin luua vajalik (indekseeritud) andmestruktuur, et järgmised tarbijad saaksid siia talletatud andmed kiiresti ja väikse kuluga kätte. Antud lahenduses on tegu valikulise komponendiga, sest Power BI võib väiksemate andmehulkade puhul

pöörduda ka otse aruandekõlbulike alusandmete poole ja sellisel juhul andmeid täiendavalt materialiseerida ei ole optimaalne.

- Power BI Desktop – Power BI Desktop platvormil valmistatakse tavaliselt ette semantiline andmemudel aruandluse kasutamiseks.
- Power BI – Power BI on ettevõttes kasutusel olev ärianalüütika tööriist, mis võimaldab andmeid visualiseerida ja luua interaktiivseid aruandeid.

Lahendus rahuldab funktsionaalsed nõuded.

Hindame vaadeldavat arhitektuuri vastavust mittefunktsionaalsetele nõutele:

- Andmete täpsus ja ajakohasus: On tagatud ootuspärasel tasemel.
- Kiirus ja sagedus: MS Fabricu näol on tegu kiire platvormiga mis on loodud suurteks ja reaaliajapõhisteks andmete liigutamiseks, sealjuures on peetud silmas masinõppe rakendustele vajaminevaid andmekiiruse parameetreid. Toote külge on võimalik rakendada komponente nagu Data Activator mis registreerib kiiresti muutunud andmetele ning uuendab kirjeid PBI aruandluses.
- Tulevikukindlus ja skaleeritavus: Tegum on kindlasti kõige tulevikukindlama ja skaleeritavama lahendusega. Tegum on suurema andmearhitektuuri ökosüsteemiga millesse ka MS viimastel aastatel on panustanud ja seda tööd kindlasti jätkab.
- Maksumus: Tegum on täna kindlasti vaadeldavatest valikutest kõige kallima tootega. Hinnastusloogika on küllaltki erinev teistest variantidest, sest platvorm automatiseerib ära palju käsitööd.

Tabel 4. MS Fabric platvorm hinnastus.

Üksus	Kuluühik	Tarbimine	Igakuine kulu (€)
Dataverse andmebaas[12]	40€/GB/Kuus	On kaasas D365 litsentsiga, võib vajada lisamahtu	1200,0
MS Fabric	Arvutusvõimsus Alguses F4 0,9 €/minut	Oleneb sagedusest	650,0

PBI litsentsid	PBI Pro litsents / 9,5€	Kasutaja kohta (5)	47,5
Lahenduse juurutamine	20 spetsialisti töötundi	20 x 60€ / 60 kuud	20,0
Kokku			1917,5

Positiivse poole pealt võib eeldada, kiiret, odavamast ja lihtsamast toote kasutuselevõtmist, seetõttu ei tohiks kuluda palju tippspetsialisti tööaega.

- Andmete turvalisus: Terviklik platvorm mis täidab kõige uuemaid turvanõudeid.
- Mugavus ja kasutatavus: Platvorm on selgelt kõige mugavam ja hõlpsam kasutamiseks ja juurutamiseks. Palju seadistamisele ja andmehaldusele kuluvat tööd on MS automatiseerinud ja tootesse sisse ehitatud. Platvorm võimaldab selle kasutusel võtta ilma ühtegi rida koodi kirjutamata selleks välja arendatud keskkonnas. Uute komponentide lisamine ja liidestamine on samuti läbi MS Onelake kihi väga lihtsaks tehtud.

4 Vaadeldavate lahenduste võrdlus ja analüüs

Bakalaureusetöö eesmärk on analüüsida ja võrrelda vaadeldavaid andmearhitektuuri lahendusi, kuid vältimaks subjektiivse arvamuse põhist analüüsi on oluline toetuda metoodilisele ja teaduslikule võrdlemise protsessile. Selleks sobib hästi Saaty meetod ehk AHP (analüütiliste hierarhiate meetod).

4.1 Saaty meetod

Saaty meetod ehk AHP põhineb kogemusel, mille omandas selle arendaja T.L. Saaty. See töötati välja vastusena avastusele, et keerukate otsuste tegemiseks on märkimisväärne puudus lihtsasti mõistetavast ja hõlpsasti rakendatavast meetodist. Saaty meetod põhineb (otsustajate poolt) objektide paarikaupa võrdlemisel. [13][14].

4.2 AHP samm sammult

AHP rakendamiseks tuleb esmalt defineerida probleem, paika panna eesmärk, kriteeriumid (mõjurid) ja alternatiivid (valikud). Nendest moodustub hierarhia e puu, kus juureks on eesmärk, selle all on kriteeriumid ning lehtedeks on alternatiivid. Kriteeriumeid tuleb võrrelda paarikaupa, andes iga paari puhul hinnangu Saaty fundamentaalskaalal. Pärast mõjurite hindamist ja nende suhtelise olulisuse leidmist peab iga mõjuri korral kõiki alternatiive samuti omavahel paarikaupa võrdlema. Nii saadakse teada lõppotsus. Ka alternatiividele tuleb Saaty fundamentaalskaalal hinnangud anda [13]. AHP protsess eeldab ka tundlikkuse kontrollimist.

4.2.1 Probleemi hierarhia koostamine

Probleem jagatakse eesmärgi, kriteeriumite, alamkriteeriumite ja alternatiivide hierarhiaks. See on loov ja otsustusprotsessi võtmetähtsusega osa. Hierarhia struktureerimine on analüüsi alustala. See näitab, kuidas ühe taseme elemendid on seotud alumise tasemega. Hierarhia edastab seoseid igal tasandil kuni lehtsõlmedeni (alternatiivide võrdlus) [14].

4.2.2 Andmete kogumine

Ekspertidelt kogutakse andmed alternatiivide paarivõrdluste kohta kvalitatiivses skaalas. Skaala (vt Tabel 5) sisaldab võrdlusi nagu: võrdtähtis, mõõdukas paremus, oluline paremus jne. Võrdlused teisendatakse numbrilisteks väärtusteks [14].

Tabel 5. Saaty fundamentaalskaala [13].

Intensiivsus	Definitsioon	Selgitus
1	Võrdtähtis	Võrreldavad pole mõjus eristatavad
3	Mõõdukas paremus	Üks kriteerium on teisest eristatavam.
5	Oluline paremus	Ühel kriteeriumil on teisega võrreldes tugev eelistus.
7	Väga tugev paremus	Ühel kriteeriumil on teisega võrreldes väga tugev eelistus.
9	Ekstreemne paremus	Tugevaim võimalik paremus või eelistus.

2, 4, 6, 8 on kahe kõrvutise otsustuse vahelised hinnangud.

4.2.3 Võrdlustest maatriksi loomine

Võrdlused organiseeritakse ruudukujulisesse maatriksisse, kus diagonaalelemendid on väärtusega 1. Kui element „i“ on olulisem kui element „j“, siis maatriksi vastav element on >1 . Vastupidises suunas on väärtus vastastikune (näiteks $1/x$) [14].

4.2.4 Omavektor ja kaalu määramine

Omavektor ja normeeritud kaaluvektor määravad kriteeriumide suhtelise tähtsuse. Vektorite elemendid tähistavad kaalusid vastavalt alternatiividele [14].

4.2.5 Konsistentsuse analüüs

Konsistentsus indeks CI (*Consistency Index*) arvutatakse valemiga:

$$CI = \frac{(\lambda_{\max} - n)}{(n-1)}$$

λ_{\max} on hindamismaatriksi suurim omaväärtus ja n on maatriksi suurus. CI võrdlemine juhusliku indeksi RI (*Random Index*) väärtusega CI/RI, annab konsistentsisuhetaru CR (*Consistency Ratio*). $CR < 0,1$ näitab head konsistentsust [14].

4.2.6 Alternatiivide hindamine

Iga alternatiivi hinnang saadakse, korrutades alternatiivi ja alamkriteeriumide kaalu. Tulemused summeeritakse, et saada üldine reiting.

AHP võimaldab määrata igale alternatiivile kaalu, lähtudes nende suhtelisest tähtsusest ühise kriteeriumi suhtes [14].

Käesolevas töös kasutatakse AHP võrdluse teostamiseks tasuta tarkvara Priority Estimation Tool [15]. Vaja selleks oli vaja tutvuda ka juhendmaterjalidega [16].

4.3 AHP analüüs

Käesolevas peatükis määratakse kriteeriumid (mõjurid) ja võrreldakse alternatiive (valikuid). Seejärel hinnatakse alternatiive kriteeriumite lõikes ja hinnatakse nende tundlikust.

4.3.1 Kriteeriumid ja nende suhteline olulisus

Kriteeriumid tulenevad peaaesjalikult funktsionaalsetest ja mittefunktsionaalsetest nõuetest (Peatükk 2.3.1). Kõik vaadeldavad alternatiivid täidavad esitatud funktsionaalsed nõuded, seetõttu keskendutakse AHP analüüsis mittefunktsionaalsete nõuete võrdlemisele. Kriteeriumid on vastavalt:

- Andmete täpsus ja ajakohasus
- Tulevikukindlus ja skaleeritavus
- Maksumus
- Kiirus ja sagedus
- Andmete turvalisus
- Mugavus ja kasutatavus

Kriteeriumite paariviisilise võrdlemise tulemused on tabelis (vt. Tabel 6) loodud autori parimate hinnangute kui ka vaadeldava tootmisettevõtte vastutava töötaja sisendile ning baseeruvalt järgmised:

Tabel 6. Kriteeriumite paariviisiline võrdlus.

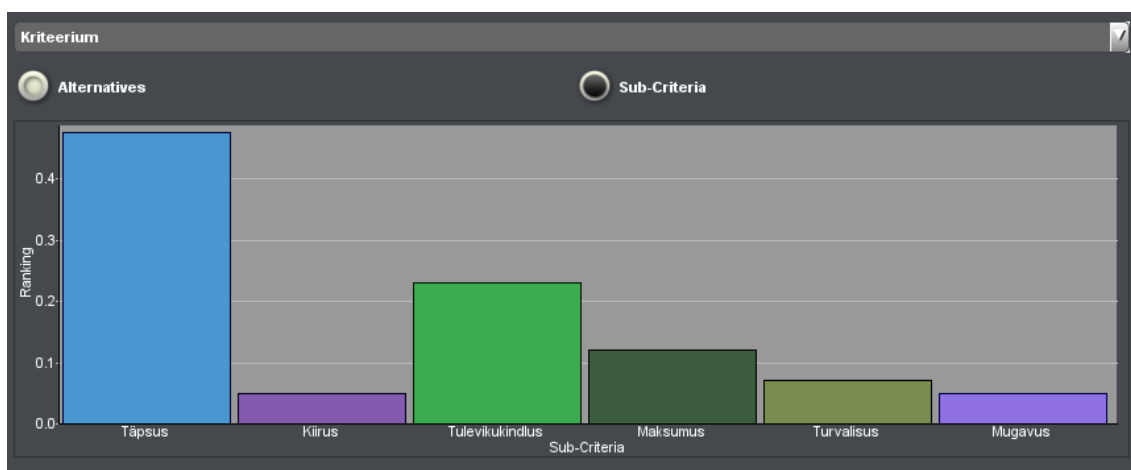
	Täpsus	Kiirus	Tuleviku-kindlus	Maksumus	Turvalisus	Mugavus
Täpsus	1,00	5,00	5,00	5,00	5,00	5,00
Kiirus	0,20	1,00	0,25	0,33	0,50	1,00
Tuleviku-kindlus	0,20	4,00	1,00	4,00	4,00	4,00
Maksumus	0,20	3,00	0,25	1,00	3,00	3,00
Turvalisus	0,20	2,00	0,25	0,33	1,00	2,00
Mugavus	0,20	1,00	0,25	0,33	0,5	1,00

Tulemusena saadi järgnevad kriteeriumite olulisuse kaalud (vt Tabel 7).

Tabel 7. Kriteeriumite kaalud.

Kriteeriumid	Kaalud
Täpsus ja ajakohasus	0,475
Tulevikukindlus ja skaleeritavus	0,230
Maksumus	0,122
Turvalisus	0,072
Mugavus ja kasutatavus	0,051
Kiirus ja sagedus	0,051

Graafiliselt on kaalud esitatud tulpdiaagrammina (vt Joonis 6) all.



Joonis 6. Kriteeriumite kaalud tulpdiaagrammis.

Kõige olulisemaks peetakse andmete täpsust, seejärel andmearhitektuuri lahenduse tulevikukindlust. Edasi vastavalt maksumust, turvalisust. Kasutusmugavust ja andmete uuendamise kiirust hindame võrselt, kuid madala prioriteetsusega. See vastab ka vaadeldava toomissettevõtte ärinõuetele ja soovidele.

4.3.2 Alternatiivide võrdlus vaadeldavate kriteeriumite suhtes.

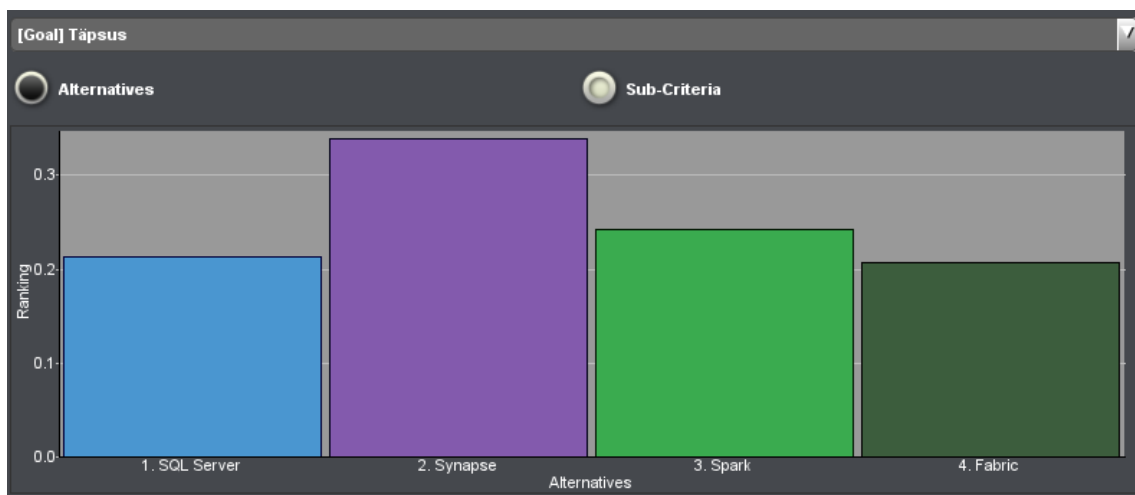
Kõik vaadeldavad kriteeriumid (mõjurid) tulenevad mittefunktsionaalsetest nõuetest ja on pikemalt selgitatud peatükis 2.3.1.

Täpsus ja andmete ajakohasus on kõige kaalukam kriteerium. Andmed peavad olema korrektsed, terviklikud ja liikuma ettenähtud aja jooksul läbi andmearhitektuuri lahenduse.

Tabel 8. Täpsuse kriteeriumi hinnangud.

Täpsus	SQL Server	Synapse	Spark	Fabric
1. SQL Server	1,00	0,50	0,50	2,00
2. Synapse	2,00	1,00	2,00	1,00
3. Spark	2,00	0,50	1,00	1,00
4. Fabric	0,50	1,00	1,00	1,00

Nagu on vaadeldav (vt Tabel 8), siis lahenduse 2. Synapse ja 3. Spark puhul võib täheldada suuremat täpsust, siiski ei ole vahed 4 alternatiivi osas märkimisväärsed. Täpsusel hinnati andmete valimise ja protsessimise paindlikust ja andmete uuendamiseks mõeldud protseduuride seadistamise suhtelist lihtsust. Lahendus 1. SQL Server on seotud lokaalse serveri ja mitmete killustunud andmeladude kasutamisega, ei ole kõige sirgjoonelisem ja lihtsasti arusaadavam lahendus. Lahendus 4. mis rakendab täielikult MS Fabric platvormi funktsionaalsust ei saa samuti kõige täpsemaks pidada, sest on kasutaja eest osaliselt varjatud ja andmehõive on automatiseeritud, samas ei ole lahendus ehk kõige kuluefektiivsem ja optimaalseim ressursse silmas pidades. Tabel 8 hinnangud on visuliseeritud tulpdiaagrammina (vt Joonis 7) all.



Joonis 7. Valikute võrdlus täpsuse kriteeriumi suhtes.

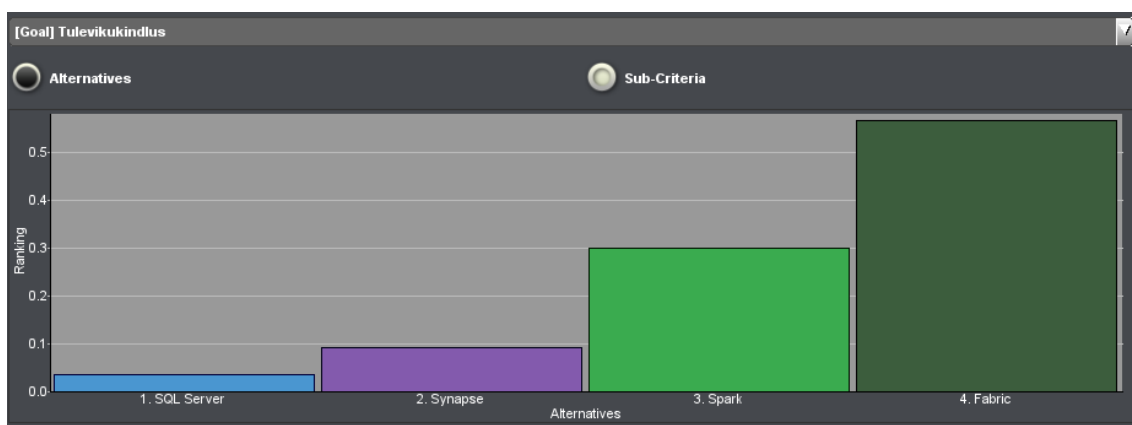
Tulevikukindlus on järjekorras tähtsuselt 2 kriteerium. On oluline, et täna loodav lahendus ei jääks mõne aastaga ajale jalgu. Seda eriti silmas pidades vaadeldava äriettevõtte plaane loodavat lahendust tulevikus teistes grupi ettevõtetes juurutada ja andmekasutust ühildada.

Tabel 9. Tulevikukindluse kriteeriumi hinnangud.

Tulevikukindlus	SQL Server	Synapse	Spark	Fabric
1. SQL Server	1,00	0,20	0,143	0,11
2. Synapse	5,00	1,00	0,167	0,14

3. Spark	7,00	6,0	1,00	0,33
4. Fabric	9,00	7,0	3,0	1,00

Nagu tabeli (vt Tabel 9) Siin on hinnangud üsna sirgjooneliselst uuemalt lahendusel vanema poole. Lahendus 4. MS Fabric platvorm on toode mida MS arendab edasi veel mitmeid aastaid, samas kui mõned vanemad lokaalsete serverite tooted ja litsentsid järkjärgult kasutamisest eemaldatakse. Järjekorras teise parima tulemuse antud kriteeriumi lõikes on 3. Spark Pool töövihikutega andmeid töötlev lahendus, mis sobib eelduslikult hilisemas etapis hõlpsaks ja kiireks MS Fabric platvormi kasutuselevõtuks, sest sisaldab juba selle komponente. Vaadeldava tootmisettevõtte kontekstis tuleb mõtestada ka tulevikuarenguid ja tekkivaid vajadusi. Hetkel töötab ettevõtte 5s riigis üle maailma ja üha olulisem on ühildada andmebaase ja äriprotsesse, vältida geograafilist andmete silostumist ja eriilmsete äritarkvarade paljusust. Vanemad lokaalsed süsteemid on seetõttu hääbumas. Tabel 9 hinnangud on visuliseeritud tulpdiaagrammina (vt Joonis 8) all.



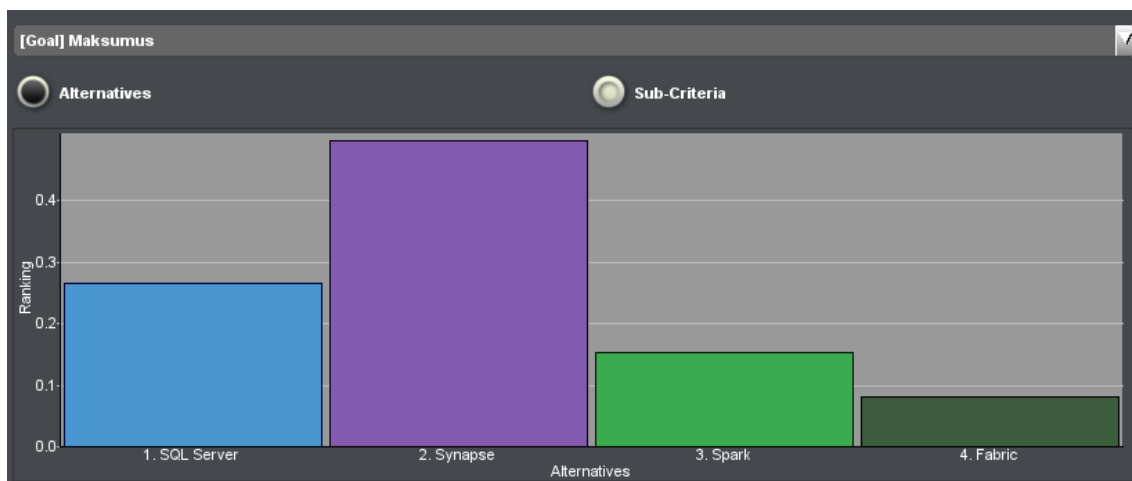
Joonis 8. Valikute võrdlus tulevikukindluse kriteeriumi suhtes.

Maksumus on tähtsuselt kolmas kriteerium. Siin on oluline mõista, et esialgu kallimana tunduv lahendus võib tuleviku perspektiivi ja tekkivaid ärinõudeid ning vajadusi silmas pidades osutada hoopis optimaalsemaks ja soodsamaks lahenduseks. Käesolevalt esitatud hinnangud maksumuse suhtes on siiski lihtsustatud ja selliseid hüpoteetilisi muutusi ei arvestata – lähtutakse arvutatud kuutasudest.

Tabel 10. Maksumuse kriteeriumi hinnangud.

Maksumus	SQL Server	Synapse	Spark	Fabric
1. SQL Server	1,00	0,50	2,00	3,00
2. Synapse	2,00	1,00	3,00	6,00
3. Spark	0,50	0,33	1,00	2,00
4. Fabric	0,33	0,17	0,50	1,00

Nagu ülalolevast tabelis (vt Tabel 10) on näha, siis maksumus on üsna sirgjooneline võrdlus ja iga andmearhitektuuri lahenduse juures peatükk 3.2 on autor teinud ka vastavad võrdlustabelid. Kuigi täpseid andmemahte ja hindu ei ole võimalik täiuslikult määrata, siis on võimalik võrrelda võrdsetel alustel kuutasu hinna komponente ja kasutatavaid teenuseid. Autor on hinnanud realistlikuks salvestusruumiks 200 GB. Hindade valideerimiseks on enamjaolt olemas MS avalikud allikad ja hinnakirjad. Oluline on arvutusse kaasata ka juurutusest tulenev spetsialisti tööhind, et hinnata võrreldavaid suurusjärke on jagatud investeering 60 kuu ehk 5 aasta peale. Spetsialisti tunnitasu arvestamiseks või analoogse teenuse sisseostmise kulu hindamiseks kasutas autor nii enda tööalaseid teadmisi töötades samas valdkonnas kui ka statistikaameti vastavat avalikku palgainfot [17]. Tabelist järeldub, et kõige odavam lahendus on lahendus 2. Synapse Serverless eeldatava kuu hinnaga 275,5 €, sellele järgneb lahendus 1. SQL Server hinnaga 717,5 €, seejärel lahendus 3. Spark Pool hinnaga 857,0 € ning kõige kallim on MS Fabricu kuutasuhinnaga 1917,5 €. Kuigi sellisena võrreldes on pilt kulude osas üsna selge, peab lõpliku otsuse hindamisel kindlasti arvestama ettevõtte võimalike tuleviku soovide ja vajadustega, et mitte investeerida valesse lahendusse mida peab mõne aja möödudes hakkama lisanduva maksumusega ümber kujundama või välja vahetama. Tabel 8 hinnangud on visuliseeritud tulpdiagrammina (vt Joonis 9) all.



Joonis 9. Valikute võrdlus maksumuse kriteeriumi suhtes.

Turvalisus ja ei ole antud võrdluses küll kõrge kaaluga kriteerium, kuid kvaliteetsest turvasemest võib pigem mõelda kui tänapäevase andmearhitektuuri lahenduse vaiketasemest. Tuleb negatiivselt esile, kui mõni lahendus ei vasta etteantud turva ja küberkaitse nõuetele.

Tabel 11. Turvalisuse kriteeriumi hinnangud.

Turvalisus	SQL Server	Synapse	Spark	Fabric
1. SQL Server	1,00	0,50	0,50	0,33
2. Synapse	2,00	1,00	1,00	0,50
3. Spark	2,00	1,00	1,00	0,50
4. Fabric	3,00	2,00	2,00	1,00

Nagu on näha ülal paiknevast tabelist (vt Tabel 11), siis turvalisuse suhtes on kõige kehvem tulemus lokaalsel SQL serveril ja kõige tugevam MS Fabricul. Ettevõtete lokaalsetesse serveritesse sissemurdmist, andmebaaside ja tarkvarade kaaperdamist on Eesti tööstusettevõteteski viimastel aastatel mitmeid juhtunud [18]. Seega ei pea neid näiteid kaugelt otsima. Lisaks on täiendav tegevus füüsilise riistvara hoolduse ja andmete säilimise eest vastutada õnnetusjuhtumite puhul. Pilvetehnoloogia ja eeskätt terviklik platvorm nagu MS Fabric omab kindlasti selles valguses eeliseid. Tabel 11 hinnangud on visuliseeritud tulpdigrammina (vt Joonis 10) all.



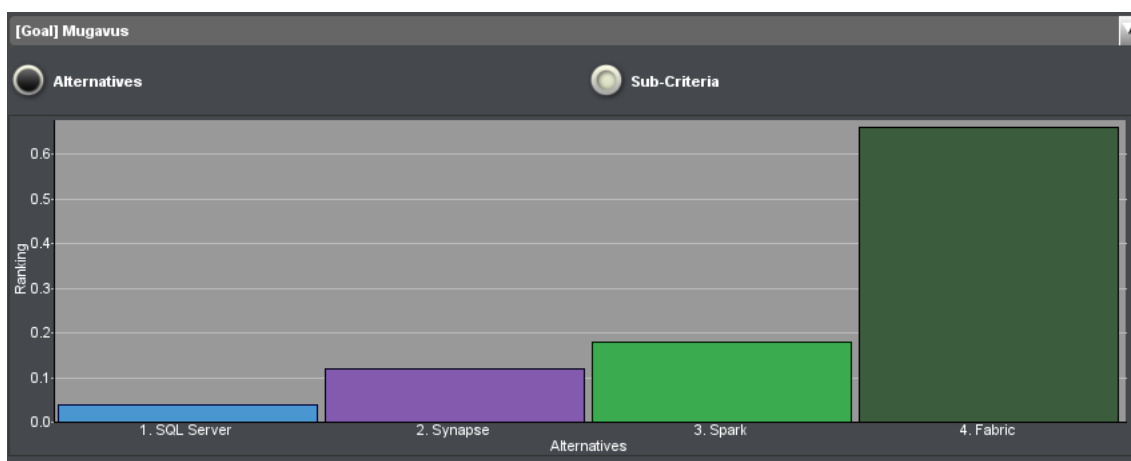
Joonis 10. Valikute võrdlus turvalisuse kriteeriumi suhtes.

Kasutusmugavus on prioriteedilt eelviimase kaaluga kriteerium. Sellegipoolest on andmete haldamise ja andmetarbijate seadistamise töö ressursside nõudlik ja kulukas, eriti juhtudel kui vaadeldavatest lahendustest osad vähendavad käsitöö miinimumini või elimineerivad sootuks.

Tabel 12. Kasutusmugavuse kriteeriumi hinnangud.

Kasutusmugavus	SQL Server	Synapse	Spark	Fabric
1. SQL Server	1,00	0,20	0,20	0,11
2. Synapse	5,00	1,00	0,50	0,14
3. Spark	5,00	2,00	1,00	0,20
4. Fabric	9,00	7,00	5,00	1,00

Nagu ülal antud hinnangutest nähtub (vt Tabel 12), siis kasutusmugavuse osas saab kõige nõrgemad tulemused lokaalne SQL Server, ning kõige parema kasutusmugavusega – lihtsam, on MS Fabric. Hinnangu andmisel on lähtutud tavastase inimese tasemest, kes tingimata ei hooli soovist lahendus ise detailsetest komponentidest kokku panna. Samuti on oluline arenduste ja seadete läbiviimise kiirus, siin on kindlasti eelis terviklike kasutajakeskkondadega lahendustel. Tabel 12 määratud hinnangud on visuliseeritud tulpdiaagrammina (vt Joonis 11) all.



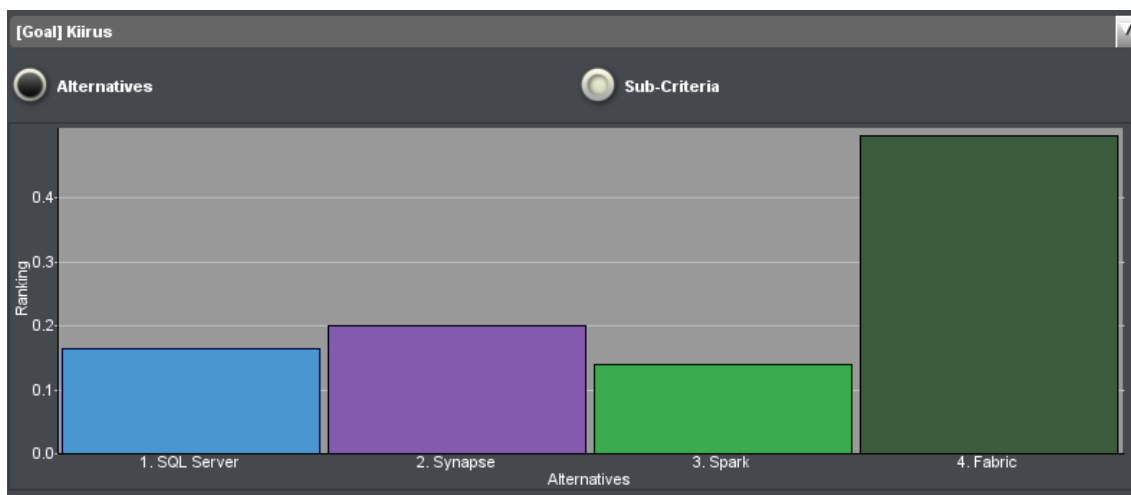
Joonis 11. Valikute võrdlus mugavus kriteeriumi suhtes.

Kiirus kriteeriumina on täna kõige madalama kaaluga, sest enamik andmeid uuendatakse korra ööpäevas. Tootmisettevõtte kasutab küll operatiivtarkvara, kuid selle andmeliikumine on täna vaadeldavast andmeliikumise voost eraldi. Tulevikus võiks loomulikult loodav lahendus lahendada ka operatiivandmete tarbijatele vajalikud nõudmised.

Tabel 13. Kiiruse kriteeriumi hinnangud.

Kiirus	SQL Server	Synapse	Spark	Fabric
1. SQL Server	1,00	1,00	1,00	0,33
2. Synapse	1,00	1,00	2,00	0,33
3. Spark	1,00	0,50	1,00	0,33
4. Fabric	3,00	3,00	3,00	1,00

Nagu järeldub hinnangutest (vt Tabel 13), siis eristuvalt kiiremaks andmeühenduseks võib pidada MS Fabric platvormi. MS Fabric sisaldab komponente nagu *Data Activator*, mis võimaldab vaadeldavaid ärianalüütika aruandeid uuendada reaalsajas. Lahendus 2. on kiirem lahendus 3. kuna puudub vajadus luua Delta/Parquet faile läbi Spark Pool töövihikute. Tabel 13 hinnangud on visuliseeritud tulpdiaagrammina (vt Joonis 12) all.



Joonis 12. Valikute võrdlus kiirus kriteeriumi suhtes.

4.3.3 Hindamismaatriksi konsistentsuse kontroll

Konsistentsi saame kontrollida läbides eelkirjeldatud arvutuse (vt peatükk 4.2.5)

Kriteeriumite paariviisiline võrdlus AHP meetodis eeldab, et hindamismaatriks on järjepidev, st sisemiselt loogiline ja vastab ootustele. Konsistentsuse hindamiseks kasutatakse konsistentsuse indeksit (CI) ja konsistentsuse suhet (CR). Antud juhul, lähtudes tabelis esitatud võrdlustest ja Priority Estimation Tool[15] poolt arvatud suurustest, saame leida λ_{\max} järgnevalt[19]:

Hindamismaatriks (Tabel 14) tuleneb Tabel 6 antud paariviisilistest hinnangutest.

Tabel 14. Hindamismaatriks.

1,00	5,00	5,00	5,00	5,00	5,00
0,20	1,00	0,25	0,33	0,50	1,00
0,20	4,00	1,00	4,00	4,00	4,00
0,20	3,00	0,25	1,00	3,00	3,00
0,20	2,00	0,25	0,33	1,00	2,00
0,20	1,00	0,25	0,33	0,5	1,00

Seejärel kasutame juba leitud kaale kaaluvektoriks(w) (vt Tabel 15).

Tabel 15. Kaaluvektor(w).

0,475	0,051	0,230	0,122	0,072	0,051
-------	-------	-------	-------	-------	-------

Korrutame maatriksi ja kaaluvektori ridade kaupa sarnaselt esimesele reale:

$$(1,00 \times 0,475) + (5,00 \times 0,051) + (5,00 \times 0,230) + (5,00 \times 0,122) + (5,00 \times 0,072) + (5,00 \times 0,051) = 3,105$$

Terve maatriksi läbi arvutades saame maatriksi ja vektori korrutise ($A \times w$) (vt Tabel 16)

Tabel 16. Maatriksi ja vektori korrutis ($A \times w$).

3,105	0,331	1,509	0,797	0,469	0,331
-------	-------	-------	-------	-------	-------

Seejärel jagame iga rea korrutise vastava kaaluvektori elemendiga $(A \times w)_i/w_i$. Esimese rea (kriteerium: Täpsus) näide $3,105/0,475=6,536$. Kõikide ridade tulemus vastavalt (vt Tabel 17)

Tabel 17. Iga rea suhete tulemus.

6,536	6,485	6,561	6,529	6,511	6,485
-------	-------	-------	-------	-------	-------

Lõpuks saame leida aritmeetilise keskmise mida otsime.

$$\lambda_{\max} = (6,536 + 6,485 + 6,561 + 6,529 + 6,511 + 6,485) / 6 = 6,518$$

$$\lambda_{\max} = 6,52.$$

$$n=6$$

Konsistententsuse indeksi saame arvutada valemiga:

$$CI = \frac{(\lambda_{\max} - n)}{(n-1)} = \frac{(6,52-6)}{5} = 0,104$$

Järgnevalt arvutame konsistententsuse suhte $CR=CI/RI$, kus RI on juhuslik indeks ja $n=6$ korral $RI=1,24[20]$. Tulemus $CR=0,084$.

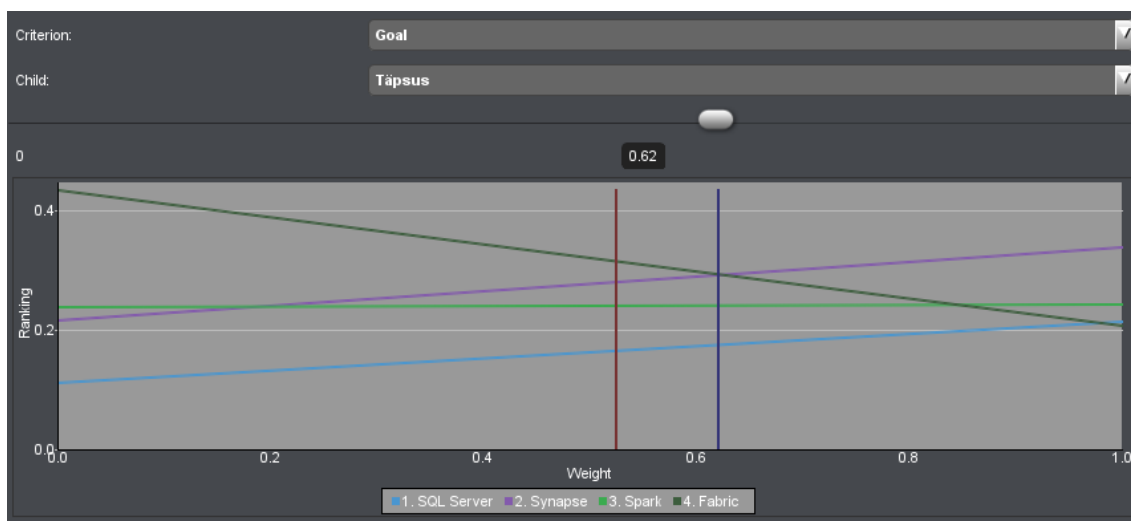
$CR = 0,084$ ja jääb alla kriitilise piiri $CR < 0,1$ mis näitab, et maatriks on piisavalt konsistentne. Seega on kriteeriumite võrdlused usaldusväärsed ning neid saab kasutada otsustusprotsessis. Konsistententsuse suhe kinnitab, et hindamismudeli sisemine loogika on

kooskõlas AHP meetodika standarditega. See tagab kaalukate otsuste tegemisel vajaliku täpsuse ja kindluse.

4.3.4 Tundlikkuse analüüs

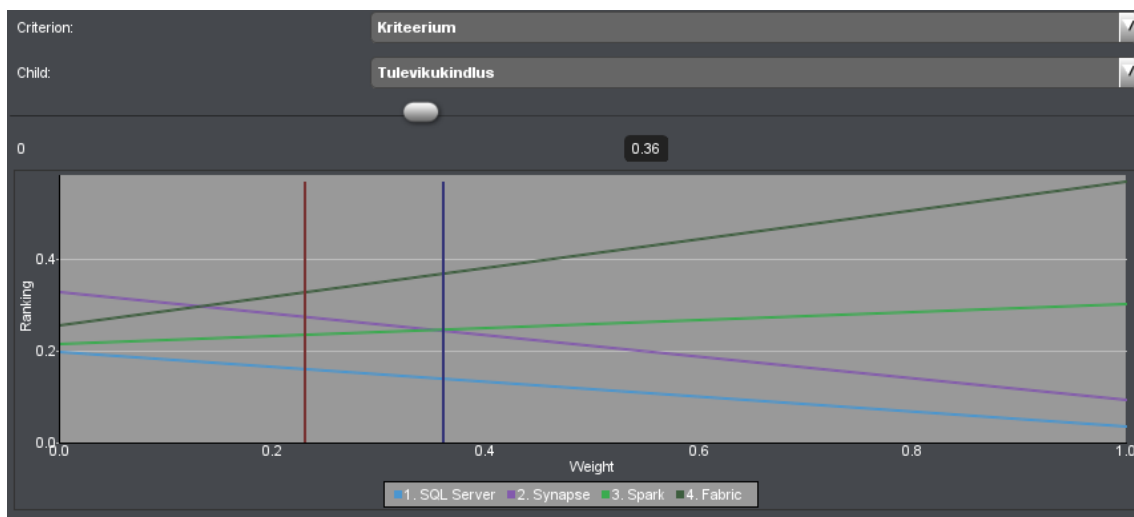
Kui ollakse tõestanud hindamismudeli vastamist konsistentsuse nõuetele, siis on mõistlik hinnata ka saadud kriteeriumite kaalude mõju eelistustele.

Joonis 13 näitab täpsuse kriteeriumi tundlikkuse analüüsi graafikut. Kui täpsuse osakaalu tõsta 0,53 pealt 0,62 peale, siis oleks parim lahendus 2. Synapse. Kaalude suhe teistesse kaaludesse enne muudatust $0,52/(1-0,52)=1,08$ on ning pärast muudatust $0,62/(1-0,62)=1,6$ seega Täpsuse kriteeriumit tuleks veelgi suurendada teiste kriteeriumite suhtes $1,6/1,08=1,48$ korda.



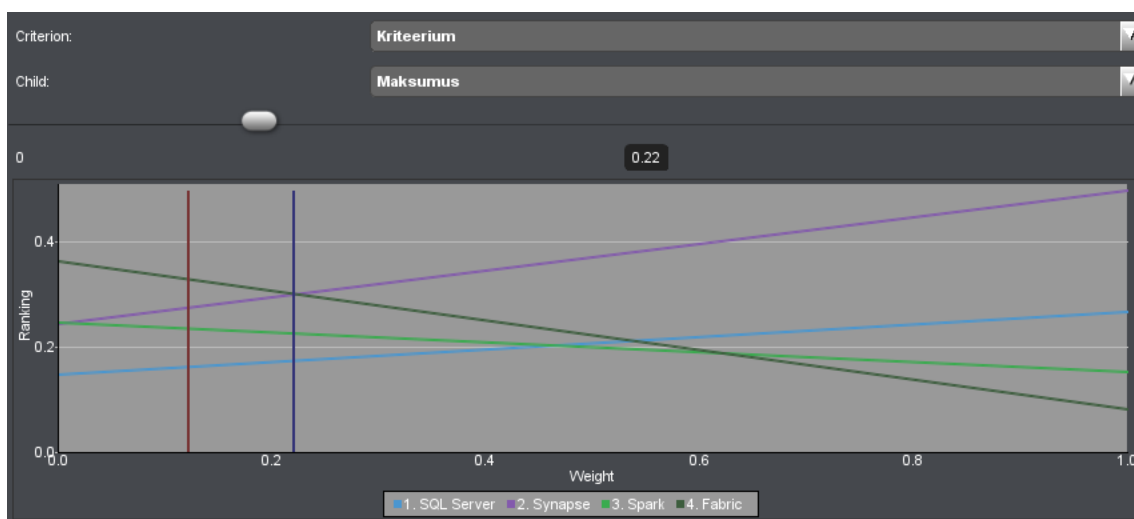
Joonis 13. Täpsuse kriteeriumi tundlikkuse graafik.

Joonis 14 esitab tulevikukindluse kriteeriumi tundlikkuse analüüsi graafikut. Kui tulevikukindluse kriteeriumi osatähtsust tõsta algsest väärtusest 0,23 kuni 0,36, toimub alternatiivide järjestuses muutus. Muutuse tulemusena saab teise parima lahenduse alternatiiv 2. Synapse asemel alternatiiv 3. Spark. Kaalude suhe teiste kriteeriumite suhtes enne muudatust arvutatakse valemiga: Enne muudatust, $0,23/(1-0,23)=0,3$ Pärast kaalu tõstmist 0,36 arvutatakse suhe samal viisil $0,3/(1-0,36)=0,563$ seega tulevikukindluse kriteeriumit tuleks veelgi suurendada teiste kriteeriumite suhtes $0,563/0,3 = 1,88$ korda.



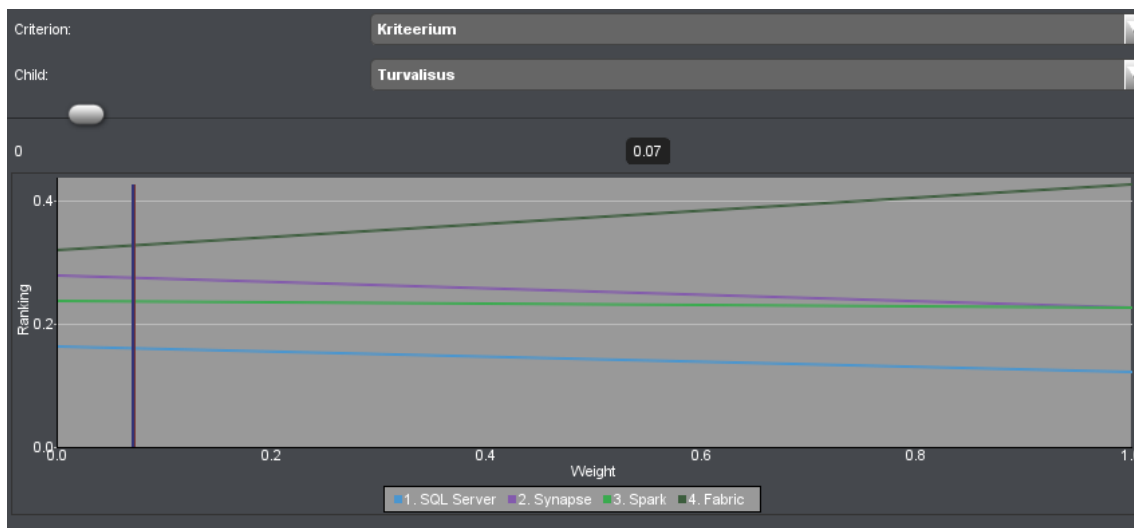
Joonis 14. Tulevikukindluse kriteeriumi tundlikkuse graafik.

Joonis 15 esitab maksumuse kriteeriumi tundlikkuse analüüsi graafikut. Kui maksumuse kriteeriumi osatähtsust tõsta algsest väärtusest 0,12 kuni 0,22, toimub alternatiivide järjestuses muutus. Muutuse tulemusena saab parimaks alternatiiviks 2. Synapse, mis varasemalt polnud esikohal. Kaalude suhe teiste kriteeriumite suhtes enne muudatust arvutatakse valemiga $0,12/(1-0,12)=0,136$. Pärast kaalu tõstmist 0,22 arvutatakse suhe samal viisil $0,22/(1-0,22)=0,282$. Seega maksumuse kriteeriumit tuleks veelgi suurendada teiste kriteeriumite suhtes $0,282/0,136=2,07$ korda. Adekvaatselt hinnates tuleviku stsenaariumeid ei peaks mitte maksumuse kaalu tõstma vaid langetama, sest olulisemaks võib muutuda lahenduse omadus seda odavamalt uuendada.



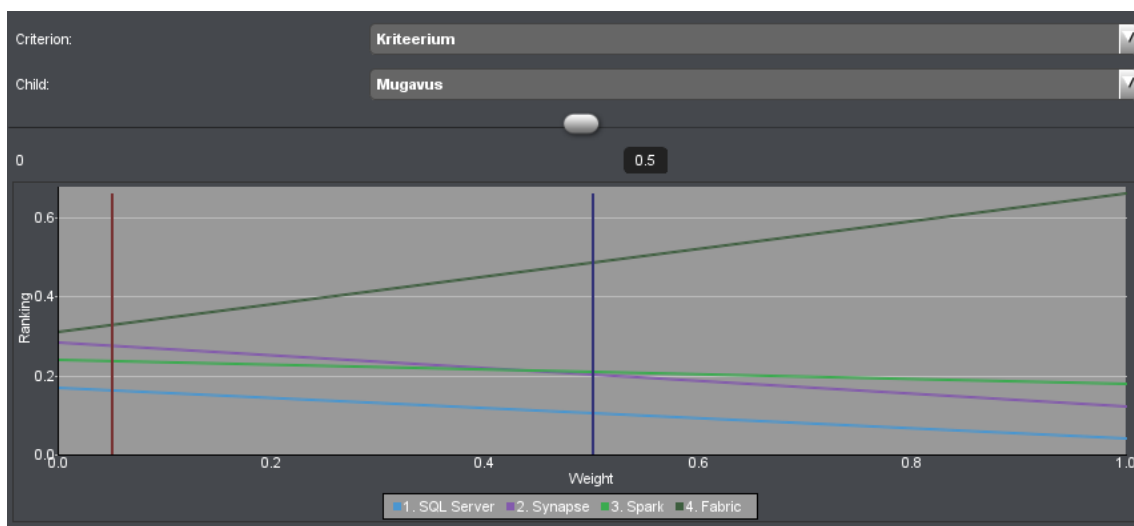
Joonis 15. Maksumuse kriteeriumi tundlikkuse graafik.

Joonis 16 esitab turvalisuse kriteeriumi tundlikkuse analüüsi graafikut. Siit on näha, et kaalu suurenedes sisulisi muutusi ei esine eelistustes ei esine.



Joonis 16. Turvalisuse kriteeriumi tundlikkuse graafik.

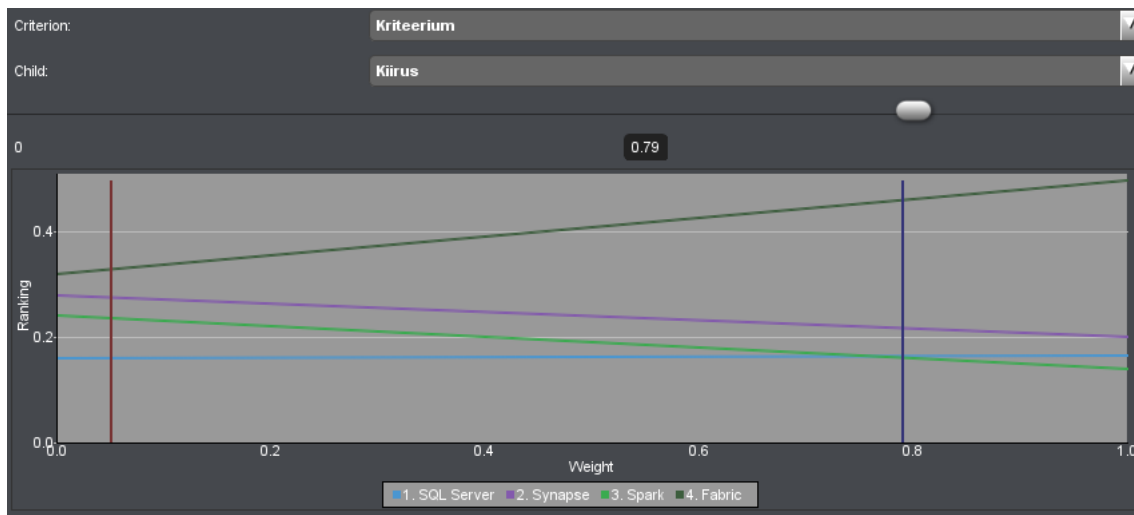
Joonis 17 esitab mugavuse kriteeriumi tundlikkuse analüüsi graafikut. Kui mugavuse kriteeriumi osatähtsust tõsta algsest väärtusest 0,05 kuni 0,5, toimub alternatiivide 1 ja 3 eelistuse järjestuses. Esikohal on alternatiiv 4. MS Fabric. Kaalude suhe teiste kriteeriumite suhtes enne muudatust arvutatakse valemiga $0,05/(1-0,05)=0,053$. Pärast kaalu tõstmist 0,5 arvutatakse suhe samal viisil $0,5/(1-0,5)=1$. Seega mugavuse kriteeriumit tuleks veelgi suurendada teiste kriteeriumite suhtes $1/0,053=18,87$ korda.



Joonis 17. Mugavuse kriteeriumi tundlikkuse graafik.

Joonis 18 esitab kiiruse kriteeriumi tundlikkuse analüüsi graafikut. Kui kiiruse kriteeriumi osatähtsust tõsta algsest väärtusest 0,05 kuni 0,79, toimub alternatiivide

järjestuses muutus. Muutuse tulemusena vahetavad 2. ja 3. koht omavahel positsioonid, kus alternatiiv 3. Spark muutub paremaks kui alternatiiv 2. Synapse. Kaalude suhe teiste kriteeriumite suhtes enne muudatust arvutatakse valemiga $0,05/(1-0,05)=0,053$. Pärast kaalu tõstmist 0,79 arvutatakse suhe samal viisil $0,79/(1-0,79)=3,76$. Seega kiiruse kriteeriumit tuleks veelgi suurendada teiste kriteeriumite suhtes $3,76/0,053=70,94$ korda.



Joonis 18. Kiiruse kriteeriumi tundlikkuse graafik.

Kokkuvõttes näitab tundlikkuse analüüs AHP meetodika kontekstis, kuidas erinevate kriteeriumite kaalude muutmine eraldiseisvalt võib mõjutada alternatiivide paremusjärjestust ja otsustusprotsessi tervikuna. Analüüs kinnitab, et kriteeriumite osakaalu muutmine juba 1,48-2,07 korda võib tuua kaasa muutusi alternatiivide järjestuses, nagu näha näiteks täpsuse, tulevikukindluse ja maksumuse kriteeriumide puhul. Samas tuleb mõista, et need on hetkel kõige kõrgema kaaluga kriteeriumid ja hinnanguid korrigeerides nende suhtelist kaalu pigem langetatakse kui tõstetakse. Mõne kriteeriumi, nagu turvalisuse, puhul kaalu muutmine sisulisi muudatusi ei tekita. Tundlikkuse analüüsi tulemused aitavad mõista, millised kriteeriumid on otsustusprotsessis kõige mõjukamad ja kuidas nende prioriteetsust tuleks vastavalt vajadusele kohandada.

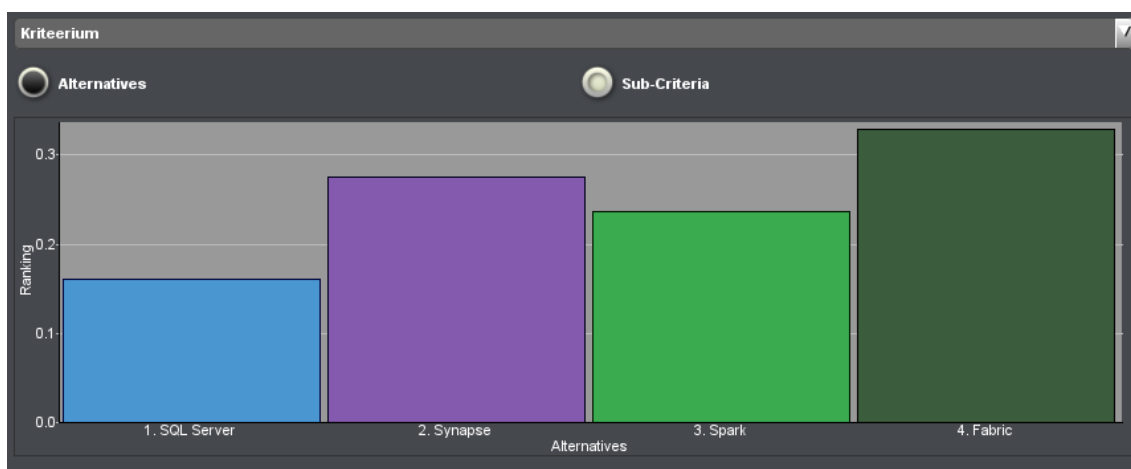
4.3.5 Valikute analüüs kriteeriumite üleselt kokku

Valikute kõiki aspekte koos vaadelda on käesoleva analüüsi põhiline eesmärk. Arvutatud kaalud on vastavad järgmisele tabelile (vt Tabel 18)

Tabel 18. Alternatiivide arvatud sobivusjärjestus.

Lahendus	Arvatud paremus
MS Fabric	0,328
Synapse Serverless	0,275
Spark Pool	0,236
Lokaalne SQL Server	0,161

Sama informatsioon on nähtaval ka graafilise tulpdiaagrammina (vt Joonis 19)



Joonis 19. Alternatiivide arvatud sobivusjärjestus graafikul.

Nagu järeldub, siis ette antud kriteeriumite tähtsuse ja kriteeriumite hinnangute kombinatsioonina on kõige parem valik MS Fabric (kaal 0,328), teine on Synapse Serverless (kaal 0,275), kolmas on Spark Pool (kaal 0,236) ja neljas variant on lokaalne SQL Server (kaal 0,161). Vahed ei olnud väga suured, aga see oli ka ootuspärane, sest lahendused täidavad sama funktsiooni sealjuures täidavad kõik kehtestatud funktsionaalseid nõudeid. Suuremad erisused tulevad sisse mittefunktsionaalsete nõuete osas ja nendel baseerudes on võimalik kõige paremini sobiv lahendus välja selekteerida. Saadud AHP tulemust võib kindlasti hinnata usaldusväärseks ja realistlikuks.

5 Analüüsi tulemused ja järeldused

Analüüsi eesmärk oli valideerida ja aru saada milline vaadeldavatest andmearhitektuuri lahendustest on kõige paremini sobituv vaadeldava tootmisettevõtte ärinõuete ja vajadustega pikaajaliselt. Lisaküsimus oli aru saada kuidas suhestub alternatiivsete lahendustega Microsoft kõige uuem andmeanalüütika platvorm MS Fabric.

5.1 Saadud tulemus

Eelmises peatükis (vt peatükk 4) läbitud analüüsi tulemusena võib öelda, et valitud nelja lahenduse seast kõige paremini vaadeldava tootmisettevõtte praegusi ja tuleviku ärinõudeid katab ühtlasi ka kõige uuem tehnoloogia – MS Fabric.

See on väärtuslik informatsioon, sest vaadeldavate tehnoloogiate kasutamine ja juurutamine on pikaajaline ja suhteliselt kallis protsess. Uute ERP süsteemide juurutamise perioodil võib minna kogu ajaline ressurss sellesse projekti. Ärianalüütika, mis siis, et ärikriitiline, ei pruugi saada soovitud tähelepanu ja allokeeritud analüüsivõimekust. Sellised analüüsid nagu käesolev bakalaureusetöö võib sellisel juhul säästa märkimisväärselt kuluvaid ressursse ja aidata otsustajatel kiiremini jõuda õigete otsusteni. Lisaks annab see tulemus otsustajale enesekindlust teha julgemaid otsuseid valides uuemaid turul pakutavaid tehnoloogiaid.

Paremuselt teiseks sobivaimaks tulemuseks sai lahendus 2 – Synapse Serverless. See oli kaugel kõige madalama maksumuse, ehk kõige parema tulemusega lahendus maksumuse kriteeriumi lõikes. Ka tehnoloogiliselt on tegu konkurentsivõimelise lahendusega, kuid mitte rohkem kui lahendus 3. Spark Pool ja 4. MS Fabric. Lahenduse miinuspooleks on kindlasti kallim ja pikem juurutusprotsess.

Paremuselt kolmanda tulemuse saavutas lahendus 3. Spark Pool. Tegum on sobiva ülemineku arhitektuuriga Synapse tehnoloogia pealt MS Fabric platvormi peale kuid ka ilma domineerivate eelisteta.

Neljanda ehk valikutest paremuselt viimase positsiooni sai maksimaalselt olemasolevat loodud andmehõive ja töötuse infrastruktuuri säilitav lokaalne SQL Serveritele ja andmebaasidele baseeruv lahendus. Suureks väljakutseks on antud lahenduse ülalpidamine ja uuenevate ärinõuetega toimetulemine. Positiivse poole pealt on tegu

kriteeriumi täpsus kõige parema tulemusega. Arusaadavalt on selliste lokaalsete serverite andmeliikumise tsüklite seadistamine ja andmete sisu määramine võimalik väga täpselt paika panna. Iseasi kas selleks on alati vajadus ja oskused.

Saaty meetodi läbiviimisel valideerisime antud hinnangute konsistentsust ja tundlikust. Hinnangud vastasid konsistentsuse nõuetele, sest arvutatud konsistentsi suhe jäi alla 1. Tundlikkuse osas jäid 3 kaalukamat näitajat: täpsus, tulevikukindlus, maksumus kordaja 1,5-2 juurde mis tähendab, et see kriteerium võib muuta järjestust mõõdukal tasemel kuid ei ole liialt domineeriv. Kriteeriumid turvalisus, mugavus ja kiirus olid kordajatega 18 ja rohkem ning kaalude muutmine paremusjärjestust ei mõjuta. Otsustusprotsess on piisavalt stabiilne ja usaldusväärne.

5.2 Autori poolsed järeldused

Autori hinnangul saavutas käesolev bakalaureusetöö seatud eesmärgid. Vaadeldi nelja kõige realistlikumat andmearhitektuuri lahendust mida ka konkreetne tootmisettevõtte võiks kaaluda ja kaalub. Kirjeldati ja analüüsiti läbi neli ideelist lahendust ja jõuti põhjendatud soosikuni.

Seatud hüpotees, kas uuemad tehnoloogiad omavad eeliseid vanemate ees osutus käesoleva kaasuse baasilt tõeseks. Siiski on sellisel puhkudel palju nüansse ja iseärasusi, iga erineva kaasuse puhul tuleb neid hinnata erinevalt ja ei saa öelda, et üldpildis on alati uued lahendused paremad kui olemasolevad või vanemad versioonid.

Töö maht ja skoop olid hästi seatud, sest etteantud aja ja töömahuga ei oleks olnud võimalik võtta juurde andmearhitektuuri alternatiive või oleks jäänud analüüs oluliselt pealiskaudsemaks. Õigesti seatud valikute ja analüüsi kitsendused olid põhjendatud. Sellele vaatamata ületas bakalaureusetöö leheküljelist mahtu 12 lehekülje võrra. Autor põhjendab seda vajadusega struktuursed 4 põhiteemat piisava detailistmega lahti kirjutada ja analüüsida, et töö ei jääks pealiskaudseks ja umbmääraseks.

Väga positiivseks peab autor isikliku õppimist ja antud teema osas paljude aspektide selgeks tegemist. Nagu varasemalt kirjeldatud on MS Fabric üsna uus tehnoloogia enda uute ja kiiresti muutuvate võimalustega. Avalikult ei ole sellest tehnoloogiast palju infot või väljaandeid peale tootja välja antud infomaterjalide, koolituste, seminariettekanete.

Tööl on loomulikult ka nõrkusi milleks on asjaolu, et andmearhitektuure ja tehnoloogilisi lahendusi on rohkem kui neli. Sisuliselt võiks eri tootjate või eritehnoloogiate osas valideerida mitukümmend lahendust, kuid selline analüüs ei ole antud töömahu juures teostatav.

Töö puudus on asjaolu, et autor ei tööta ärianalüütika arendaja, konsultandi või programmeerijana, see tähendab, et autoril puudus võimalus vaadeldavatesse korporatiivtarkvarade ja -komponentide täisseedistustesse või demolahendustesse sisse minna, et tõsta analüüsi detailiastet. Lisada tehnilisi skeeme või jooniseid. Sisuliselt võiks igast vaadeldavast lahendusest või komponendist eraldi luua mitu akadeemilist uurimistööd.

Edasiulatuvalt jätkutegevustena soovitab autor andmearhitektuuri valikul mõelda enda ettevõtte lühiajalistest ja pikaajalistest arendusplaanidest mis seonduvad andmete ja andmetarbijatega. Uued arendused on ressursi mahukad ja üsna kallid. Samas pakuvad kiirelt arenevad tehnoloogilised AI ja ärianalüütika võimalused olulist võitu töötajate tööaja kokkuhoiu või protsesside kiirema ja täpsema läbimise kontekstis. Taustal tekitavad tulevikus tarkvarad ja seadmed järjest rohkem andmeid ja neid on mõistlik ärielistel eesmärkidel ka ära kasutada. Seega on kasu potentsiaal suur ja väga suur.

6 Kokkuvõte

Korrektsete andmete kogumine, töötlemine ja kasutamine ärieesmärkide saavutamiseks on tänapäeva tootmisettevõtete konkurentsitingimustes vältimatu tegevus. Ärianalüütika kasutamine, et võtta vastu korrektseid ja õigeid otsuseid kiiresti on konkurentsieelis ja eesmärk mille poole peaks iga kaasaegne tootmisettevõtte püüdlema.

Käesoleva bakalaureusetöö eesmärk oli valida kõige sobivam andmearhitektuuri lahendus ärianalüütika kasutamiseks tootmisettevõttele, kes on otsustanud kasutada MS D365 ERP süsteemi. Lisaks oli eesmärk võrrelda autori valitud lahendusi Microsofti uusima andmeanalüütika platvormiga MS Fabric.

Analüüsi tulemusena leiti, et MS Fabric on parim lahendus, kuna see rahuldab kõige paremini ettevõtte praeguseid ja eeldatavaid tuleviku ärinõudeid. See on oluline, kuna uue tehnoloogia juurutamine on pikaajaline ja kulukas protsess.

Teiseks parimaks lahenduseks osutus Synapse Serverless, millel kõige suurem eelis oli suhteline madalaim hind, kuid juurutusprotsess oli pikem. Kolmas koht kuulus Spark Pool'ile, mis sobis üleminekuks Synapse tehnoloogia pealt MS Fabric platvormile. Kõige vähem sobivaks osutus lokaalne SQL Server lahendus, mis ei suuda hästi kohanduda muutuvate ärinõuetega ja mis vajab põhjalikumad seadistustööd.

Analüüsi tulemused näitasid, et MS Fabric vastas kõigile kriteeriumidele, nagu täpsus, tulevikukindlus ja maksumus, ning oli vastupidav muudatustele. Analüüsitavaid lahenduste valikut kontrolliti ka Saaty meetodil, mis kinnitas tulemuste usaldusväärsust.

Autori hinnangul on töö täitnud oma eesmärgid, andes selge ülevaate neljast sobivast andmearhitektuuri lahendusest, mida tootmisettevõtted kes kasutavad või hakkavad kasutama MS D365 ERP süsteemi võiksid kaaluda.

Kasutatud kirjandus

- [1] Thomas H. Davenport, „Analytics 3.0,” 2013, [Võrgumaterjal]. <https://hbr.org/2013/12/analytics-30> [Kasutatud 14.10.2024].
- [2] A. McAfee ja E. Brynjolfsson, „Big Data: The Management Revolution,” 2012, [Võrgumaterjal]. <https://hbr.org/2012/10/big-data-the-management-revolution> [Kasutatud 14.10.2024].
- [3] E Brynjolfsson, L. Hitt, H. Kim, „Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?,” 2011, [Võrgumaterjal]. https://a51.nl/sites/default/files/pdf/SSRN_id1819486.pdf [Kasutatud 14.10.2024].
- [4] “Data Lake Explained: A Comprehensive Guide to Its Architecture and Use Cases,” 2023. [Võrgumaterjal]. <https://www.altexsoft.com/blog/data-lake-architecture/> [Kasutatud 04.11.2024].
- [5] “Azure Data Factory pricing,” 2024. [Võrgumaterjal]. <https://Azure.Microsoft.com/en-us/pricing/details/data-factory/ssid/> [Kasutatud 30.11.2024].
- [6] “Microsoft SQL Server 2012 Licensing Guide,” 2013. [Võrgumaterjal]. https://download.microsoft.com/download/7/3/c/73cad4e0-d0b5-4be5-ab49-d5b886a5ae00/sql_server_2012_licensing_reference_guide.pdf [Kasutatud 04.11.2024].
- [7] “Access lake databases using serverless SQL pool in Azure Synapse,” 2023. [Võrgumaterjal]. <https://learn.microsoft.com/en-us/azure/synapse-analytics/metadata/database> [Kasutatud 21.11.2024].
- [8] “Azure Synapse Analytics pricing,” 2024. [Võrgumaterjal]. <https://Azure.Microsoft.com/en-us/pricing/details/Synapse-analytics/> [Kasutatud 30.11.2024].
- [9] Rakesh Chanda “Understanding Delta Lake Table vs Parquet,” 2024. [Võrgumaterjal]. <https://medium.com/@rakeshchanda/understanding-delta-lake-table-vs-parquet-0424368dc0c1> [Kasutatud 30.11.2024].

- [10] “Microsoft Fabric pricing,” 2024. [Võrgumaterjal]. <https://Azure.Microsoft.com/en-us/pricing/details/Microsoft-fabric/> [Kasutatud 30.11.2024].
- [11] „Link your Dataverse environment to Microsoft Fabric and unlock deep insights,” 2024. [Võrgumaterjal]. <https://learn.microsoft.com/en-us/power-apps/maker/data-platform/azure-synapse-link-view-in-fabric> [Kasutatud 04.12.2024].
- [12] “Dynamics 365 Licensing Guide,” 2024. [Võrgumaterjal]. <https://download.microsoft.com/download/9/6/7/96706B15-1CBE-47B7-AB9E-6BC31A377BBB/Dynamics%20365%20Licensing%20Guide%20Feb%202024.pdf> [Kasutatud 24.11.2024].
- [13] L. Võhandu, „Subjektiiivsetest hinnangutest objektiivsete tulemusteni,” Tallinna Tehnikaülikool, Tallinn, 1998.
- [14] N. Bhushan ja K. Rai, „Strategic Decision Making: Applying the Analytic Hierarchy Process,” Springer, 2004. [Võrgumaterjal]. https://issuu.com/victore.cardozodelgado/docs/strategic_decision_making._applying [Kasutatud 04.11.2024].
- [15] „Priority Estimation Tool,” [Võrgumaterjal]. [https://sourceforge.net/projects/priority/..](https://sourceforge.net/projects/priority/) [Kasutatud 01.11.2024].
- [16] Sajid Siraj, „PriEsT Beginner’s Guide,” [Võrgumaterjal]. <https://priority.sourceforge.io/priest-user-guide.pdf> [Kasutatud 04.11.2024].
- [17] „Palgarakendus,” 2024. [Võrgumaterjal]. <https://palgad.stat.ee/> [Kasutatud 25.11.2024].
- [18] K. Kaaver „Meie süsteemis oldi sees mitu kuud.“ Küberrünnaku läbi elanud ettevõtja räägib loo perefirmade kolmest kõige raskemast nädalast,” 2023, [Võrgumaterjal]. <https://arileht.delfi.ee/artikkel/120190212/meie-susteemis-oldi-sees-mitu-kuud-kuberrunnaku-labi-elanud-ettevotja-raagib-loo-perefirmade-kolmest-koige-raskemast-nadalast> [Kasutatud 04.12.2024].
- [19] Dr Stuart Burge, „The Systems Engineering Tool Box,” 2014. [Võrgumaterjal]. <https://www.burgehugheswalsh.co.uk/Uploaded/1/Documents/Analytic-Hierarchy-Process-Tool-v2.pdf> [Kasutatud 30.11.2024].
- [20] H. A. Donegan and F. J. Dodd, „A Note on Saaty’s Random Indexes,” 1991, [Võrgumaterjal]. <https://www.sciencedirect.com/science/article/pii/089571779190098R/pdf?md5=c3>

bf05e2eeb4d9ae041bdbb3340588e&pid=1-s2.0-089571779190098R-main.pdf

[Kasutatud 04.11.2024].

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Ats Raigla

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Andmearhitektuuri analüüs ärianalüütika kasutamiseks tootmisettevõttes“, mille juhendaja on Tarvo Treier.
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

26.12.2024

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.