

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Marko Lindeberg 193669IAAB

Meeskonnapõhine infrastruktuuri kuluarvestus Entigo OÜ kliendi näitel

Bakalaureusetöö

Juhendaja: Kristiina Hakk

PhD

Kaasjuhendaja: Rein Rimmel

Rakenduskõrgharidus

Tallinn 2022

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: Marko Lindeberg

16.05.2022

Annotatsioon

Käesoleva bakalaureusetöö eesmärk on leida ja teostada tehniline lahendus Entigo OÜ kliendi näitel meeskonnapõhiseks kuluarvestuseks, mille abil muuta pilveteenuste kulud arendusmeeskondadele läbipaistvamaks.

Probleemiks on kõrged pilveteenuste kulud, mille kontrolli alla saamiseks on esmalt vaja selget ülevaadet kuludest. FinOps kultuuripraktika pakub selle lahendamiseks välja meeskonnapõhise kuluarvestuse, mis ongi antud lõputöö fookuses.

Selleks mõtestatakse lahti taustainfo ning defineeritakse lõputöö eesmärgi täitmiseks vajalikud tehnilised nõuded. Analüüsi käigus tehakse ülevaade olemasolevatest töövahenditest, kirjeldatakse nende omadusi ja puudusi ning hinnatakse nende sobivust lõputöös seatud nõuetega. Selle tulemusel valitakse välja sobivaim töövahend tehnilise lahenduse teostamiseks.

Tehnilise teostuse käigus valmib lahendus, mis muudab pilveteenuste kulud arendusmeeskondadele läbipaistvamaks, pakkudes selleks meeskonnapõhist kuluarvestust. Lahendus saab olema ettevõtteülene, et seda oleks võimalik kasutada ka teiste ettevõtete jaoks.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 49 leheküljel, 7 peatükki, 21 joonist, 1 tabelit.

Abstract

Team Based Cost Accounting of the Cloud Infrastructure on the Example of Entigo OÜ's Customer

The aim of this Bachelor's thesis is to find and implement a technical solution for team-based cost accounting on the example of Entigo OÜ's customer that will make the costs of cloud services more transparent to its development teams.

The problem is the high cost of the cloud services along with the lack of clear overview of these costs that would help to address the problem. FinOps cultural practice suggests team-based cost accounting to address the excessive cloud costs. Team-based cost accounting is also the focus of current thesis.

In the theoretical part of this thesis, technical requirements are defined for the desired solution. In the course of the analysis, an overview of the existing tools is done and the features and disadvantages of these tools are assessed to evaluate the suitability for the technical implementation. As a result, the most suitable tool is selected to implement the solution.

During the technical implementation, a solution will be implemented that will make the costs of cloud services more transparent to development teams by providing team-based cost accounting.

Although the current thesis is on the example of Entigo OÜ's customer, the technical solution will be implemented so that it can be used by other companies.

The thesis is in Estonian and contains 49 pages of text, 7 chapters, 21 figures, 1 table.

Lühendite ja mõistete sõnastik

Amazon Athena	<i>Serverless</i> päringuteenus S3 andmete analüüsimiseks
API	<i>Application Programming Interface</i> , rakendusliides
Argo CD	Deklaratiivne GitOps'i pideva tarne tööriist Kubernetesele
ARN	<i>Amazon Resource Name</i> , unikaalne AWS-i ressursside nimi
<i>AssumeRole</i>	API käsk STS kaudu ajutise pääsumandaadi saamiseks
AWS	<i>Amazon Web Services</i> , enimlevinud pilveplatvorm maailmas
AWS CloudFormation	Teenus AWS-i ressursside automaatseks seadistamiseks
AWS Glue	<i>Serverless</i> andmetöötlusteenus
AWS Glue Crawler	Teenus erinevatest allikatest andmete lugemiseks
AWS Glue Database	Konteiner, mis sisaldab tabeleid
AWS Lambda	<i>Serverless</i> sündmustepõhine arvutusteenus
Billing and Cost Management	AWS-i veebiteenus kulude optimeerimiseks
<i>Bucket</i>	<i>Container for objects</i> , konteiner objektide jaoks
CI/CD	<i>Continuous Integration/Continuous Delivery</i> , pidev integreerimine, pidev tarne
<i>Cloud native</i>	<i>Architected to run in a cloud</i> , pilvearhitektuur
CNCF	<i>Cloud Native Computing Foundation</i> , avatud lähtekoodiga pilvandmetöötluse tarkvara sihtasutus
CSV	<i>Comma-Separated Values</i> , komaeraldusega väärtused
CUR	<i>AWS Cost & Usage Report</i> , AWS-i kulu- ja kasutusaruanded
DevOps	<i>Development and Operations</i> , arendus ja käitus
EC2	<i>Elastic Compute Cloud</i> , AWS-i virtuaalserver
ECR	<i>Elastic Container Registry</i> , hallatud kujutisfailide register
EKS	<i>Elastic Kubernetes Service</i> , AWS-i hallatud Kubernetese teenus
<i>Feature Flags</i>	Tarkvaraarenduse tehnika funktsioonide käitusaegseks sisse- ja väljalülituseks ilma tarkvarakoodi muutmata
FinOps	Pilvepõhise finantsjuhtumise kultuuripraktika
<i>Frontend</i>	Tarkvarakomponendi kliendipoolne keskkond
GitOps	Pilvepõhiste rakenduste CI/CD juurutamisviis

gRPC	<i>Google Remote Procedure Call</i> , Google-i poolt arendatud kaugprotseduurikutse
Helm	Kubernetese juurutustööriist rakenduste ja teenuste loomiseks ning seadistamiseks
<i>HorizontalPodAutoscaler</i>	Kubernetese ressurss, mis võimaldab töökoormuse jälgimise abil automaatselt rakendusi skaleerida
HTTPS	<i>Hypertext Transfer Protocol Secure</i> , turvaline hüperteksti edastuse protokoll
IAM	<i>Identity and Access Management</i> , identiteedi- ja pääsuhaldus
<i>Ingress</i>	Kubernetese objekt, mis haldab välist juurdepääsu klastrisisestele teenustele
JSON	<i>JavaScript Object Notation</i> , JavaScripti objektide notatsioon
kaun	<i>Pod</i> , Kubernetese väiksemad ja lihtsamad juurutavad objektid. Sisaldab endas ühte või mitut konteinerit
Kubernetes	Avatud lähtekoodiga süsteem konteinerrakenduste juurutamiseks, haldamiseks ning automatiseerimiseks
Label	Märgend, virtuaalne teabesilt
MIT	<i>Massachusetts Institute of Technology</i> , Massachusettsi Tehnoloogiainstituut
MSK	<i>Managed Streaming for Apache Kafka</i> , AWSi hallatud andmevoo käsitsemise tarkvara Apache Kafka
<i>Policy</i>	<i>An object in AWS that defines resource or entity permissions</i> , AWSi objekt, mis määrab ressursside või olemite õigused
PromQL	<i>Prometheus Query Language</i> , Prometheusi päringukeel
RBAC	<i>Role-Based Access Control</i> , rollipõhine pääsu reguleerimine
RDS	<i>Relational Database Service</i> , relatsioonandmebaasiteenus
<i>Reserved Instance</i>	Arveldussoodustus eeltellitud/reserveeritud virtuaalserveritele
S3	<i>Simple Storage Service</i> , AWSi objektipõhine salvestusteenus
SaaS	<i>Software as a Service</i> , tarkvara teenusena, pilvetarkvara
<i>Serverless</i>	Serverivaba andmetöötlus, kus pilveteenuse pakkuja vastutab koodilõigu täitmise eest
<i>Spot Data Feed</i>	<i>Spot Instance</i> kasutusinfot sisaldav andmevoog
<i>Spot Instance</i>	Virtuaalserveri tüüp, mis võimaldab soodustust kasutust mitte leidnud arvutusvõimsuse näol
SSO	<i>Single sign-on</i> , ainulogimisega pöördus
STS	<i>Secure Token Service</i> , turvatalongiteenus
YAML	<i>YAML Ain't Markup Language</i> , andmete jadastuse keel

Sisukord

1 Sissejuhatus	11
2 Probleemi kirjeldus	13
2.1 Teoreetiline taust	13
2.2 Probleem	14
2.3 Eesmärk	15
2.4 Metoodika	15
3 Loodava lahenduse analüüs	16
3.1 Lahenduse nõuete määramine	16
3.2 Olemasolevad töövahendid	20
3.2.1 AWS Cost Explorer	20
3.2.2 AWS Cost & Usage Report	22
3.2.3 CAST AI	24
3.2.4 Kubecost	26
3.2.5 Harness Cloud Cost Management	28
3.2.6 Crane + FAdvisor	31
3.2.7 Vantage	33
3.2.8 Olemasolevate töövahendite võrdlus ning sobiva töövahendi valik	35
4 Pakutud lahendus	38
4.1 Paigalduskeskkonna taustainfo	38
4.2 Lahenduse arhitektuur	38
4.3 Tehnoloogia valik	41
4.3.1 Tehnoloogia taristu haldamisel	42
4.3.2 Tehnoloogia töövahendi haldamisel	42
5 Tehniline teostus	43
5.1 Kubecosti paigaldamine	43
5.1.1 Mallimine ning manifestide kirjeldamine	43
5.1.2 Töövahendi paigaldamine	44
5.1.3 Töövahendi esmane testimine	44
5.2 AWS kulude sidumine Kubecostiga	45

5.2.1 CUR loomine ning selle integreerimine Amazon Athena teenusega	46
5.2.2 Spot Data S3 ning Spot Data Feed andmevoo seadistamine	47
5.2.3 Kubernetese ja AWS teenuste märgiste lisamine	48
5.2.4 Kulujaotuse märgiste aktiveerimine	51
5.2.5 IAM poliitikate, kasutajate ja rollide loomine	52
5.2.6 IAM rolli sidumine Kubecostiga	53
5.2.7 Töövahendi lõplik seadistus	53
5.3 Tekkinud probleemid ning nende lahendused	54
6 Tulemused	57
6.1 Valitud töövahendi sobilikkuse hindamine	57
6.2 Töövahendi funktsionaalsuse hindamine.....	57
6.3 Võimalikud edasiarendused tulevikus	58
7 Kokkuvõte	59
Kasutatud kirjandus	60
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	67
Lisa 2 – Kubernetese juurutusressursi märgiste lisamise näide	68
Lisa 3 – <i>Spot Data Feed</i> ja IAM pääsumandaadiks vajalike IAM rollide loomise TerraFormi skriptid.....	69
Lisa 4 – Teenuskonto sidumine IAM rolliga ja selle kontrollitud tulemus	70
Lisa 5 – Meeskonnapõhine Kubernetese klastri ning summeritud AWSi teenuste (External veerg) kulujaotus ühe keskkonna kohta kuu lõikes (eemaldatud kuluväärtustega).....	71
Lisa 6 – Meeskonnapõhine detailne AWSi teenuste kulujaotus seitsme päeva lõikes (eemaldatud kuluväärtustega).....	72
Lisa 7 – Rakenduspõhine kulujaotus ühe tunni lõikes (eemaldatud kuluväärtustega) ...	73
Lisa 8 – Rakenduse- ja meeskonnapõhine prognoositud kuluarvestus kuu lõikes, sorteeritud protsessori ressursikasutuse järgi (eemaldatud kuluväärtustega)	74

Jooniste loetelu

Joonis 1. FinOps'i pidev elutsükkel [16].	18
Joonis 2. AWC Cost Exploreri kuluaruanne kasutatud teenuste kohta.	21
Joonis 3. AWS Cost and Usage Reports näidisraportid [86].	23
Joonis 4. CAST AI veebirakenduse ülevaade näidisklastriist [32].	25
Joonis 5. Kubecost veebiliidese avalehe disain.	27
Joonis 6. HCCM näidis Kubernetese kuludest [91].	30
Joonis 7. Crane koondpaneel lisatud klastritest [92].	31
Joonis 8. FAdvisor kulude koondpaneel nimeruumi järgi jeenides [54].	32
Joonis 9. Vantage-i näidisaruanne AWS-i kulude kohta [60].	35
Joonis 10. Kubecost töövahendi komponentide diagramm.	40
Joonis 11. Töövahendi vaikeväärtuste mallimine ning failide lisamine repositooriumisse.	43
Joonis 12. Annotatsioon Kubernetese ressursi kustutamise vastu [79].	44
Joonis 13. Kubecost rakenduse ülevaade Argo CD veebiliidises.	45
Joonis 14. CUR ja Athena integreerimiseks kasutatud CloudFormationi malli loodud ressursid.	47
Joonis 15. Spot Data Feed andmevoo loomine ning selle kirjeldamine JSON-vormingus AWS käsurautiliidi abil [71].	48
Joonis 16. Kubecosti konfiguratsioon väliste teenuste märgiste võtmete defineerimiseks.	49
Joonis 17. Käskude jada ECR repositooriumite märgiste lisamiseks ning ühe repositooriumi märgiste kontroll koos selle väljundiga JSON-vormingus.	51
Joonis 18. Kulujaotuse märgiste aktiveerimine kasutajaliidese kaudu.	52
Joonis 19. Kubecosti diagnostika AWSi teenustega integreerimisel.	54
Joonis 20. AWS Athena päring koos selle kirjelduse ning staatusega.	55
Joonis 21. STSi kaudu identifitseerimine.	56

Tabelite loetelu

Tabel 1. Olemasolevate töövahendite võrdlus lõputöö nõuetega.	37
--	----

1 Sissejuhatus

Pilveteenuseid ning Kubernetesest kasutatakse ettevõtete poolt järjest rohkem. Kubernetes on avatud lähtekoodiga süsteem rakenduse konteinerite juurutamise, skaleerimise ja haldamise automatiseerimiseks [1]. Eurostati 2021. aasta andmetel kasutab 41% Euroopa ettevõtetest pilvetehnoloogiaid, mis on viieprotsendiline kasv võrreldes 2020. aastaga [2]. Sihtasutuse CNCF (*Cloud Native Computing Foundation*) iga-aastase uuringu tulemustest selgub, et Kubernetesest kasutatakse või plaanitakse kasutada 96% vastanute seas [3].

Entigo OÜ kliendiks on tarkvaraarendusega tegelev ettevõtte, kes pakub oma teenust mitmekümnetes erinevates riikides. Ettevõtte kasutab oma infrastruktuurina AWS (*Amazon Web Services*) poolt hallatud Kubernetese teenust EKS (*Elastic Kubernetes Service*) koos teiste AWSi hallatud pilveteenustega. Aprillis 2022 läbiviidud analüüsi põhjal on ettevõttes arendatavaid mikroteenuseid üle 140 ning neid arendatakse kümne meeskonna poolt üheksas erinevas regioonis. Ettevõtte suureks probleemiks on pilveteenuste kulude kontrolli all hoidmine ning selle tagamiseks on vaja selget kuluarvestust. Keerukus lisandub meeskonna- ja rakenduspõhiste kulude arvestusel Kubernetese ressursside näol, mida arvestatakse pilveteenuste pakkuja poolt vaid ühe summeeritud teenuse kuluna. Detailsemaks Kubernetese kuluarvestuseks vajalikke tööriistu pilveteenuste pakkujal ei ole.

Käesoleva lõputöö eesmärk on taustauuringu ning eksisteerivate lahenduste analüüsi käigus leida ning realiseerida tehniline lahendus, mille abil muuta pilveteenuste kulud arendusmeeskondadele läbipaistvamaks, võimaldades näha meeskonnapõhiseid infrastruktuurikulusid eksporditud raportina või koondpaneelina (*dashboard*) mõnes graafilise liidesega keskkonnas, mida hiljem FinOps raamistiku juurutamisel kasutada.

Peatükis 2 tuuakse välja lõputöö teemaga seotud teoreetiline taust, kirjeldatakse probleem, eesmärk ning meetodid kuidas eesmärgi täitmiseni jõuda. Peatükis 3 defineeritakse lahenduse jaoks vajalikud nõuded, analüüsitakse ning võrreldakse olemasolevaid lahendusi nende iseärasuste ja puuduste leidmiseks ning valitakse lõputöö lahenduse jaoks sobiv töövahend. Peatükis 4 tutvustatakse paigalduskeskkonna taustainfot, kirjeldatakse ning põhjendatakse tehnoloogia valikut ning mõtestatakse lahti lahenduse arhitektuur. Peatükis 5 kirjeldatakse detailselt töövahendi paigaldamist, AWSi teenuste kulude liidestamist paigaldatud töövahendiga ning tekkinud probleeme ja nende lahendusi. Peatükis 6 hinnatakse valitud töövahendi sobilikkust lõputöö eesmärgi täitmisel ja selle funktsionaalsusi ning tuuakse välja võimalikud edasiarendused tulevikuks. Peatükk 7 annab ülevaate, mida käesoleva lõputöö käigus tehti, missuguse lahenduseeni jõuti ning kuidas on seda võimalik ettevõtteüleselt ka mujal kasutada.

2 Probleemi kirjeldus

Käesolevas peatükis tuuakse välja lõputööga seotud teoreetiline taust, mis annab ülevaate pilvekulude kontseptsioonist ning sellega kaasnevatest probleemidest ning aitab selle abil lahti mõtestada antud lõputöös käsitletava probleemi ning eesmärgi. Samuti kirjeldatakse metoodika, mille abil lõputöö eesmärgi saavutamiseni jõuda.

2.1 Teoreetiline taust

Lõputöö teema, probleemi ja eesmärgi paremaks mõistmiseks selgitatakse missuguses pilveteenuste keskkonnas käesolev probleem on tekkinud ning kuhu suunas soovitakse edasi liikuda. Alustuseks on hea lahti seletada pilvandmetöötluse termin koos selle kontseptsiooniga.

Pilveandmetöötluses tähendab sõna „pilv“ teenusepakkujat, kes pakub teenuseid interneti kaudu [4]. Pilveandmetöötluse kontseptsioon pärineb 1961. aastast MIT (*Massachusetts Institute of Technology*) professori John McCarthy poolt, kes ütles: „Andmetöötlust võidakse kunagi korraldada üldkasutatava teenusena, nagu telefonisüsteem. Iga tellija peab maksma ainult selle tarbimise eest, mida ta tegelikult kasutab, kuid samal ajal on tal juurdepääs väga suurele süsteemile.“ [5].

Üks esimesi pilvandmetöötluse kontseptsiooni juurutanud ettevõtteid oli Salesforce, kes hakkas 1990-ndate aastate lõpus pakkuma SaaS (*Software as a Service*) pilvetarkvara [4].

Amazon.com Incorporation, kes alustas e-kaubandusega tegelevat ettevõtlust 1995. aastal, kasutas raamatute müümiseks interneti taristut. Kliendid pidasid interneti teel ostmist mugavaks ning kiire laienemise tõttu kasvasid ka nõuded infrastruktuuri suhtes, eriti pühade-eelsetel aegadel. Nende arvutuskeskuste kasutustegur oli madal, alla kümne protsendi ning Amazon nägi kasutamata arvutusvõimsuse müümises ärivõimalust, mille tulemusel loodi 2006. aastal tüdarettevõtte nimega Amazon Web Services [6]. Samal aastal hakkas AWS pakkuma ettevõtetele IT-taristu teenuseid veebiteenusena – praegu tuntud kui pilveandmetöötlus. Täna pakub AWS ülimalt usaldusväärset ja skaleeruvat

taristuplatvormi pilves, mille teenuseid kasutavad sajad tuhanded ettevõtted 190 erinevas riigis [7].

Pilvandmetöötluse kasutuselevõtt on kiiresti kasvav trend. Eurostati andmetel kasutavad 2021. aasta seisuga 41% Euroopa Liidu ettevõtetest pilvandmetöötlust [2].

Pilvandmetöötluse populaarsuse kasvuga suureneb ka *cloud native* arhitektuuri ja tehnoloogiate lähenemine ning kasutuselevõtt, tarkvaraarenduses eelistatakse monoliitsete rakenduste asemel konteinerites mikroteenuseid [8]. Konteinerrakenduste puhul on oluline tähtsus ka sellel, kuidas rakendused omavahel suhtlevad. Selle saavutamiseks kasutatakse süsteeme, mis tegelevad konteinerite haldamise, juhtimise ning automatiseerimisega. Enimlevinud konteinerrakenduste orkestreerimise süsteem on vabavaraline Kubernetes [9]. Datadogi „2021 Container Report“ küsitluse tulemustest selgub, et peaaegu 90% Kubernetese kasutajatest kasutavad pilvhallatud teenuseid, mis on 70% rohkem kui aasta varem. Nendest ligi 30% kasutavad AWSi poolt hallatud Kubernetese teenust EKS (*Elastic Kubernetes Service*) [10].

Pilvetehnoloogia laialdase kasutusega on üks suuremaid probleeme kulude kontrolli all hoidmine. Suurandmete (*Big data*) valdkonnas tuntud ettevõtte Pepperdata avalikustas 750 IT-professionaali poolt vastatud küsitluse tulemused, millest selgus, et enam kui kolmandikul ettevõtetest on pilve-eelarve ületatud kuni 40% ning üks kaheteistkümnest ületab ka selle [11].

Entigo OÜ kliendi näitel (edaspidi ettevõtte) viidi kiirkorras läbi taristukulude analüüs eesmärgiga pilveteenuste kasutuse pealt kokku hoida 10%. Mõningate ülevarustatud (*over-provisioned*) ressursside vähendamise abil sai eesmärk täidetud, kuid see tõstatas teema, kuidas proaktiivselt taristukulud kontrolli all hoida. Üheks tõhusamaks viisiks pilvekulude haldamisel peetakse FinOps kultuuripraktikat. See on tava, mille abil tuuakse pilve muutuvasse kulumudelisse finantsvastutus, võimaldades meeskondadel ise oma pilvekulusid hallata [12].

2.2 Probleem

Lõputöös käsitletakse kõrgete pilveteenuste kulude probleemi ettevõtte näitel, mille lahendamiseks pakub FinOps kultuuripraktika välja finantsvastutuse viimise

meeskondadeni [13]. Tegemist on mitmeetapilise protsessiga, mille esimene etapp on meeskonnapõhiste kulude välja näitamine.

AWSi hallatud teenuste puhul on küll võimalik kasutada märgiseid (*tag*) ning selle abil kulud AWSi töövahenditega meeskonna vaates raporteerida, kuid peamine probleem seisneb EKS teenuses, mida AWSi kulude vaates raporteeritakse kui ühte konkreetset teenust. Tegelikult kasutatakse Kubernetese ressursse 140 rakenduse ja kümne meeskonna poolt, mille kulusid ei ole hetkel võimalik meeskondade lõikes eristada.

2.3 Eesmärk

Lähtudes varasemalt välja toodud probleemist, on lõputöö eesmärk muuta pilveteenuste kulud arendusmeeskondadele läbipaistvamaks läbi meeskonnapõhise kuluarvestuse ning raporteerimise. Selleks on vaja leida ja teostada tehniline lahendus, mis võimaldab meeskonnapõhist kuluarvestust Kuberneteses ja ülejäänud AWSi teenustes. Kuigi probleemi käsitletakse ettevõtte näitel, on oluline, et lahendus oleks ettevõtteülene ning seda saaks vajadusel kasutada ka teiste Entigo OÜ klientide jaoks.

2.4 Metoodika

Eesmärgi täitmiseks teostatakse analüüs, mille abil kirjeldatakse vajadused ning nõuded, millele loodav lahendus vastama peab. Tehakse taustauuring juba olemasolevate lahenduste leidmiseks ning nende ülevaate käigus analüüsitakse, kas need vastavad lõputöös paika pandud nõuetele või mitte. Pärast olemasolevate lahenduste ülevaadet ning analüüsi, teostatakse kokkuvõtlik võrdlus, kus hinnatakse töövahendite omadusi ning valitakse sobivaim töövahend, mille abil realiseeritakse tehniline lahendus.

3 Loodava lahenduse analüüs

Selles peatükis analüüsitakse erinevaid võimalusi lõputöö probleemi lahendamiseks. Võrreldakse pilveteenuse pakkuja AWSi olemasolevaid kuluahalduslahendusi ning kolmanda osapoole nii vabavaralisi kui ka omandatarkvaralisi lahendusi pilveteenuste kulude haldamiseks. Saadud tulemuste abil tehakse kindlaks, kas olemasolevad lahendused vastavad lõputöös seatud nõuetele ning on võimalik ettevõtte puhul kasutusele võtta või tuleb sobiv lahendus ise välja töötada.

3.1 Lahenduse nõuete määramine

Nõuete määramisel on arvestatud, et probleemi käsitletakse ettevõtte näitel. Nõuded pannakse paika kliendi soovide ja vajaduste järgi, et parendada kontrolli AWS kulude kasvu osas.

Selle saavutamiseks viidi 2022. aasta esimeses kvartalis läbi hoogtöö kokkuhoiukohtade leidmiseks ning selle tulemusel saavutati umbes 20% pilveteenuste kulude vähenemine kuu lõikes.

Selleks, et muuta kulud läbipaistvamaks ning motiveerida meeskondi ise oma kulusid jälgima, tunneb ettevõtte huvi FinOps kultuuripraktika rakendamise vastu ning sellest tingituna seati 2022. aasta teise kvartali eesmärgiks meeskonnapõhise raporteerimise võimekuse arendamine ja meeskondadele kulude nähtavaks muutmine.

FinOps'i kultuuripraktika põhilistest nõuetest ja aspektidest aru saamiseks, on vaja lahti mõtestada selle elutsüklil koos lühida kirjeldusega igast etapist. FinOps'i elutsüklil koosneb kolmest peamisest etapist, mida pidevalt läbitakse. Nendeks etappideks on *inform*, *optimize* ja *operate*.

Inform – see on FinOps'i teekonna esimene etapp, mis annab organisatsioonidele ja meeskondadele ülevaate kuludest ja võimaldab eelarve koostamist ning kulude prognoosimist [14]. Selles etapis hakatakse aru saama oma kuludest ning sellest, mis neid kulusid põhjustab. Andes meeskondadele peaaegu reaalajas ülevaate nende kuludest, luuakse jagatud vastutus, mille abil saavad meeskonnad teada, mille peale nad täpselt kulutavad ning miks. Jagatud ressursside kulud jaotatakse sobivalt – võrdselt või arvutatud kasutusmõõdikute järgi. Luuakse tulemuskaardid (*scorecards*), mis

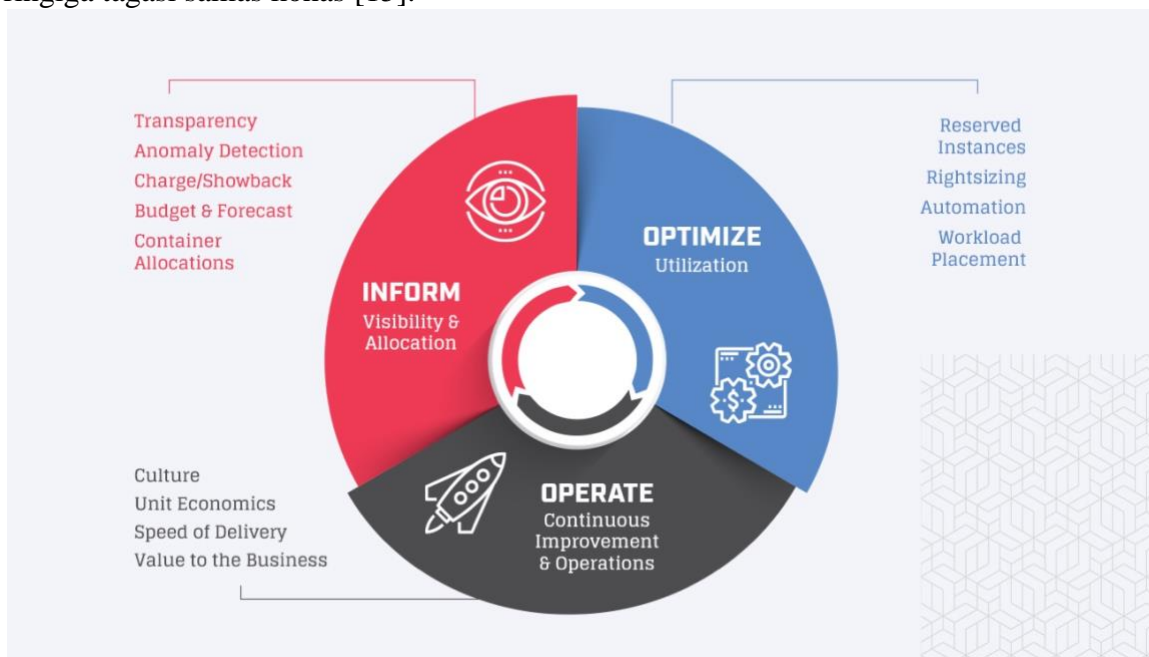
võimaldavad meeskondadel vaadelda, kuidas läheb erinevates projektides kulude, kiiruse ja kvaliteedi optimeerimine. Kulukaartide abil on võimalik tõhusalt leida valdkondi või teenuseid mida on võimalik optimeerida [15].

Optimize – FinOps'i teekonna teine etapp, mis keskendub kulude vältimisele ning kulude optimeerimisele. Selles etapis tuvastatakse alakasutatud teenused, anomaaliad ning koostatakse sellekohased raportid. Kasutamata ressursse on võimalik tuvastada nii serveripargi, andmebaaside, salvestusseadmete kui võrguseadmete põhjal.

Hinnatakse *reserved instance* või *spot instance* tüüpi serverite kasutuselevõtu tasuvust ning muid kasutusalaalseid allahindlusi. Võrreldakse hindu ja töökoormuse paigutust ning vajadusel vaadatakse erinevate pilveteenuste pakkujate poole, et võrrelda hinnavalikuid [15].

Operate – kui optimeerimise etapp seab eesmärgid parendamiseks, siis käitamisfaas seab paika vajalikud protsessid. Siin ei seata eesmäärke, vaid tehakse otsuseid ja plaane varem seatud eesmärkide saavutamiseks. Selleks võib olla optimeerimise etapis tuvastatud ressursside alakasutusest tingitud serveritüüpide vahetamine odavamate vastu, mõne projekti puhul pooljuhtketta-põhise (*solid state drive*) salvestusseadme asendamine odavamaga magnetkõvaketta vastu. See etapp rõhutab ka protsesside pidevat täiustamist ning ressursside optimeerimise automatiseerimist [15].

Joonisel 1 on välja toodud FinOps'i elutsükkel, mida saab mõtestada kui pidevat ringikujult võidusõidurada – võimalik igal ajal või kohas teele asuda ning lõpuks olete ringiga tagasi samas kohas [15].



Joonis 1. FinOps'i pidev elutsükkel [15].

Antud lõputöö raames keskendutakse FinOps'i raamistikus *inform* etapile, mis rõhub peamiselt sellele, et meeskondadel oleks peaaegu reaalajas ülevaade ning kontroll oma pilvekasutuse kulude üle.

Tuginedes FinOps'i esimesele etapile *inform*, on oluline paika panna märgistamisstrateegia, mis on kulude jaotamisel üks olulisemaid nõudeid. See võimaldab lisaks tavapärasele kulujaotusele paika panna strateegia jagatud ressursside kulu osas [13], mis mängib koos märgiste strateegia olulist rolli lõputöö nõuete määramisel.

FinOps'i raamistiku järgi on vajalik finantsvastutuse viimine meeskondadeni [16]. Ka lõputöö eesmärgist tuleneva vajaduse järgi ei piisa tüüpilistest kuluraportitest, on oluline, et kulusid saab jagada meeskonnapõhiselt nii Kuberneteses kui ka ülejäänud AWSi teenustes, kaasa arvatud salvestusseadmete, koormusjaoturite ning võrguliikluse kulude näol.

Finantsvastutuse toomine meeskondadeni eeldab ka õigeaegset ning pidevalt kuluraportite väljanäitamist kõigile meeskondadele või äriüksustele. On oluline, et

meeskonnad saaksid soovi korral endale kohandatud aruandeid luua ning tekiks võimalus olemasolevate andmete pealt eelarveid koostada ja prognoose kuvada [16].

Kuna ettevõttel on kasutuses mitu erinevat regiooni koos samade meeskondade poolt arendatud rakendustega, siis on vaja meeskonnapõhist kulude jaotust regioonide üleselt, kuid soovitatavalt võimalusega välja näidata iga konkreetse regiooni kulu.

Meeskonnapõhiste kulude tervik arvestatakse kokku AWS-i teenuste ning Kubernetesse paigaldatud rakenduste summana, mistõttu oleks hea, kui summeerimist ei peaks käsitsi läbi erinevate keskkondade tegema, vaid platvorm võimaldaks AWS-i kulusid läbi API (*Application Programming Interface*) päringute või S3 (*Amazon Simple Storage Service*) kaudu lugeda.

Juriidilistest asjaoludest tulenevast ajakulust ning keerukusest, on ettevõttes SaaS teenuste kasutamine välistatud. Lähtudes ettevõtte varasemast kogemusest, pikeneks protsess mitme kuu võrra ning välistaks lõputöö eesmärgi ja ettevõtte teise kvartali eesmärgi õigeaegse teostuse.

Kuna töövahendid paigaldatakse oma infrastruktuurile, oleks hea järgida juba ettevõttes varasemalt kasutuses olevaid parimaid tavasid ning mustreid, millest üks on GitOps raamistiku kasutus.

Oluline on tähelepanu pöörata ka andmete püsitalletuse võimalusele, mis tagab, et töövahendi taaskäivitusel ei teki andmekadu ning millega hoitakse ära üleliigne ajakulu andmete uuesti alla laadimisel.

Nõuete määramisel võetakse arvesse, et probleemi käsitletakse ettevõtte näitel ning osasid nõudeid võivad mõjutada ettevõtte spetsiifika. Lähtutakse FinOps'i elutsükli poolt paika pandud soovitustest ning nõuetest, et finantsvastutus meeskondadeni tuua. Samuti arvestatakse nõuete defineerimisel funktsioonidega, mis võimaldaksid ettevõttes lõputöö tulemi kasutusele võtta ja jätkata järgmiste FinOps etappide juurutamisega tulevikus järgmiste FinOps'i etappideni jõuda.

Lõputöö nõuete kokkuvõte

Nõuded loodavale lahendusele on järgnevad:

- Lahendus võimaldab koondada kulusid erinevatest Kubernetese klastritest ja pilvekeskkonna regioonidest.
- Lahendus peab toetama kulude eristamist märgiste strateegia alusel.
- Lahendus peab võimaldama meeskonnapõhist kulude jaotust nii Kuberneteses kui ka ülejäänud AWSi kulude kombineerimisel.
- Lahendus võimaldab AWSi kulusid läbi API päringute lugeda ning need koos Kubernetese kuludega summeerida.
- Lahendust peab olema võimalik paigaldada ettevõtte infrastruktuurile.
- Lahendus peab võimaldama rakenduspõhist kuluarvestust Kubernetese ressursside kasutamise kohta.
- Lahendus peab toetama püsitalletust (teenuse taaskäivitusel peavad varasemad andmed säilima).
- Lahendus on kättesaadav kõigile ettevõtte meeskondadele.
- Lahendus on kasutatav erinevates keskkondades.
- Meeskondadel on võimalus endale sobivaid koondpaneele luua.

3.2 Olemasolevad töövahendid

Käesolevas peatükis antakse ülevaade erinevatest töövahenditest, mida lõputöö eesmärgi saavutamiseks võrreldakse. Töövahendite omadusi ning puudusi võrreldakse lõputöös defineeritud nõuetega, mille abil hiljem sobiv töövahend välja valitakse.

3.2.1 AWS Cost Explorer

AWS Cost Explorer on üks mitmetest erinevatest AWS-i poolt pakutud kuluhaldusteenustest. See on tööriist, mis võimaldab ülevaadet ja analüüsi pilveteenuste kulude ning kasutuse üle mitmete vaikearuannete komplektide abil. Cost Explorer veebiteenuse kaudu kuvatakse kuni viimase 12 kuu kulude ülevaade või järgneva 12 kuu kuluprognos [17].

AWS-i pilveteenuste kulusid ja kasutust on võimalik vaadata kas läbi Cost Explorer-i kasutajaliidese või kasutades selleks API päringuid. Veebiliidese kaudu on päringute tegemine tasuta, kuid API päringuid tehes kaasneb tasu 0,01 dollarit päringu kohta [18].

Cost Explorer võimaldab varasemalt Billing and Cost Management veebiteenuse kaudu aktiveeritud kulude jaotamise siltide (*cost allocation tags*) abil kuvada detailsemat ülevaadet kasutuskulude kohta [19]. Võti-väärtus paaride siltide kasutamine võimaldab eristada erinevate teenuste ja meeskondade kulusid Cost Explorer-i raportites.

Joonisel 2 on ekraanitõmmis AWS Cost Exploreri viimase kuue kuu teenuste aruandest, mis visualiseerib selgelt kõige suuremad kuluallikad.



Joonis 2. AWS Cost Exploreri kuluaruanne kasutatud teenuste kohta.

Lähtudes käesoleva lõputöö nõuetest, hinnatakse AWS Cost Exploreri võimalusi antud lõputöö probleemi lahendamiseks.

Olgugi, et kulude jaotamise siltide abil on võimalik kulusid eristada, tuleb tähelepanu pöörata sellele, et kui AWSi teenused asuvad erinevates kontodes, on võimalik kulusid näha kas iga konto alt eraldi või AWSi organisatsiooni juurkasutaja alt kõikide kontode kohta. Meeskondadele on võimalik anda AWS Cost Exploreri ligipääs läbi IAM (*Identity and Access Management*) poliitikate ja kasutajate, kuid kui AWSi organisatsiooni liikmekontosid (*member account*) on mitmeid, siis eraldi iga konto ligipääsu haldamine lisab keerukust ning on aeganõudev tegevus. Tuginedes AWSi turvalisuse parimatele tavadele, on rangelt soovitatud vältida juurkasutajat igapäevasteks haldustoiminguteks [20], mis välistab juurkonto kaudu kuluraportitele ligipääsu andmise.

Ettevõtte kontekstis AWS Cost Explorer Kubernetese kulude arvestuseks ei sobi, sest kuluraportil näidatakse välja vaid EKS teenuse kogukulu, mitte detailsemat kulu kasutatud ressursside järgi.

3.2.2 AWS Cost & Usage Report

AWS Cost & Usage Report (edaspidi CUR) sisaldab endas kõige põhjalikumat AWSi kulu- ja kasutusandmete kogumit, sealhulgas täiendavaid metaandmeid AWSi teenuste, krediidi, tasude, maksude, allahindluste, kulukategooriate ning reserveeritud virtuaalserverite (*reserved instances*) kohta [21].

CUR võimaldab kulu- ja kasutusaruannete salvestamist S3 *bucket*-i abil. Aruanded võimaldavad kulude ülevaadet toodete, tooteressursside, kasutajate poolt defineeritud siltide (*tag*) alusel ja ajaliselt tunni, päeva või kuu alusel. Aruannet värskendatakse AWSi poolt kord päevas ning aruanne on kättesaadav AWS CUR kasutajaliidese kaudu CSV (*Comma-Separated Values*) vormingus.

Aruandeid on võimalik alla laadida ja kuvada arvutustabelitarkvara abil või kasutades päringute jaoks S3 APIt. CUR aruandeid luuakse, loetakse, muudetakse ja kustutatakse AWSi kasutajaliidese või AWS CUR API kaudu [22] ning aruannete kasutamiseks on vajalik AWS S3 *bucket*-i olemasolu, kuhu aruanded salvestatakse. Sellest tingituna on aruandluse esmasel seadistusel nõutud olemasoleva S3 *bucket*-i defineerimine või uue loomine.

Joonisel 3 on näha AWS CUR raportite kasutajaliidese vaade, kus kuvatakse olemasolevaid raporteid ning võimalusi uute loomiseks.

Dashboard
Bills
Cost Explorer
Budgets
Reports
Cost Allocation Tags
Payment Methods
Payment History
Transactions (beta)
Consolidated Billing
Preferences
Credits
Tax Settings

AWS Cost and Usage Reports

As of November 15, 2018, you can now configure your Cost and Usage Report to integrate with Amazon Athena, making it easy to analyze your billing data using standard SQL queries. [Learn more](#)

[Create report](#) [Edit](#) [Delete](#)

<input type="checkbox"/>	Report name	S3 bucket	Time granularity	Data last refreshed
<input type="checkbox"/>	Hourly_CostUsageReport	reporting	Hourly	02/06/2019, 7:59 AM
<input type="checkbox"/>	Daily_CostUsageReport	reporting	Daily	02/06/2019, 7:58 AM

Other Reports

- [Analyze your cost and usage using AWS Cost Explorer](#)
AWS Cost Explorer lets you dive deeper into your cost and usage data to identify trends, pinpoint cost drivers, and detect anomalies. [Learn more](#)
- [Monitor your Reserved Instance \(RI\) utilization using AWS Cost Explorer](#)
This report allows you to visualize your RI utilization, providing insight into opportunities for increasing your RI usage efficiency. [Learn more](#)
- [AWS Usage Report](#)
A CSV report that provides usage data for select AWS services.

Joonis 3. AWS Cost and Usage Reports näidisraportid [85].

Kuigi CUR kujutab endas kõige põhjalikumat kulu- ja kasutusandmete kogumit [21] ning võimaldab siltide abil kulude eristamist, ei võimalda see detailset Kubernetese kulude ülevaadet ega suuda kulusid visualiseerida, vaid kasutab selleks pilvepõhist ärianalüüsi (*business intelligence*) teenust AWS QuickSight või kolmanda osapoole tarkvara, mis on suuteline CUR andmeid lugema ja visualiseerima [23]. AWS QuickSight teenus on tasuta ning tuginedes AWSi hinnakalkulaatorile, tuleks 30 mitteregulaarse kuluraportite lugeva ja viie kuluraporteid koostava meeskonnaliikme eest igakuise maksumusena tasuda ligikaudu 125 dollarit [24].

Lähtudes taaskord AWSi turvalisuse parimatest tavadest, ei ole mõistlik AWS QuickSight teenust kasutada juurkasutajana [20], mis tähendaks, et sarnaselt AWS Cost Explorerile, tuleks ülevaate saamiseks tagada kuluraportite ligipääsud AWS kontode põhisel.

Kuigi CUR kasutamine AWSi liikmekontodel on lubatud 2020. aasta detsembrist [25], tähendaks see ettevõtte kontekstis, kus liikmekontosid on üle kümne, olulist IAM ligipääsuõiguste halduse suurenemist. Lisaks tuleks kuluraportid igast keskkonnast kokku koguda ning need seejärel kusagil mujal summeerida.

3.2.3 CAST AI

CAST AI on täisteenust pakkuv pilveautomaatika platvorm, mis pakub võimsaid automatiseerimisfunktsioone Kubernetese töökoormuse optimeerimiseks. CAST AI meeskonna väitel võimaldab nende platvorm erinevate tööstusharude ettevõtetel säästa 50–90% oma pilvekuludest [26].

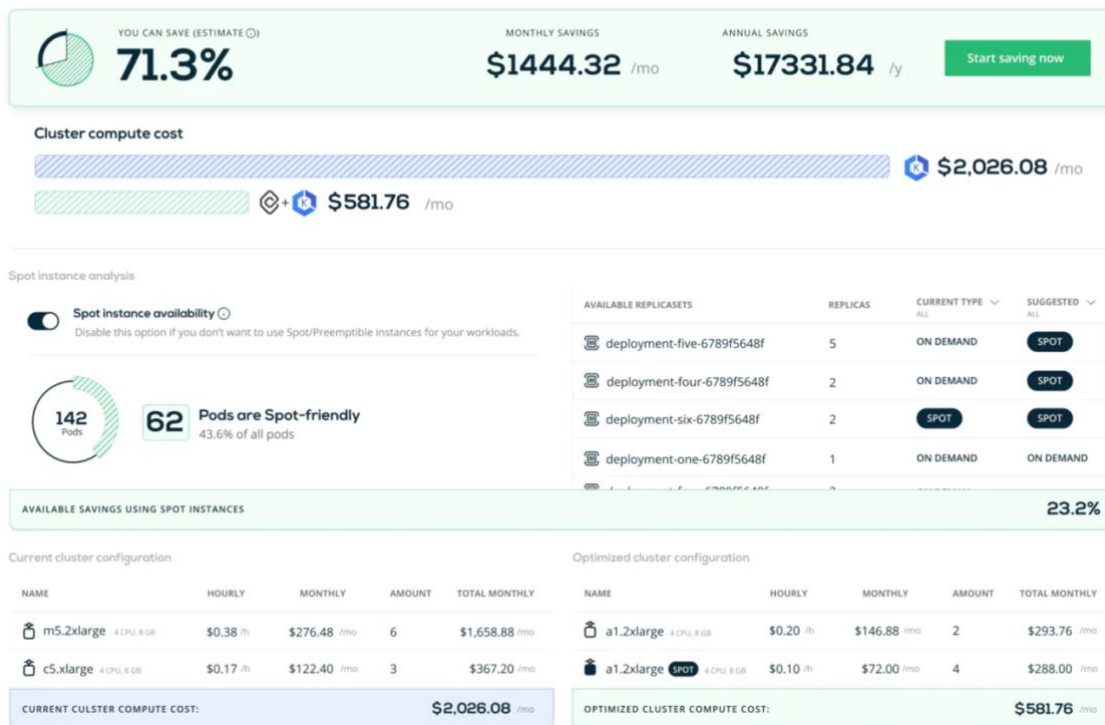
CAST AI lahendus pakub Kubernetesega seotud ressursside ja kulude optimeerimist, kasutades selleks masinõppe abil juhitavat optimeerimismootorit, mis rakendab Kubernetese klastritega seatud muudatusi reaalsajas [27]. CAST AI võimaldab Kubernetese klastris erinevate ressursside skaleerimist, kasutades selleks HorizontalPodAutoscaler objekte [28].

Hinnakirja alusel pakutakse nelja põhikategooriasse kuuluvat teenust, milleks on Starter, Growth, Growth PRO ning Enterprise. Ainus tasuta kategooria on Starter, mis pakub masinõppe alusel teostatud Kubernetese klastri analüüsi ning piiramatus koguses kuluaruandlust. Ülejäänud lisade ja rohkemate funktsioonide kasutamiseks on vajalik kasutada tasulisi kategooriaid [29].

Käesoleva lõputöö eesmärgi on saada ülevaade kuludest, mitte optimeerida Kubernetese klastris ressursside kasutust ning seetõttu on antud töö raames piisav *Starter* kategooria. Kubernetese ressursside ülevaate saamiseks on vajalik Kubernetese klastrisse paigaldada avatud lähtekoodiga pistikprogramm *castai-agent* [30]. Pistikprogrammi paigaldamiseks käivitatakse lühike skript, mille käigus luuakse selle tööks vajalikud Kubernetese ressursid.

Pärast edukat paigaldamist on võimalik CAST UI veebirakenduse kaudu kontrollida oma Kubernetese klastriga seotud detailset kuluaruandlust. Joonisel 4 on näha CAST AI avalehe disaini. Avalehel kuvatakse Kubernetese klastrisest kulude aruanne koos nende võimalike optimeerimismeetoditega.

Available savings



Joonis 4. CAST AI veebirakenduse ülevaade näidisklastrist [31].

Lähtudes taaskord lõputöö nõuetest ning võrreldes neid CAST AI dokumentatsiooni järgi kirjeldatud võimalustega, jõuti järelduseni, et antud töövahend võimaldab küll Kubernetese klasteri detailset kuluarvestust, kuid ülejäänud pilveteenuste kuluarvestust see ei paku [29].

Tegemist on SaaS tüüpi töövahendiga, mis ei ole lõputöö nõuete järgi sobiv lahendus, sest ei võimalda oma taristule paigaldamist. See-eest võimaldatakse ühe konto alla koondada mitmeid Kubernetese klastreid koos võimalusega kulused klasteripõhiselt agregeerida, mis vastab ühele lõputöö nõuetele [26].

Tuginedes CAST AI poolt väljastatud võrdlusele konkureeriva töövahendiga selgus, et erinevalt konkureerivast töövahendist ei ole võimalik Kubernetese kulused meeskonnapõhiselt koondada, vaid võimaldatakse kuluarvestust nimeruumi, juurutuse (*deployment*) või teenuse põhisele [26].

Alates 2022. aasta jaanuarist on CAST AI väljalasketate (*release notes*) kohaselt võimalik kuluarporteid eksportida ka tabelarvutusprogrammidele sobival CSV vormingus [32].

3.2.4 Kubecost

Kubecost on vabavaraline avatud lähtekoodiga tarkvaraprojekt Kubernetese kulude optimeerimiseks, mis kasutas oma esimeses iteratsioonis Karl Stoney poolt 2018. aastal välja töötatud Kubernetese kulude ülevaadet Prometheusi ja Grafana monitooringu pinu (*stack*) baasil [33].

Pärast aastapikkust arendust ja täiustamist, toodi 2019. aastal Ajary Tripathy ja Webb Browni poolt turule Kubecost, mis võimaldab reaalsajas Kubernetese kulude jaotust nimeruumi, teenuste, konteinerite, meeskondade, osakondade või toodete alusel.

Täielikuks kulude ülevaate tagamiseks seotakse Kubernetese klastrisisesed ressursikulud (keskprotsessor, muutmälu, salvestusseadmed, koormusjaoturid jms.) pilveplatvormide klastriväliste kuludega nagu RDS (*Relational Database Service*), MSK (*Managed Service for Apache Kafka*) ja S3, kasutades selleks pilveplatvormide arveldusteenuste APIsid [34].

Kuigi Kubecosti puhul on tegemist avatud lähekoodiga vabavaraga, jagunevad töövahendi kategooriad kolmeks, milleks kaks on tasulised:

- Free – tasuta kategooria, mis võimaldab ühe Kubernetese klastri lisamist töövahendi kohta, 15-päevast andmesäilitust, kulujaotust ning monitooringut koos prognoosidega, koondatud Kubernetese ja ülejäänud pilveteenuste kulud (AWS, Microsoft Azure ja Google Cloud Platform pilveplatvormidel)
- Business – lisaks Free kategooriale piiramatu arv Kubernetese klastreid ühes kasutajaliideses, 30-päevane andmesäilitus, garanteeritud teenustasemelepe.
- Enterprise – lisaks Business kategooriale SSO (Single Sign-On) tugi, piiramatu andmesäilitus, kohandatud ressursside hinnakujundus.

Kubecosti vaikimisi installeerimise käigus paigaldatakse *DaemonSet* tüüpi ressurss, mis tagab, et klastri kõikide sõlmede (*node*) peal käivitatakse võrguliikluse meetrikat koguv ning kogutud meetrikat agregeerimis- ja analüüsimiskeskonda edastav kaun. Kogutud meetrika analüüsi käigus saadakse ülevaade kolmest erinevast võrgukulu liigist:

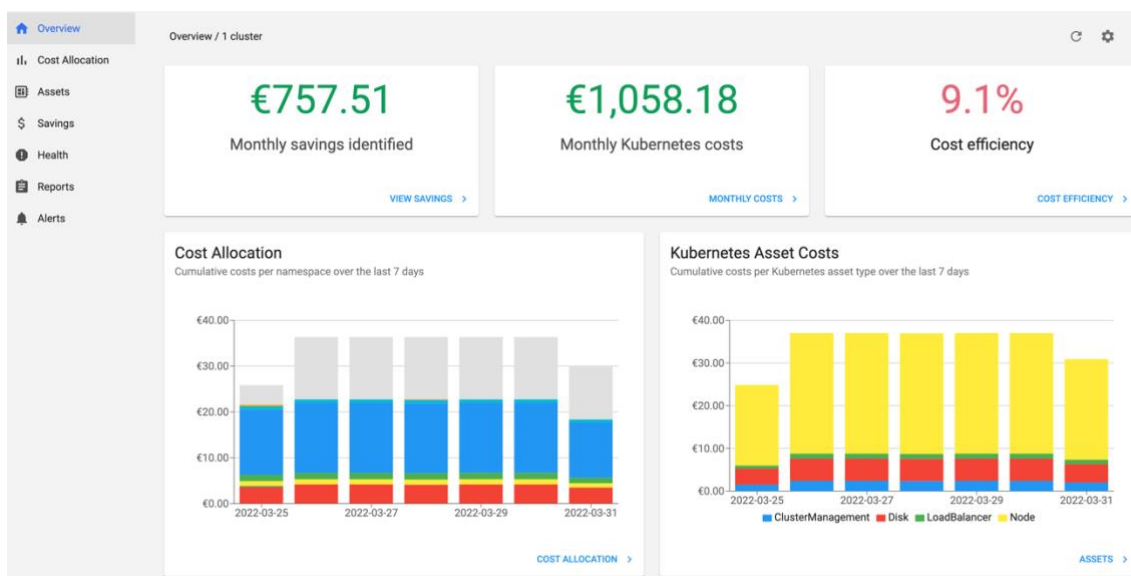
- *Internet Egress* – võrgu sihtkoht on klastriväline.

- *Cross Region Egress* – võrgu sihtkoht on klastrisisene, kuid ei asu samas pilveplatvormi regioonis.
- *Cross Zone Egress* – võrgu sihtkoht on klastrisisene ning samas pilveplatvormi regioonis, kuid erinevas pilveplatvormi tsoonis.

Selline klassifitseerimine on oluline, sest sellised võrgukulud korreleeruvad enamiku pilveteenuste pakkujate võrgumudelitega [35]. Kubecosti tutvustuslõikude järgi genereeritakse automaatne statistika, mille abil võimaldatakse 30-50% kokkuvõtte infrastruktuurikuludelt vältides ressursside üle- või alavarustamist [36].

Ametliku Kubecosti dokumentatsiooni järgi võimaldatakse aktiivset hoiatuste (*alerts*) edastamist Prometheus Alertmanager ja Slack teenuste abil [37].

Joonisel 5 on näha Kubernetesi kulujaotuse ülevaadet Kubecosti veebiliidese kaudu.



Joonis 5. Kubecost veebiliidese avalehe disain.

Lisaks Kubecosti veebiliidesele, eraldatakse *cost-model* eksportija käitamisega erinevaid kulumõõdikuid (*cost metrics*), mida on võimalik Prometheusi seireteenusesse edastada ning PromQL päringute kaudu Grafana teenuse abil visualiseerida [38].

Võrreldes lõputöö nõudeid Kubecosti dokumentatsiooniga selgus, et antud töövahend võimaldab kulude eristamist märgistamisstrateegia järgi [39]. Seejärel on kulud võimalik agregeerida varasemalt loodud märgistuste järgi selliselt, et Kubernetesi ressursside ning ülejäänud AWSi teenuste kulud kuvatakse summaarselt, mis vastab

kahele lõputöö nõudele [40]. Märjiste strateegia abil on võimalik koondada lisaks Kubernetese rakenduse otsestele ressursikuludele ka lisakulud salvestusseadmete, koormusjaoturite ning võrguliiklusest tingitud kulude kohta.

Kubernetese kulude summeerimiseks ülejäänud AWSi teenuste kuludega kasutatakse taustal AWS CUR teenust, mille käigus loetakse API päringute kaudu S3 *bucket*-ist kuluandmed. Saadud sisendi abil teostatakse vajalikud päringud AWS Athena teenusele, mille tulemused kirjutatakse omakorda Kubecosti S3 *bucket*-isse. Seejärel edastatakse tulemused API päringute Kubecost, kus need veebiliidese koondpaneelil visualiseeritakse [41].

Antud töövahendi puhul installeeritakse see oma infrastruktuurile, nii nagu on loodava töövahendi nõuete puhul defineeritud. Pärast Kubecosti paigaldust on ligipääs veebiliidesele, millega tuuakse meeskondadeni nende pilvekasutuse kulud koos võimalusega endale sobivate koondpaneelide loomisega.

Tasuta versiooni puhul võimaldatakse kogu AWSi organisatsioonikulude kuvamist CUR andmete järgi ning nende eristamist meeskondade ja muude varasemalt defineeritud märjastuste alusel, kuid piiranguna sätestatakse vaid ühe Kubernetese klatri lisamine kasutajaliidesele. Tasuta versiooni kasutamine eeldab kõigi klatriite kulude ülevaate saamiseks Kubecosti veebiliidese paigaldamist kõikidesse klatriitesse.

Tasulise versiooni puhul võimaldatakse mitme Kubernetese klatri koondamist ühte Kubecosti veebiliidesele [42].

3.2.5 Harness Cloud Cost Management

Harness on tarkvara tarne platvorm, millega kasutatakse DevOps'i protsesside lihtsustamiseks masinõpet. Harnessi platvormi kuuluvad CI/CD (*Continuous Integration/Continuous Delivery*) ahela jaoks vajalikud tooted, pilvekulude haldus, Feature Flags haldus ja palju muud [43]. Antud lõputöö raames analüüsitakse Harnessi pilvekulude halduse toodet Harness Cloud Cost Management (edaspidi HCCM), mis pakub erinevate võimaluste, lisade ja kasutajatoe astmetega kolme hinnakategooriaga toodet: – Free, Team ning Enterprise [44].

- Free – Kuni 250 000 dollari suuruse pilvekulude haldus ja optimeerimine, koondpaneelid, ühe kuu pikkune andmesäilitus, pilvekulude optimeerimine

automaatse peatamise (Intelligent Cloud Auto Stopping) abil, integreeritud platvormi CI/CD ja Feature Flags toodetega.

- Team – Lisaks tasuta kategooriale piiramatute pilvekulude haldus, klastrite monitooring, kohandatud koondpaneelid, viie aasta pikkune andmete säilitamine, garanteeritud teenustasemelepe.
- Enterprise – lisaks Team kategooriale mitme organisatsiooni tugi, SSO (Single Sign-On) tugi, revisjonipäevikud (audit trails), IP-põhiste piirangute (whitelisting) haldus, kohandatud rollid ja ligipääsude haldus.

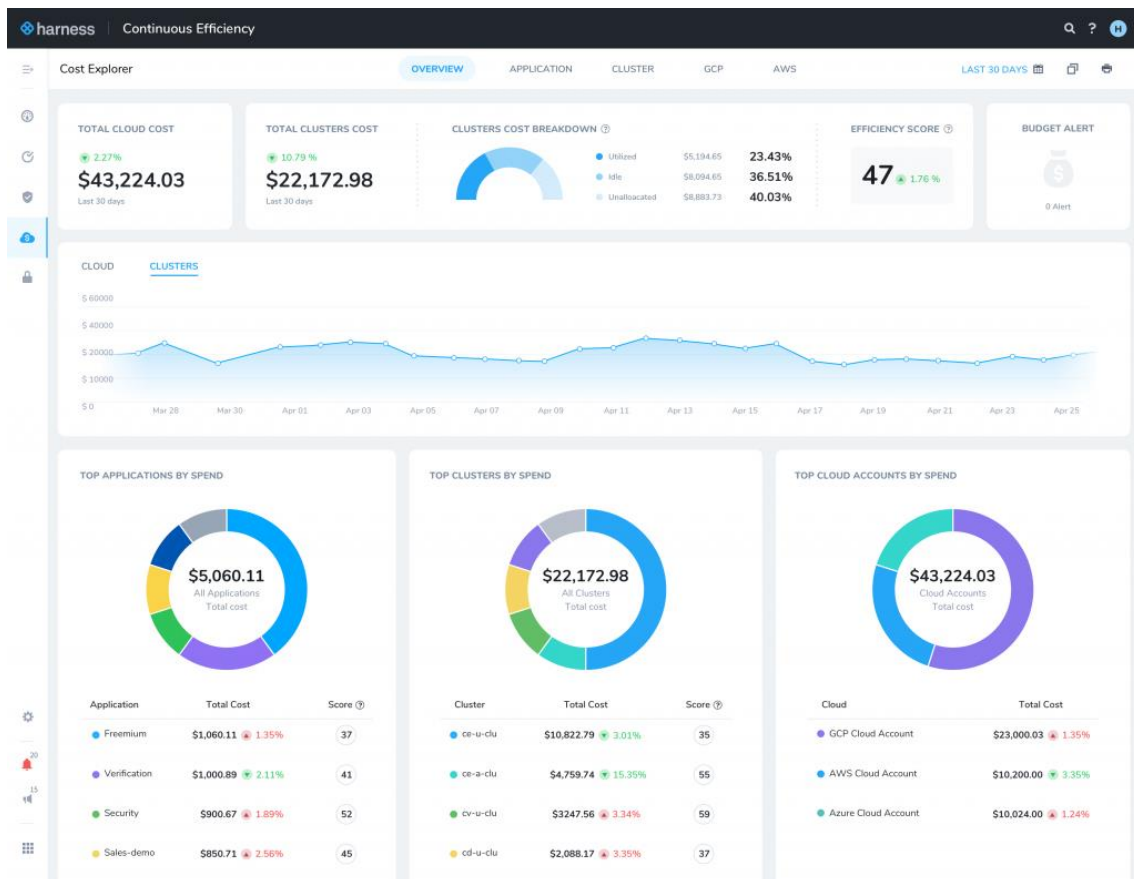
HCCM võimaldab pilvekulude haldust kasutades pilveplatvormidega suhtlemiseks API päringuid. Kuluhalduse minimaalseks seadistamiseks AWSi pilveplatvormil, on vajalik täita mõningad eeltingimused [45]:

- CUR ja selle loomisega kaasnev S3 bucket.
- IAM (Identity and Access Management) poliitikate ja rollide loomine CUR raportite lugemispääsuks ning Harness S3 bucket-isse kirjutuspääsuks.

Seejärel seadistatakse pilveplatvormi kontole ligipääs, defineerides varasemalt loodud CUR raportite S3 *bucket*-i nime ja IAM rolli ARNi (*Amazon Resource Name*). Sellega võimaldatakse ülevaade AWS-i ressurssidest ning nende kasutusest.

Kubernetesi klastrite täpsemaks kulude ülevaateks on eeltingimusteks Harness Delegate-i ja Kube Metrics Serveri pistikprogrammide varasem paigaldus. Harness Delegate-i paigaldusel on võimalik nimeruumipõhised õigused, nimeruumiülene lugemisõigus või nimeruumiülene (üle kogu Kubernetesi klastrite) lugemis- ja kirjutusõigus. Seejärel paigaldatakse Kubernetesi *StatefulSet* tüüpi ressurss, mis suhtleb läbi HTTPS (*Hypertext Transfer Protocol Secure*) ja gRPC (*Google Remote Procedure Call*) protokollide Harnessi lõppseadmetega. Lõppseadmetega suhtluseks on vajalik väljuva võrguliikluse lubamine Harnessi IP aadressidele [46].

Joonisel 6 on näha HCCM veebirakenduse disaini, kus on kuvatud selge ülevaade Kubernetesi klastrite kogukulust ning detailsem ülevaade rakenduste kuludest.



Joonis 6. HCCM näidis Kubernetese kuludest [89].

Kuigi HCCM võimaldab Kubernetese ressursside kuluülevaadet klastrisse paigaldatud Kube Metrics Server ja Harness Delegate pistikprogrammide abil ning AWSi teenuste kulusid loetakse läbi API päringute, on vajalik ettevõtte juurkonto sidumine kolmanda osapoole kontoga, kasutades selleks kontoüleseid IAM rolle [47].

Harnessi dokumentatsioonist selgub, et töövahend võimaldab agregeerida informatsiooni erinevatest klastritest ja regioonidest ning kulusid klastri, nimeruumi või teenuse järgi filtreerida [48], mis vastab kahele lõputöö nõudele. Kahjuks ei ole toetatud kulujaotus märgistussiltide alusel, mis välistab sellisel kujul meeskonnapõhise kulujaotuse võimaluse.

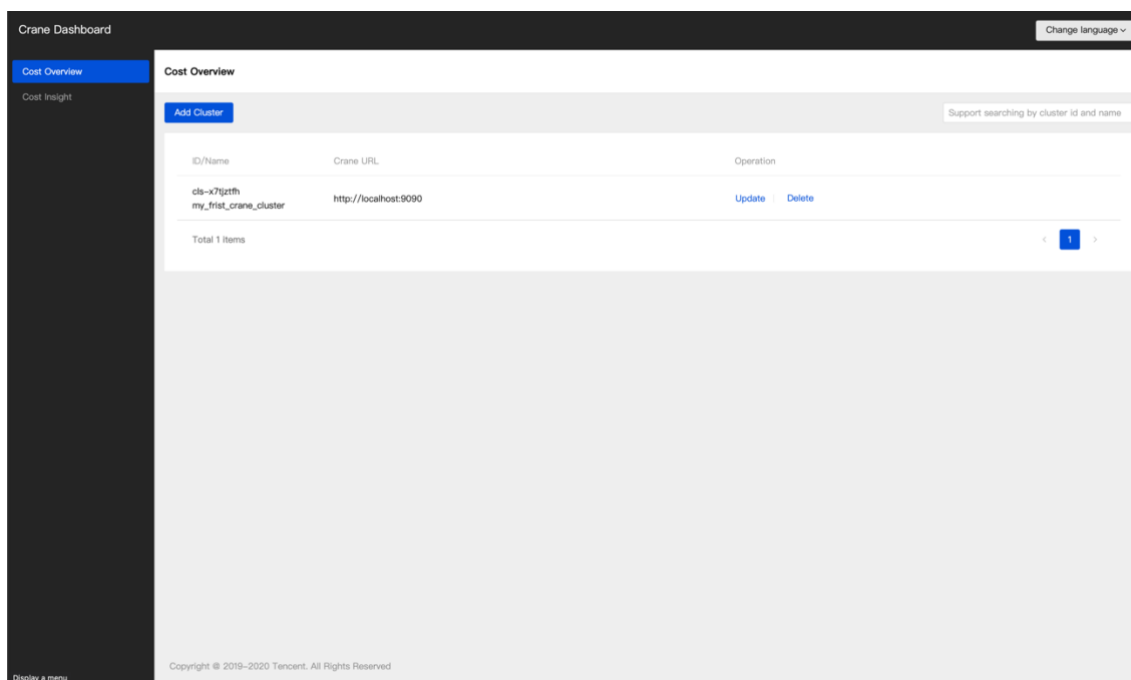
Käesoleva lõputöö nõuetes on välja toodud, et SaaS teenuseid ei ole juriidilistel asjaoludel võimalik kasutada, sest see muudaks antud töövahendi kasutuselevõtu keerulisemaks ja ajamahukamaks. Seetõttu tõttu ei oleks võimalik täita ettevõtte teise kvartali ega lõputöö eesmärke.

3.2.6 Crane + FAdvisor

Crane (Cloud Resource Analytics and Economics) on *cloud native* avatud lähtekoodiga projekt, mis haldab Kubernetese kulusid. See on inspireeritud FinOps'i kontseptsioonidest [49].

Crane-i eesmärk on pakkuda terviklikku projekti, mis aitab Kubernetese kasutajatel pilveressursside kasutamisel kokku hoida, pakkudes rikkalike funktsioonide komplekti, seal hulgas aegridade (*time series*) prognoosimine seireandmete põhjal, ressursikasutuse- ja kulude läbipaistvus, ressursikasutuse- ja kulude optimeerimine Kubernetese ressursside ümberpaigutuse (*reallocation*), kasutusnõuete ja kordistamise (*requests and replication*) abil ning tõhus autoskaleerimine (*effective autoscaling*). [50]

Joonisel 7 on näha Crane-i paigalduse käigus installeeritud veebirakendust, mis võimaldab sinna alla Kubernetese klastreid koondada ja nende kulusid raporteerida.



Joonis 7. Crane koondpaneel lisatud klastritest [90].

FAdvisor (FinOps Advisor) võimaldab hinnata ning välja tuua iga Kubernetese kauna, kontaineri või muu ressursi kulu kasutajate poolt defineeritud siltide siltide (*label*) abil. FAdvisori kulumudel on mõeldud kulude prognoosimiseks, mitte kuluarvelduse asendamiseks.

See on kogum erinevatest meetrika eksportijatest, kogujatest ja visualiseerijatest ning rakendusliidese serverist, mille installeerimise käigus paigaldatakse järgnevad komponendid:

- Eksportijad – koguvad FinOps-i poolt soovitud meetrikat kuluarvestuse jaoks.
 - Kube-State-Metrics – teenus, mis kuulab Kubernetes API serverit ning genereerib erinevat meetrikat Kubernetes ressurside oleku kohta [51].
 - Node-Exporter – Prometheusi eksportija riistvara ja operatsioonisüsteemi meetrika jaoks [52].
- Prometheus – avatud lähtekoodiga süsteemide jälgimise tööriistakomplekt.
- Grafana – avatud lähtekoodiga veebirakendus, mis võimaldab jälgida, analüüsida ning visualiseerida rakenduste meetrikat.
- Rakendusliidese server - koondab kogutud andmed ning arvutab pilvekasutuse kulud.

FAdvisorit on võimalik paigaldada ka ilma Crane-i installeerimata, mille käigus paigaldatakse vaid eksportijate kogum, millest saadud meetrika on võimalik Prometheusi abil kokku koguda ning Grafana kaudu visualiseerida.

Joonisel 8 on näha FAdvisor meeskonna poolt loodud Grafana koondpaneeli, mis annab ülevaate Kubernetesi klasteri ressursikuludest- ja kasutusest nimeruumi vaates [53].



Joonis 8. FAdvisor kulude koondpaneel nimeruumi järgi jeenides [53].

2022. aasta märtsi seisuga on erinevatest pilveplatvormidest toetatud vaid Tencent Cloud ning teiste pilveplatvormide toetamine 2022 esimese poolaasta tegevuskavadesse välja

toodud ei ole [54], mis tähendab, et AWSi kulusid ei ole võimalik läbi API päringute lugeda.

Crane koos FAdvisor paigaldusega võimaldab küll Kubernetese ressursside rakenduse ning meeskonnapõhist kuluarvestust varasemalt defineeritud märgiste strateegia abil, kuid ametlikus dokumentatsioonis ega näidispildidel ei ole informatsiooni salvestusseadmete, koormusjaoturite ning võrguliiklusega seotud kulude kohta. See tähendab, et tegelikkuses ei toimu kulujaotus nii detailselt nagu mõne konkureeriva töövahendi puhul.

Antud töövahend paigaldatakse oma taristule ning see on kättesaadav kõigile meeskondadele läbi Prometheusi või Crane veebileidese. Sellest tingituna vastab see kahele lõputöö nõudele.

3.2.7 Vantage

Vantage on 2020. aastal Ben Schaechteri ja Brooke McKimi poolt asutatud pilvekulude haldamise platvormi pakkuv ettevõtte. Vantage koondab pilvetaristu kulud kõigi pilveteenuste pakujate lõikes, et pakkuda tsentraliseeritud ülevaadet pilve kogukuludest. Tegemist on SaaS tüüpi lahendusega ning seetõttu on panustatud väga palju dokumentatsioonile, et selgitada, kuidas teatud funktsioonid ja süsteemid töötavad [55].

Vantage kasutab AWS-ga suhtlemiseks kontoülest (*cross-account*) IAM rolli, mille abil võimaldatakse pilveplatvormi kuluülevaadet, mis tähendab, et Vantage ei vaja AWSi APIga suhtlemiseks juurdepääsumandaate (*access credentials*) ega paroole. Tegemist on AWSi poolt ametlikult soovitatud ja eelistatud meetodiga kontodevaheliste tegevuste jaoks. Seda meetodit kasutavad ka AWSi sisemised teenused ja süsteemid [56].

Kuigi tegemist on kommertstarkvaraga, on üks neljast kategooriast siiski tasuta [57]:

- Starter – Ainus tasuta kategooria, erinevate pilveplatvormide tugi, kahe kuu pikkune andmesäilitus, kuluraportite, -prognooside ja -soovituste võimalus. Kuni 2500 dollari ulatuses jälgitavaid pilvekulutusi.

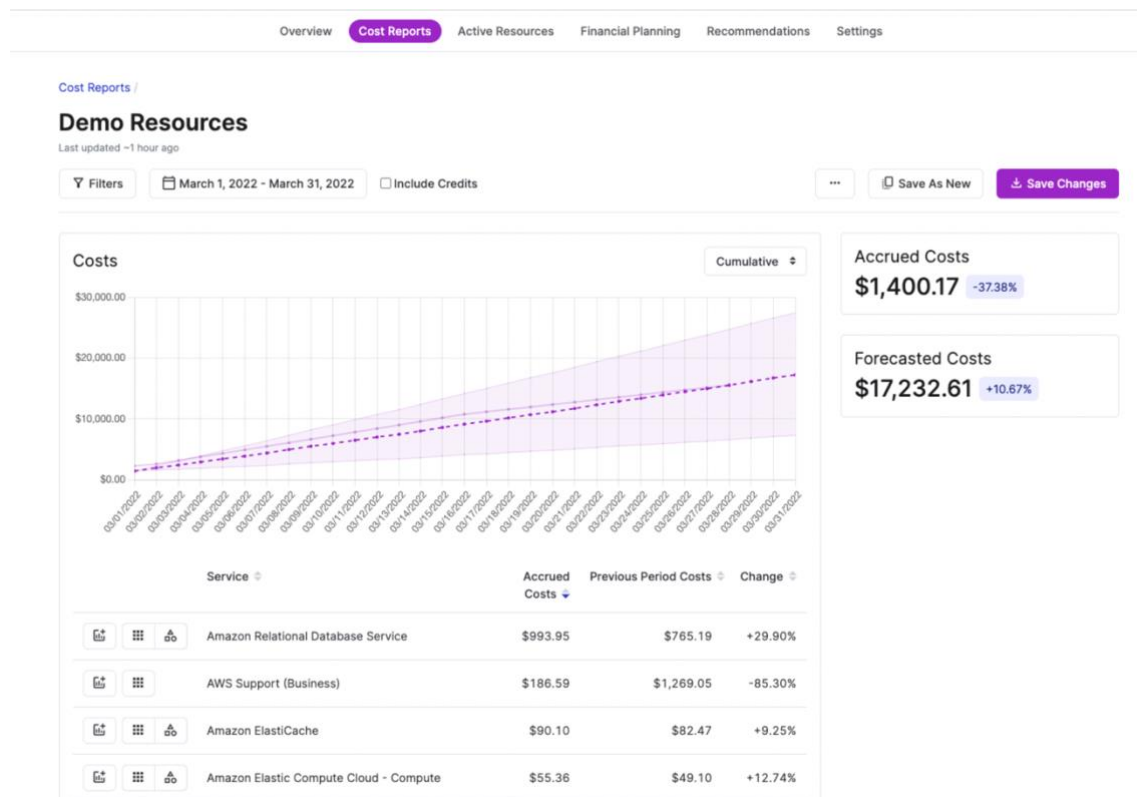
- Pro – Lisaks Starter kategooriale kuue kuu pikkune andmesäilitus, igakuine kulude kokkuvõte, garanteeritud teenustasemelepe ning kuni 7500 dollari ulatuses jälgitavaid pilvekulusid.
- Business – Lisaks Pro kategooriale aasta pikkune andmesäilitus, Kubernetese kulujaotuse võimekus, personaalne teenusehaldur ning kuni 20 000 dollari ulatuses jälgitavaid pilvekulusid.
- Enterprise – Lisaks Business kategooriale veel RBAC (Role-Based Access Control) ja SSO tugi ning rohkem kui 20 000 dollari ulatuses jälgitavaid pilvekulusid.

Kulude jaotamine toimub läbi automaatselt kogutud kasutusmeetrika, eeldusel, et kasutatakse AWSi poolt hallatud Kubernetese teenust EKS ning CloudWatch Container Insights teenust meetrika kogumiseks ja edastamiseks, mis omakorda vajab CloudWatch pistikprogrammi paigaldust Kubernetese klastrisse. Ametliku dokumentatsiooni kohaselt on hetkel toetatud vaid AWSi hallatud Kubernetese teenus EKS ning see välistab võimaluse kliendi enda poolt hallatud Kubernetese või Amazoni hallatud konteinerteenust (*Elastic Container Service*) teenuse kulude jaotuse [56].

Vantage-i tasuline *Business* kategooria pakub Kubernetese kulujaotust klastri, nimeruumi, rakenduse või märgiste alusel mis võimaldaks ka meeskonnapõhist kuluarvestust. Siiski eeldab iga märgis (*label*) Vantage-i meeskonna poolt käsitsi defineerimist [58], mis suurendab ajakulu ning töövahendi kasutamise keerukust olukordades, kus tekib vajadus uute märgiste lisamiseks või olemasolevate muutmiseks.

Lisaks on ettevõtte kontekstis SaaS teenuste kasutamine juriidilistel asjaoludel välistatud, kuna on ajamahukas ja võib osutada keeruliseks.

Joonisel 9 on näha näidisaruannet AWSi kulude kohta, tuues välja kõigi teenuste kogukulu vastavalt kuupäevale, kuluprognooši jooksva kuu lõikes ning kulude ülevaate teenuste lõikes [59].



Joonis 9. Vantage-i näidisaruanne AWS-i kulude kohta [59].

Iga kuluaruande loomisel koostatakse masinõppemudeli abil kulude prognoos järgneva kuu lõikes. Vantage pakub kuluraportite edastamist läbi e-maili teenuse ning Slacki ja Microsoft Teamsi suhtluskanalite.

Ettevõtte kontekstis, kus pilveplatvormi regioone on üheksa, ületab pilvekulude summa tugevalt 20 000 dollari piiri, millest tingituna sobiks töövahendi kategooriana ainult *Enterprise* lahendus.

3.2.8 Olemasolevate töövahendite võrdlus ning sobiva töövahendi valik

Analüüsi käigus selgus, et paljude lahenduste puhul täideti enamus nõuetest kas tasuta või tasuliste kategooriate abil ja kõiki töövahendeid on võimalik kasutada erinevates keskkondades. Siiski on analüüsitud töövahenditel ka mitmeid puudusi. Kuigi HCCM, Vantage ja CAST AI seadistamisel paigaldatakse Kubernetese klastrisse pistikprogrammid, on taustal tegemist siiski SaaS teenusega, mille kasutamine on antud

lõputöös viidatud probleemile välistatud. Lisaks võimaldab Vantage ainult osaliselt oma taristule paigaldamist ja kulude eristamist märgiste abil.

Mõlemad AWSi teenused ei vastanud nõuetele, sest kuigi võimaldatakse märgiste abil meeskonnapõhiste kulude eraldamist, kuvatakse Kubernetesi kulusid summeeritult ühe teenuse näol ning sellest tingituna ei ole võimalik ainult neid teenuseid kasutades lõputöö eesmärki täita.

Crane + FAdvisor võimaldas küll paigaldamist oma taristule, kuid lõputöö kirjutamise hetkel toetatud ainult üks pilveplatvorm – Tencent Cloud [60].

Tabelis 1 on välja toodud olemasolevate töövahendite võrdluse kokkuvõte ning nende vastavus lõputöös seatud nõuetele. Nõudeid hinnati väärtusega 0-2, mis liideti kokku ning mille abil arvutati punktisumma. Suurima punktisummaga töövahend osutus valituks. Nõuete väärtused jagunesid järgnevalt:

- 0 – ei vasta nõudele;
- 1 – vastab osaliselt nõudele;
- 2 – vastab täielikult nõudele.

Kahe lahenduse puhul on töövahendite võrdluses (Tabel 1) kasutatud lühendeid **CF** (Crane + FAdvisor) ning **CE** (AWS Cost Explorer). Tabeli päisesse on lisatud teenuse nimed, tabeli jalusesse nõuetele vastavuse punktisumma ning tabeli vasakpoolses veerus on kirjeldatud lõputöö nõuded.

Tabel 1. Olemasolevate töövahendite võrdlus lõputöö nõuetega.

	CE	CUR	CAST AI	Kubecost	HCCM	CF	Vantage
Kulude koondamine erinevatest klastritest ja regioonidest	2	2	2	2	2	2	2
Kulude eristamine märgiste alusel	2	2	0 [26]	2	0	2	1 [58]
Kombineeritud meeskonnapõhine kulujaotus	0	0	0 [29]	2	2	0	2
Kulude lugemine läbi API päringute	0	0	0	2	2	0 [54]	2
Paigaldus oma taristule	0	0	1 [30]	2	1 [45]	2	1 [58]
Kubernetese rakendus põhine kuluarvestus	0	0	2	2	2	2	2
Püsitalletuse tugi	2	2	2	2	2	0	2
Kättesaadav kõigile meeskondadele	2 [20]	2 [20]	1 [29]	2	2	2	1 [57]
Meeskondadel on võimalik koondpaneelide loomine	2	2 [23]	0	2	1 [44]	2	2 [57]
Kasutatav erinevates keskkondades	2	2	2	2	2	2	2
Punktisumma	12	12	10	20	16	14	17

Võttes aluseks olemasolevate töövahendite vastavuse lõputöös sätestatud nõuetele koos kokkuvõtliku töövahendite võrdluse tabeli punktisummaga (Tabel 1), jõuti järeldusele, et lõputöö eesmärgiks sobivaim töövahend on Kubecost. Kubecost vastas täielikult kõigile loodava lahenduse nõuetele ning on ka AWSi enda poolt soovitatud tööriist Kubernetese kulujaotuseks [61]

4 Pakutud lahendus

Käesolevas peatükis kirjeldatakse lühidalt paigaldatava Kubecost töövahendi keskkonna taustainfot, tehnoloogia valikut ning lahenduse arhitektuuri. See aitab lahti mõtestada, kuidas viidi läbi lahenduse planeerimine.

4.1 Paigalduskeskkonna taustainfo

Selleks, et paremini mõista väljapakutud lahendust, selle planeerimist ning realiseerimist, tehakse ülevaade keskkonnast, kuhu antud lahendust paigaldama hakatakse.

Kogu pilvetaristu ja Kubernetese mikroteenuste haldus on teostatud taristu koodina meetodil (*Infrastructure as Code*), kasutades selleks TerraForm ning YAML keeli. Mõlema keele puhul kirjeldatakse ressursse ning nende soovitud olekuid deklaratiivse koodina [62] [63] ning seda hoiustatakse versioonihaldussüsteemi Git repositooriumis (edaspidi Git).

Ressursside soovitud olekud realiseeritakse läbi pideva tarne (*CD*) tööriistade. Pilvetaristu puhul kasutatakse täielikult hallatud AWS CodePipeline teenust, mis võimaldab automatiseerida tarkvara või infrastruktuuri väljalaskmiseks vajalikke samme [64]. Selle abil käivitatakse TerraFormi kood ning saavutatakse pilveteenuste soovitud olekud.

Kubernetesi hallatud mikroteenuste puhul realiseeritakse ressursside soovitud olek Argo CD kaudu, kasutades selleks YAML keeles defineeritud manifeste. Argo CD on deklaratiivne GitOps'i pideva tarne tööriist Kubernetesele [65].

4.2 Lahenduse arhitektuur

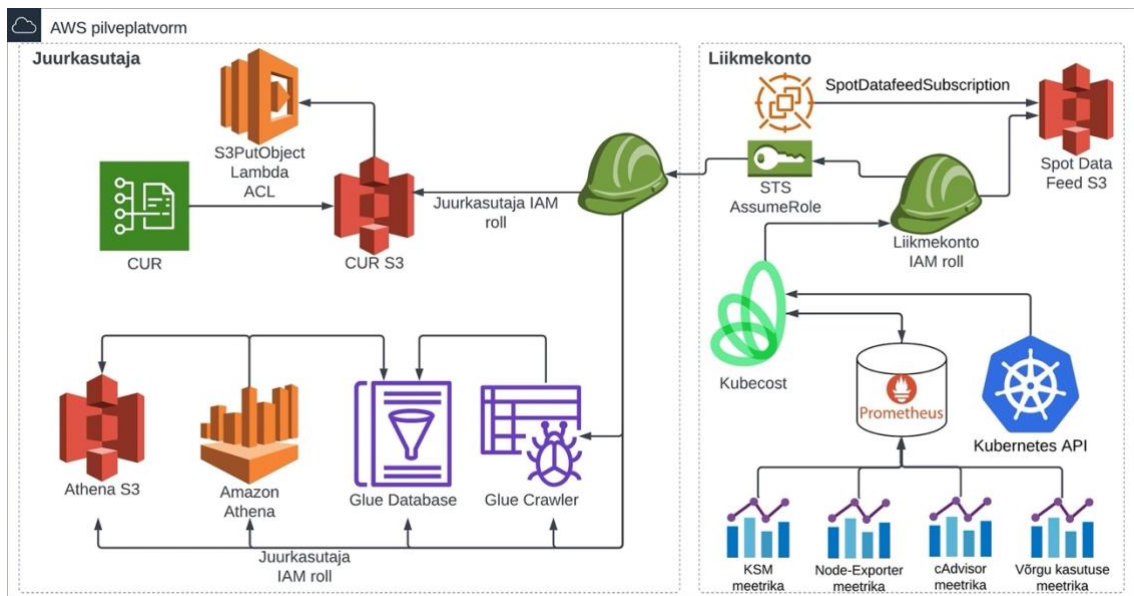
Võttes arvesse, et töövahend paigaldatakse Kubernetesesse ning sellega seotakse välised AWS-i teenuse kulud, koosneb arhitektuur järgnevatest komponentidest (Joonis 10):

- Kubecost
 - Cost-model – teostab kulujaotuse arvutusi ning loeb ja kirjutab andmeid Prometheusi;

- Frontend – käitab Nginx pöördproksit, haldab marsruutimist cost-modeli ja serveri suunas;
- Server – server mis vatab API päringutele.
- Kubernetes API – liides Kubernetese klastrite haldamiseks, loomiseks ja konfigureerimiseks.
- Prometheus – aegridade andmesalv monitooringu meetrika jaoks
 - Kube-State-Metrics – edastab Kubernetese API meetrikat;
 - Node-Exporter meetrika – edastab sõlmede kasutusmeetrikat;
 - cAdvisor meetrika – edastab konteinerite ressursikasutuse meetrikat;
 - Võrgu meetrika – edastab võrgukasutuse meetrikat detailsemaks kulujaotuseks.
- AWS integratsioon
 - Kulude jaotamise sildid – võimaldab kulude arvutamist siltide abil
 - CUR
 - S3 bucket CUR andmete jaoks – konteiner kuhu uuendatakse kord päevas CUR aruanded;
 - Glue Crawler – teenus mille abil edastatakse ajakohased andmed CUR S3 bucket-ist Glue andmebaasi;
 - Glue Database – andmebaas, kus hoiustatakse CUR vajalikke andmeid Amazon Athena päringute tegemiseks;
 - Amazon Athena – teenus, mis teostab andmebaasipäringuid AWS Glue andmebaasist;
 - S3 bucket Athena päringute vastuste jaoks – siin hoiustatakse Amazon Athena poolt teostatud päringute vastuseid, mida Kubecost kuluarvestuseks kasutab.
 - Spot Instance kasutuskulud
 - Spot Data S3 bucket – konteiner, kus hoiustatakse spot instance kulu- ja kasutusandmeid;
 - Spot Data Feed tellimus (subscription) – teenus, mis edastab spot instance serverite kulu- ja kasutusandmed S3 bucket-isse.
 - IAM
 - Kasutaja (testimise eesmärgil) – testimise eesmärgil loodud IAM kasutaja, millega oleks võimalik teenustele ligipääsu kontrollida;

- Rollid – IAM rollid, mida Kubecosti rakendus teenustele ligipääsemiseks kasutab;
- Poliitikad – IAM poliitikad, mis tagavad vajalikud õigused AWS teenuste kasutamiseks.

Joonisel 10 on kujutatud lõputöö lahenduse arhitektuur selliselt nagu see tehnilise teostuse peatükis realiseeritakse. Parema ülevaate andmiseks mõtestatakse lahti Kubecost töövahendi komponentide omavaheline suhe.



Joonis 10. Kubecost töövahendi komponentide diagramm.

Vastavalt Kubecosti dokumentatsioonile ning AWS-i ametlikule dokumentatsioonile, on AWS-i organisatsiooni juurkontos seadistatud CUR teenus koos sinna kuuluva S3 bucket-iga [39], [66]. Ametliku AWS CloudFormation malli abil on seadistatud AWS Lambda, Amazon S3, AWS Glue ning Amazon Athena teenused [67].

Lisaks on Kubecosti ning AWS-i ametlikku dokumentatsiooni järgides seadistatud IAM kasutajad, rollid ja poliitikad organisatsiooni juur- ja liikmekontos ning seejärel ühendatud IAM roll Kubecosti teenuskontoga [39], [68], [69].

Kuna ettevõtte kontekstis kasutatakse osades keskkondades ka *Spot Instance* tüüpi virtuaalservereid, on vastavalt Kubecosti ning AWS-i dokumentatsioonile seadistatud ka *S3 bucket*, kuhu kogutakse *Spot Data Feed* andmevoo abil nende virtuaalserverite kasutusstatistikat ja hindu [39], [70]. Sellega saavutatakse Kubecostis veelgi detailsem kulujaotus.

Joonisel 10 on näha, kuidas Kubecost pöördub eksportijate kogutud meetrika saamiseks Prometheusi teenuse poole. Saadud informatsiooni alusel teostatakse Kubernetese klasteri kulujaotuse arvestus.

AWSi teenuste kulude arvestuseks pööratakse Kubecosti IAM rolli kaudu *CUR S3 bucket*-i poole. Kuna Kubecosti töövahend ning *CUR* andmed asuvad erinevates AWSi kontodes, kasutatakse kontoüleseks ligipääsuks Kubecosti rolli kaudu *AssumeRole* API päringuid. Nii tagastatakse läbi STSi (*Secure Token Service*) ajutine pääsumandaat juurkasutaja teenustele [71].

Sellega võimaldatakse Kubecosti töövahendile *CUR* andmetele ligipääs, kust AWS Glue Crawleri abil ajakohased andmed AWS Glue andmebaasi sisestatakse. Seejärel teostatakse Amazon Athena kaudu päringud AWS Glue andmebaasi. Päringu tulemused kirjutatakse Amazon Athena tulemuste *S3 bucket*-isse, kust saadud andmed Kubecosti IAM rolli abil analüüsimiseks edastatakse.

Spot Instance tüüpi serverite hinnakirja saamiseks pööratakse Kubecosti IAM rolli kaudu AWSi organisatsiooni liikmekontos (*member account*) asuva *Spot Data Feed S3 bucket*-i poole, kust informatsioon Kubecosti analüüsimiseks edastatakse. Kuna antud *S3 bucket* asub Kubernetese klasteriga samas kontos, ei ole STSi kaudu *AssumeRole* API päringud enam vajalikud.

4.3 Tehnoloogia valik

Tehnoloogia valikul on oluline et järgitaks parimaid praktikaid ning võimaluse korral jätkataks ettevõttes varasemalt juurutatud tehnoloogia kasutamist. See kiirendab ja lihtsustab tehnilise teostuse protsessi. Samuti hoitakse tehnoloogia pinu (*stack*) ühtlasena, mis on teada ka teistele ettevõtte DevOps meeskonnaliikmetele.

4.3.1 Tehnoloogia taristu haldamisel

Kuna lõputöö eesmärgi realiseerimiseks oli vajalik uute AWS teenuste ettevalmistus (*provisioning*) ning nende haldamine, omas suurt tähtsust ka tehnoloogia valik taristu haldamiseks.

Lõputöö autor soovis jätkata juba varasemalt ettevõttes kasutatud tööriistadega TerraForm ja AWS CloudFormation, kuna tegemist on laialt levinud ning kõrgelt hinnatud töövahenditega [72]mis võimaldavad hoida ka tehnoloogia pinu ühtlasena.

4.3.2 Tehnoloogia töövahendi haldamisel

Töövahendi paigaldamine eeldas Kubernetese klastrisse installeerimist ning seetõttu oli oluline tähelepanu pöörata ka seal kasutatud tehnoloogiale.

Sarnaselt taristu haldamise tehnoloogiale, soovis lõputöö autor jätkata varasemalt kasutatud meetodite ja Argo CD tööriistaga, et tehnoloogia pinu ühtlasena hoida. Lisaks on Argo CD puhul tegemist ühe hinnatuma GitOps tööriistaga Kubernetese jaoks [73].

5 Tehniline teostus

Käesolevas peatükis kirjeldatakse, kuidas realiseeriti töövahendi esialgne paigaldus, töövahendi seadistamine ning sellele eelnenud tegevused AWSi teenuste kulujaotuse saamiseks ning kulujaotuse siltide aktiveerimine. Lisaks tuuakse välja tehnilise teostuse käigus tekkinud probleemid ning kirjeldatakse kuidas tekkinud probleemid lahendati.

5.1 Kubecosti paigaldamine

Töövahendi paigaldusel lähtuti ametlikust Kubecosti dokumentatsioonist [74], mis võimaldas rakenduse esialgse probleemivaba paigalduse. Kuna tegemist on peamiselt Kubernetese kulujaotuseks mõeldud töövahendiga, paigaldati rakendus Kubernetese klastrisse [74].

5.1.1 Mallimine ning manifestide kirjeldamine

Kuigi Kubecosti dokumentatsioonis on töövahendi soovitatud paigaldusviisiks Helmi mallide kasutamine [74] soovis lõputöö autor hoida ühtlasena kasutuses olevat tehnoloogiapinu ning seetõttu valiti manifestide kirjeldamiseks *helm template* käsk [75]. Selle abil genereeriti Kubecosti väärtuste (*values*) ja konfiguratsiooni failidest Kubernetese YAML manifestid.

Arvestades töövahendi olemust ja võrreldes seda varasemalt paigaldatud rakendustega ning ettevõtte Kubernetese klatri iseärasustega, leidis lõputöö autor, et antud lahendus tuleks paigaldada Kubernetese nimeruumi *monitoring*. Joonisel 11 on välja toodud käskude jada, mille abil genereeriti esialgsed Kubernetese manifestid ning lisati need seejärel rakenduste versioonihalduse repositooriumisse .

```
helm repo add kubecost https://kubecost.github.io/cost-analyzer/  
helm -n monitoring template cost-analyzer kubecost/cost-analyzer -f  
/kubecost/custom-values.yaml --output-dir /helm/kubecost/manifests  
git add .  
git commit -m“kubecost initial deployment files“  
git push
```

Joonis 11. Töövahendi vaikeväärtuste mallimine ning failide lisamine repositooriumisse.

Mallimise käigus defineeriti töövahend kasutama olemasolevat Prometheusi monitooringutarkvara, võimaldati püsitalletus ning defineeriti Kubernetese *Ingress* tüüpi ressurss, et võimaldada rakendusele klastriväline ligipääs.

5.1.2 Töövahendi paigaldamine

Tingituna kokkuleppest ettevõtte ja lõputöö autori vahel, paigaldati Kubecost esialgu ühte keskkonda ning selle abil realiseeriti taristu arenduseks mõeldud Kubernetese klastris valmislahendus, mida oleks võimalik tulevikus kasutada ka teistes keskkondades ning ettevõtteüleselt ka teiste klientide jaoks.

Selleks defineeriti kõigepealt *Application* tüüpi Argo CD ressurss, milles kirjeldati rakenduse iseärasused nagu failide asukoht repositooriumis, nimeruum ning automaatne sünkroniseerimispoliitika. Argo CD abil paigaldati ressursid Kubernetese klastrisse ning hoiti nende defineeritud olekuid [76]. Sünkroniseerimispoliitika puhul oli määratud, et ressursside kustutamisel Git repositooriumist, eemaldatakse ressursid ka Argo CD abil Kubernetese klastrist [77].

Joonisel 12 on näha töövahendi püsitalletusseadmele lisatud annotatsioon, millega defineeriti, et Argo CD rakenduse kustutamisel jäetakse see ressurss ikkagi alles. See garanteerib, et kogemata kustutatud Kubecosti rakenduse puhul ei tekiks andmekadu [78].

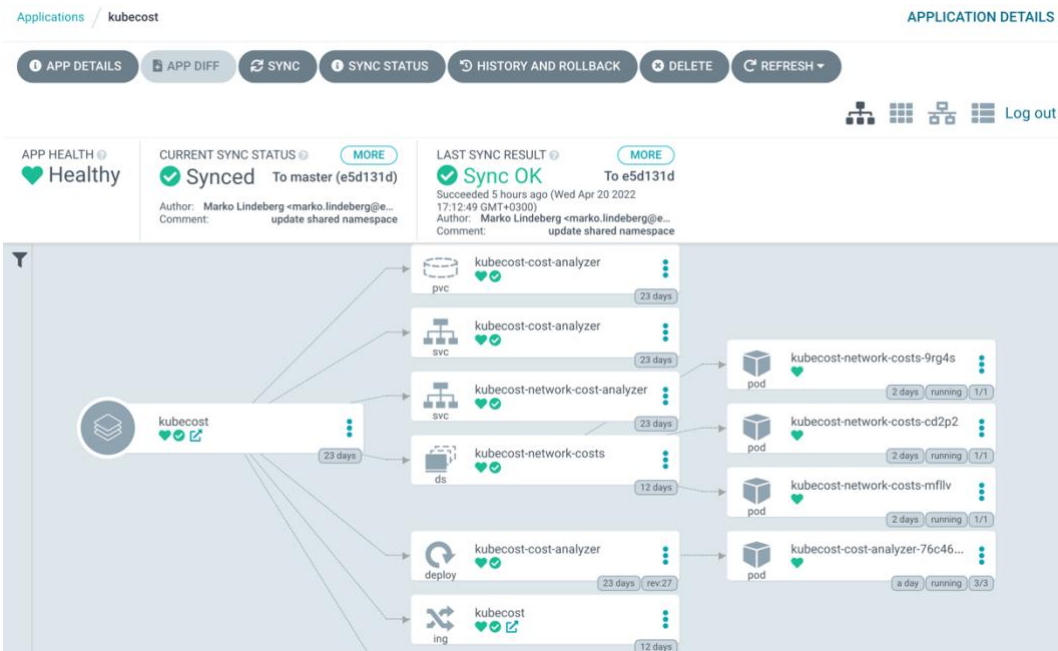
```
# Sünkroniseerimisel ei kustutata ressurssi #  
argocd.argoproj.io/sync-options: Prune=false
```

Joonis 12. Annotatsioon Kubernetese ressursi kustutamise vastu [78].

5.1.3 Töövahendi esmane testimine

Selleks, et antud töövahendi edukas paigalduses ning seadistuses kindel olla, kontrolliti lõputöö autori poolt Argo CD veebiliidese kaudu töövahendiga seotud komponentide olekut.

Joonisel 13 on näha filtreeritud kujul Kubecosti ressursside sünkroniseeritud olekut Kubernetese klastris.



Joonis 13. Kubecost rakenduse ülevaade Argo CD veebiliideses.

Pärast rakenduste paigaldust Argo CD kaudu, testiti rakenduse veebiliidese avamist läbi värskelt loodud Kubernetese *Ingress* objekti, mis kinnitas lõputöö autorile, et töövahendi esimane paigaldus õnnestus.

5.2 AWS kulude sidumine Kubecostiga

Lõputöö autor lähtus AWSi kulude ja Kubecosti sidumisel ametlikest dokumentatsioonidest [39], [41] mille abil lahenduse eeltingimusteks vajalikud ressursid loodi ning seadistati.

AWS kulude sidumine Kubecostiga koosnes järgmistest etappidest:

- CUR loomine ning selle integreerimine Amazon Athena teenusega – CUR kuluraportite loomine.
- Spot Data S3 ning Spot Data Feed andmevoo seadistamine – spot instance tüüpi virtuaalserverite kulu- ja kasutusandmete kogumine.
- Kubernetese ja AWS teenuste märgiste lisamine – kulujaotuseks vajalike märgiste lisamine.
- Kulujaotuse märgiste aktiveerimine – aktiveeriti märgised, mida CUR raportites esile tuuakse.

- IAM poliitikate, kasutajate ja rollide loomine – loodi vajalikud õigused teenuste kasutamiseks.
- IAM rolli sidumine Kubecostiga – seadistati töövahend IAM rolli kasutama.
- Töövahendi lõplik seadistus – teostati viimased töövahendi seadistused.

5.2.1 CUR loomine ning selle integreerimine Amazon Athena teenusega

Tuginedes Kubecosti dokumentatsioonile, oli AWSi kulujaotuseks vajalik liidestus CURi ning Amazon Athena teenustega, mis eeldas nende teenuste olemasolu organisatsiooni juurkontos [39].

Kuna CUR oli varasemalt juurkontost puudu, tuli lõputöö autoril see kõigepealt Kubecosti ning AWSi dokumentatsiooni järgi selliselt seadistada, et rakenduste integratsioonitüübiks oleks valitud Amazon Athena [41]. CUR esmase seadistuse käigus loodi AWSi dokumentatsiooni kohaselt ka S3 *bucket*, kuhu AWSi kuluraportid igapäevaselt kirjutatakse [79].

Pärast 24 tunni möödumist, laekusid CURi S3 *bucket*-isse esimesed andmed AWSi teenuste kulude kohta koos Amazoni poolt eelseadistatud AWS CloudFormation malliga, mille abil võimaldati Amazon Athena integratsioon CUR raporteerimisega, kasutades selleks Amazon Lambda funktsioone ning AWS Glue teenuseid [67] ning nende omavaheliseks sidumiseks vajalike IAM rolle.

Samuti lähtuti FinOps'i esimese etapi eesmärgist võimaldada meeskondadele ajakohane ülevaade nende pilvekuludest [14] ning seetõttu valiti raporti loomisel võimalus kulude agregeerimiseks just tundide, mitte päevade või kuude järgi [66].

Joonisel 14 on näha AWS CloudFormation malli abil loodud 11 ressursi ülevaadet, millega võimaldati Amazon Athena ja CUR raportite integratsioon.

Logical ID	Physical ID	Type	Status	Status reason	Module
AWSCURCrawler	AWSCURCrawler-kubecost	AWS::Glue::Crawler	CREATE_COMPLETE	-	-
AWSCURCrawlerComponentFunction	kubecost-AWSCURCrawlerComponentFunction-14M2KLRFCPBQ8	AWS::IAM::Role	CREATE_COMPLETE	-	-
AWSCURCrawlerLambdaExecutor	kubecost-AWSCURCrawlerLambdaExecutor-QEDD7NVBV52V	AWS::IAM::Role	CREATE_COMPLETE	-	-
AWSCURDatabase	athenacurfn_kubecost	AWS::Glue::Database	CREATE_COMPLETE	-	-
AWSCURInitializer	kubecost-AWSCURInitializer-dvsAw1TDIbv4	AWS::Lambda::Function	CREATE_COMPLETE	-	-
AWSCURReportStatusTable	cost_and_usage_data_status	AWS::Glue::Table	CREATE_COMPLETE	-	-

Joonis 14. CUR ja Athena integreerimiseks kasutatud CloudFormationi malli loodud ressursid.

AWS CloudFormation malli abil loodud ressursid võimaldasid CURi ja Amazon Athena vahelise automaatse integratsiooni, mis oli vajalik AWSi teenuste kulujaotuse näitamiseks läbi Kubecosti.

5.2.2 Spot Data S3 ning Spot Data Feed andmevoo seadistamine

Kubecosti ja AWSi integratsioon võimaldab *spot instance* tüüpi virtuaalserverite kuluarvestust. Kuna tegemist on olemuselt justkui börsilt ostetud arvutusjõudlusega, mida saab tunduvalt odavamalt võrreldes tavalist tüüpi EC2 virtuaalserveritega, tuleb arvestada asjaoluga, et nõudluse kasvades võivad virtuaalserverid igal ajal kaduda ning uutega asenduda [80].

Sellest tingituna muutuvad pidevalt ka virtuaalserverite hinnad ning korrektseks kulujaotuseks on vajalik ajakohase informatsiooni kättesaadavus. AWSil on selle jaoks eraldi andmevoog *Spot Data Feed*, mis edastab *spot instance* tüüpi virtuaalserverite hinna- ja kasutusstatistika varasemalt seadistatud S3 *bucket*-isse [70]. Kubecost toetab sama andmevoo alusel *spot instance* tüüpi serverite kuluarvestust [39].

Selle funktsiooni lubamiseks lõi lõputöö autor eraldi S3 *bucket*-i ning seadistas automaatse andmevoo, mis sinna *spot instance* virtuaalserverite kulu- ning

kasutusandmeid edastas. Lähtudes kasutuses olevast tehnoloogiast ning soovist hoida tehnoloogiapinu ühtlasena, kasutati *S3 bucket*-i ning selle ligipääsuõiguste loomiseks TerraFormi ressursse.

Joonisel 15 on näha käsud andmevoo loomiseks ning olemasolevate andmevoogude kuvamiseks koos JSON-vormingus redigeeritud väljundiga, kasutades selleks AWSi käsurealiidese kāske.

```
aws ec2 create-spot-datafeed-subscription \  
    --bucket spot-data-feed-dev \  
    --prefix spot  
  
aws ec2 describe-spot-datafeed-subscription  
  
{  
  "SpotDatafeedSubscription": {  
    "Bucket": "spot-data-feed-dev",  
    "OwnerId": "[REDACTED]",  
    "Prefix": "spot",  
    "State": "Active"  
  }  
}
```

Joonis 15. Spot Data Feed andmevoo loomine ning selle kirjeldamine JSON-vormingus AWS käsureautiliidi abil [70].

Antud JSON-vormingus väljundit redigeeriti, et sealt eemaldada AWSi konto identifitseerimisnumber. Andmevoo loomisel kasutatud parameeter *prefix* määrab eesliite, kuhu objektid edastatakse.

5.2.3 Kubernetese ja AWS teenuste märgiste lisamine

Üks FinOps'i kontseptsioonidest on pilveteenuste kulujaotus märgiste (*label*) abil. Selleks on soovitatud paika panna märgiste loomise strateegia [81]. Kubecost kasutab samasugust lähenemist ja korrektseks kulujaotuseks eeldatakse põhjalikku märgiste strateegiat.

Sellest tingituna töötas lõputöö autor välja ühtse Kubernetese ja AWSi ressursside märgiste kontseptsiooni, mis võimaldaks korrektset kulujaotust. Kubernetese ressursside puhul lisati märgised kõikidele koormusjaoturitele, püsitalletusseadmetele ning juurutusobjektidele ja nende mallidele. Ülejäänud AWSi teenuste puhul kõigile kasutuses olnud ressurssidele, kuhu märgiste lisamine võimalik oli.

Kubernetese märgiste lisamisel lähtus lõputöö autor Kubecosti ja Kubernetese dokumentatsiooni soovitudest [82]. Märgised defineeriti võti-väärtus paaridena järgnevalt:

- app: rakenduse nimi;
- team: meeskond, kelle poolt rakendust arendatakse;
- product: täpsustus, millise tootega on tegemist (toode võib koosneda erinevatest komponentidest);
- env: keskkond, kuhu rakendus on paigaldatud (näiteks, arenduskeskkond);
- component: defineerib, millise komponendiga on tegu (näiteks, andmebaas).

Kubernetese juurutusressursside märgiste näide, mille abil defineeriti rakenduse nimi, selle tootenimetus, meeskonna nimi, juurutuskeskkond ning missuguse komponendiga on tegemist, on välja toodud lõputöö lisades (Lisa 2). Ülejäänud AWSi teenuste märgiste lisamiseks defineeriti kõigepealt Kubecosti rakenduses väliste teenuste märgiste võtmed [39], mille abil Kubernetese ja AWSi kulud omavahel seoti.

Joonisel 16 on välja toodud Kubecosti väliste märgiste definitsioonid, millega võimaldatakse Kubernetese ja ülejäänud AWSi teenuste summeeritud kulujaotus.

```
namespace_external_label: "kubernetes_namespace"
cluster_external_label: "kubernetes_cluster" #Kubernetese klaster
controller_external_label: "kubernetes_controller"
product_external_label: "kubernetes_label_app"
service_external_label: "kubernetes_service"
deployment_external_label: "kubernetes_deployment"
team_external_label: "kubernetes_label_team" #meeskonna nimi
environment_external_label: "kubernetes_label_env" #keskkonna nimi
department_external_label: "kubernetes_label_department"
statefulset_external_label: "kubernetes_statefulset"
daemonset_external_label: "kubernetes_daemonset"
pod_external_label: "kubernetes_pod"
owner_external_label: "kubernetes_label_owner"
```

Joonis 16. Kubecosti konfiguratsioon väliste teenuste märgiste võtmete defineerimiseks.

Pärast väliste märgiste võtmete defineerimist lisati ülejäänud AWSi teenustele märgised. Kuna lõputöö eesmärgiks oli meeskonnapõhine kuluarvestus ning vaikimisi võimaldas Kubecost ka teenusepõhist kulujaotust, defineeriti vaid järgnevad võti-väärtus paarid:

- `kubernetes_cluster`: Kubernetese klasteri nimi (võimaldab tulevikus klasteriülest kulujaotust);
- `kubernetes_label_env`: keskkond, kus antud teenust kasutatakse;
- `kubernetes_label_team`: meeskond, kes seda teenust kasutab;
- `kubernetes_label_product`: täpsustus, missuguse tootega on tegemist.

Märgiste lisamisel muudeti juba varasemalt defineeritud TerraFormi ressursse, lisati muudatused Git repositooriumisse, kust need TerraForm tööriista abil realiseeriti. Kuna osad teenused olid ajalooliselt seadistatud käsitsi läbi AWSi kasutajaliidese, lisati nende teenuste märgised samuti AWSi kasutajaliidese kaudu.

Ettevõtte poolt kiirkorras läbi viidud taristukulude analüüsi käigus selgus, et üheks märkimisväärseks kuluallikaks oli AWS ECR (*Elastic Container Registry*) teenus, mis oma olemuselt eeldas märgiseid repositooriumi kaupa. Antud keskkonnas oli ECR teenuse puhul kasutuses 527 repositooriumit, mis tähendas, et märgiste lisamine läbi AWSi kasutajaliidese oleks olnud rohkelt aegavõttev ning suure eksimisvõimalusega. Sellest tingituna mõtles lõputöö autor välja alternatiivi, kasutades AWSi käsurealiidest ning selle dokumentatsiooni [83].

Joonisel 17 on näha käskude jada, mille abil lisati kolm võti-väärtus paariga märgist igale ECRi repositooriumile, seejärel kontrolliti ühele repositooriumile lisatud märgiseid koos selle väljundiga JSON-vormingus.

```

# Kõikidele repositooriumitele lisatakse märgised

aws ecr describe-repositories --region us-east-1 | jq
'.repositories[].repositoryArn' | xargs -I {} aws ecr tag-resource --
region us-east-1 --resource-arn {} \
    --tags Key=kubernetes_cluster,Value=dev
Key=kubernetes_label_team,Value=Infrastructure-DevOps
Key=kubernetes_label_env,Value=dev

# Kontrollitakse ühe repositooriumi märgiseid

aws ecr list-tags-for-resource --region us-east-1 \
    --resource-arn arn:aws:ecr:us-east-
1:[REDACTED]:repository/[REDACTED]/grafana

# Saadud tulemus JSON vormingus
{
  "tags": [
    {
      "Key": "kubernetes_cluster",
      "Value": "dev"
    },
    {
      "Key": "kubernetes_label_env",
      "Value": "dev"
    },
    {
      "Key": "kubernetes_label_team",
      "Value": "Infrastructure-DevOps"
    }
  ]
}

```

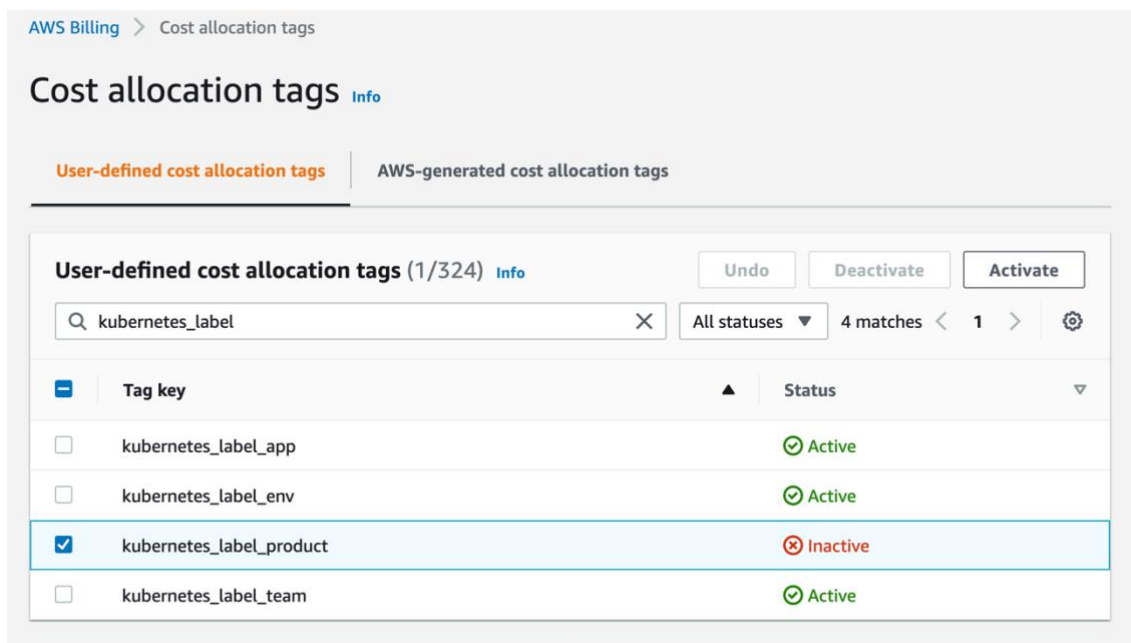
Joonis 17. Käskude jada ECR repositooriumite märgiste lisamiseks ning ühe repositooriumi märgiste kontroll koos selle väljundiga JSON-vormingus.

5.2.4 Kulujaotuse märgiste aktiveerimine

Kõik AWSi ressurssidele lisatud märgised ei ole alati lisatud kulujaotuse eesmärgil vaid pigem informatsiooni jagamiseks või mõnel muul filtreerimise eesmärgil. Selleks, et lisatud märgiseid oleks võimalik kuluarvestuses kasutada, tuli need kõigepealt läbi AWSi juurkonto aktiveerida.

Liikmekonto teenustele lisatud märgised tekkisid umbes 24 tundi pärast AWSi juurkonto Billing and Cost Management kasutajaliidese alla [84], kust need seejärel käsitsi

aktiveerida tuli. Joonisel 18 on näha AWSi juurkonto kulujaotuse märgiseid ning seda, kuidas neid aktiveeritakse.



Joonis 18. Kulujaotuse märgiste aktiveerimine kasutajaliidese kaudu.

Pärast kulujaotuse märgiste aktiveerimist tuli oodata umbes 24 tundi, et need kajastuksid CUR aruandes [85].

5.2.5 IAM poliitikate, kasutajate ja rollide loomine

Selles peatükis kirjeldatakse, kuidas loodi vajalikud IAM poliitikad, rollid ning kasutaja Kubecosti kuluarvestuseks. Töö käigus lähtuti Kubecosti dokumentatsiooni „Step 3: Setting up IAM permissions“ peatükist [41].

Kuna Kubecosti töövahend paigaldati arenduseks mõeldud Kubernetese klastrisse, mis ei ole AWSi juurkontos, oli vajalik IAMi poliitikate, rollide ning kasutaja seadistamine mõlemas kontos.

Ajaloolistel põhjustel ja AWSi soovitustest lähtudes, hallatakse juurkontos ainult hädavajalikke teenuseid ning seetõttu käib kogu seadistus käsitsi läbi AWSi kasutajaliidese. Sellest tingituna peeti mõistlikuks kasutada Kubecosti poolt toetatud AWS CloudFormation malli [86], mis võimaldas uue pinu loomisel lihtsalt ja probleemivabalt vajalikud ressursid luua, defineerides parameetritena liikmekonto ID ning Amazon Athena päringute tulemusteks loodud S3 *bucket*-i.

Malli abil loodi juurkontosse IAM roll Kubecosti töövahendi jaoks ning IAM poliitika peatükis 5.2.1 loodud Amazon Athena, Amazon S3 ja AWS Glue teenuste ligipääsuõigusteks. IAM poliitika seoti IAM rolliga, mis tagas poliitika abil defineeritud õigused.

Liikmekontos, kuhu Kubecost paigaldati, kasutati taristu halduseks TerraFormi skripte. Kuna CUR ning sellega seotud ressursid asusid AWSi juurkontos, defineeriti liikmekonto IAM poliitika kirjeldamisel STSi kaudu *AssumeRole* pääsumandaadi ligipääs juurkasutaja IAM rollile. Lisaks defineeriti ligipääsuõigused *Spot Data Feed* andmevoole.

Kahe olulisema IAM poliitika loomiseks vajalikud TerraFormi skriptid on välja toodud lõputöö lisades (Lisa 3).

5.2.6 IAM rolli sidumine Kubecostiga

Üks viimastest töövahendi paigaldamise sammudest eeldas peatükis 5.2.5 loodud IAM rolli sidumist Kubecosti teenuskontoga. Teenuskonto sidumisel IAM rolliga lähtuti AWSi dokumentatsioonist, mille abil lisati teenuskontole annotatsioon, mis võimaldab seda IAM rolli kasutada [68]. Teenuskontole annotatsiooni lisamine ning selle kontrollimise käsud koos väljundiga on esitatud lõputöö lisades (Lisa 4).

5.2.7 Töövahendi lõplik seadistus

Töövahendi seadistamise viimases peatükis kirjeldati Kubecosti manifestides keskkonnamuutujate abil CUR raportiga seotud väärtused. Defineeriti järgnevad väärtused [41]:

- `spotLabel` – spot instance tüüpi virtuaalserveri märgise võti;
- `spotLabelValue` – spot instance tüüpi virtuaalserveri märgise väärtus;
- `spotDataRegion` – AWS regioon, kus Spot Data Feed andmevoog seadistati;
- `spotDataBucket` – S3 bucket-i nimi;
- `projectID` – AWS konto ID, kus spot instance tüüpi servereid kasutatakse;
- `athenaProjectID` – AWS konto ID, kus on seadistatud CUR (juurkasutaja);
- `athenaBucketName` – Amazon Athena päringutulemuste S3 bucket;
- `athenaRegion` – AWS konto regioon, kus on seadistatud Amazon Athena teenus;

- athenaDatabase – AWS Glue andmebaas, mida kasutatakse Athena päringute tegemisel;
- athenaTable – tabeli nimi AWS Glue andmebaasis;
- masterPayerARN – AWSi juurkasutaja Kubecost IAM rolli ARN.

Pärast väärtuste defineerimist ning nende lisamist Giti repositooriumisse, taaskäivitati uue konfiguratsiooni sisse lugemiseks Kubecosti rakendus. Seejärel avati Kubecosti veebiliides ning veenduti, et AWSi teenuste integratsioon õnnestus. Joonisel 19 on näha Kubecosti õnnestunud integratsioon klastriväliste AWSi teenustega.

App Info

Kubecost v1.92.0

Node Usage over last 7 days

Max: 6 Avg: 6

Pricing Sources ¹

⚠ Savings Plan, Reserved Instance, and Out-Of-Cluster ^

- ✓ Config athenaBucketName is set
- ✓ Config athenaDatabase is set
- ✓ Config athenaProjectID is set
- ✓ Config athenaRegion is set
- ❗ no reservation data available in Athena

✓ Spot Data Feed ^

- ✓ Config awsSpotDataBucket is set
- ✓ Config awsSpotDataRegion is set

Joonis 19. Kubecosti diagnostika AWSi teenustega integreerimisel.

Üks kontrollidest on kujutatud punase hüüumärgiga, sest *reserved instance* tüüpi virtuaalsereid ei kasutata ning seetõttu puudub ka sellekohane informatsioon Amazon Athenas päringutes. Lõputöö autor suhtles selle kinnitamiseks ka Kubecosti kasutajatoega.

5.3 Tekkinud probleemid ning nende lahendused

Käesolevas peatükis antakse ülevaade tehnilise teostuse ajal tekkinud probleemidest ning nende lahendamisest.

Töövahendi paigaldamisel tekkisid probleemid AWSi teenuste kulujaotuse seadistusel. Kuigi lähtuti olemasolevast dokumentatsioonist, ei õnnestunud esimese iteratsiooni käigus Kubecosti ja CUR vaheline liidestus. Tõrkeotsingu käigus tegeleti töövahendi logide analüüsimisega, mille abil selgus, et üks Amazon Athena päring ebaõnnestus keelatud ligipääsuõiguste tõttu.

Tõrkeotsingu eesmärgil loodi juurde üks IAM kasutaja, millele kinnitati peatükis 5.2.5 loodud IAM poliitika. Loodud IAM kasutaja pääsumandaadi abil testiti AWS käsurautiliiti kasutades Amazon Athenas ebaõnnestunud päringu käitamist, mis õnnestus.

Joonisel 20 on näha AWSi käsurautiliidi kaudu tehtud päringud, millega kinnitati, et IAM õigused antud operatsiooni jaoks olid piisavad.

```
# AWS Athena päring

aws athena start-query-execution \
  --query-string "SELECT
line_item_usage_start_date,savings_plan_savings_plan_a_r_n,line_item_r
esource_id,savings_plan_savings_plan_rate FROM kubecost as cost_data
WHERE line_item_usage_start_date BETWEEN date '2022-04-08' AND date
'2022-04-09' AND line_item_line_item_type = 'SavingsPlanCoveredUsage'
ORDER BY line_item_usage_start_date DESC" \
  --work-group "primary" --query-execution-context
Database=athenacurcfn_kubecost,Catalog=AwsDataCatalog \
  --result-configuration OutputLocation=s3://aws-athena-query-
results-[REDACTED]

# Päringu tegemisest saabunud vastus koos selle ID-ga
{
  "QueryExecutionId": "05752b86-8547-47ee-9f34-cbc922b2914a"
}

# Päringu kirjeldamine, et näha selle staatust
aws athena get-query-execution \
  --query-execution-id 05752b86-8547-47ee-9f34-cbc922b2914a | jq
'.QueryExecution.Status'

# Päringu staatus
{
  "State": "SUCCEEDED",
  "SubmissionDateTime": "2022-04-09T01:32:57.154000+03:00",
  "CompletionDateTime": "2022-04-09T01:32:59.278000+03:00"
}
```

Joonis 20. AWS Athena päring koos selle kirjelduse ning staatusega.

Tänu õnnestunud Amazon Athena päringule välistati probleem IAMi poliitikaga. Järgmiseks testiti, kas Kubernetese teenuskonto, mis seoti eelmises peatükis loodud Kubecosti IAM rolliga, töötab korrektselt või mitte.

Selle testimiseks paigaldati Kubernetesesse üks AWSi käsureautiliiti sisaldav konteiner, mis määrati kasutama Kubecosti teenuskontot. Seejärel kontrolliti konteineri seest AWS STSi abil rolli identiteeti. Joonisel 21 on näha STSi abil päritud identiteeti ning sellelt saadud vastust.

```
# käsk STSi kaudu identifitseerimiseks

aws sts get-caller-identity

# STS-i poolt saadetud vastus
{
  "UserId": "AROARVWFQSE54BZIONKEH:botocore-session-[REDACTED]",
  "Account": "[REDACTED]",
  "Arn": "arn:aws:sts::[REDACTED]:assumed-role/KubecostRole-[REDACTED]/botocore-session-[REDACTED]"
}
```

Joonis 21. STSi kaudu identifitseerimine.

Saadud tulemus kinnitas, et STSi kaudu saadi korrektne pääsumandaat. Seejärel käivitati samast konteinerist Amazon Athena päring (Joonis 20), mille käigus selgus, et nii teenuskonto kui Kubecosti IAM roll töötasid korrektselt.

Mitmetunnise tõrkeotsingu käigus ning Kubecosti kasutajatoega suhtlemise abil selgus, et *Spot Data Feed* ja Athena päringu tulemuste S3 *bucket*-ite defineerimine toimis erinevalt, sest ühe puhul eeldati eesliidet *s3://* ning teise puhul mitte, kuid ametlikus dokumentatsioonis sellekohane täpsustus puudus. Seetõttu paluti lõputöö autori poolt ka vastava täpsustuse lisamist töövahendi dokumentatsiooni.

Pärast korrektuuride tegemist ning töövahendi konteineri taaskäivitamist, õnnestus Kubecosti ja CUR raporteerimise vaheline liidestus.

Tehnilise teostuse käigus rohkem probleeme ei tekkinud ning töövahendi paigaldamine õnnestus.

6 Tulemused

Käesolevas peatükis hinnatakse, kas lõputöös seatud nõuete defineerimise abil ning olemasolevate lahenduste analüüsi käigus valiti välja sobiv töövahend lõputöö eesmärgi täitmiseks. Lisaks hinnatakse tehnilise teostuse käigus valminud meeskonnapõhise kuluarvestuse lahenduse funktsionaalsust ning tuuakse välja võimalikud edasiarenduse suunad tulevikus.

6.1 Valitud töövahendi sobilikkuse hindamine

Sobiliku töövahendi valimisel teostati ülevaade olemasolevatest vabavaralistest ning omandatarkvaralistest lahendustest, teostati nende analüüs ning võrreldi nende funktsionaalsusi lõputöös defineeritud nõuetega. Võrdluse käigus selgus, et Kubecost vastas kõigile nõuetele ning omas suurel hulgal ametlikku dokumentatsiooni, mis lisas kindlust soovitud eesmärgi täitmiseks ning seetõttu leidis lõputöö autor, et välja valitud töövahend sobis meeskonnapõhise kuluarvestuse saavutamiseks.

6.2 Töövahendi funktsionaalsuse hindamine

Töövahendi funktsionaalsuse hindamisel lähtuti lõputöö kirjeldatud probleemist ning loodava lahenduse analüüsi peatükis defineeritud nõuetest ning võrreldi neid valminud lahenduse funktsioonidega. Lõputöö tulemusena valmis tehniline lahendus, mis andis meeskondadele detailse ülevaate nende pilvekasutuse kuludest Kuberneteses ja ülejäänud AWSi teenustes. Kubernetese puhul eristati kulusid protsessori, muutmälu, salvestusseadmete, koormusjaoturite ja võrguliikluse järgi ning ülejäänud AWSi teenuste puhul teenuse nime järgi (Lisa 5, 6).

Samuti võimaldas lahendus regiooniülest AWSi teenuste kulujaotust ning Kubernetese rakenduspõhist kulujaotust meeskonna vaates (Lisa 7, 8). Kõigile meeskondadele võimaldati virtuaalse privaattõrgu kaudu ligipääs töövahendile, millega tagati võimalus kohandatud kuluraportite loomiseks.

Valminud lahendus täitis kõiki lõputöös püstitatud nõudeid ning sellega täideti ettevõtte eesmärk meeskonnapõhise kuluarvestuse võimaldamise näol. Lisaks oli teostatud lahendus ettevõtteülene, ning seetõttu on seda võimalik ka teiste ettevõtete jaoks

kasutada. Kuna antud lahendus toetab ka Microsoft Azure ning Google Cloud Platform pilveplatvormide kuluarvestust, ei ole see seotud vaid AWSiga. Seetõttu on seda võimalik kasutada ka ettevõtetel, kelle pilvetaristuks on Microsoft Azure või Google Cloud Platform teenused.

6.3 Võimalikud edasiarendused tulevikus

Valminud lahenduse üheks edasiarenduseks võiks olla RBAC õiguste ja SSO juurutamine Kubecosti tasulise versiooni abil, mis võimaldaks raportitele rollipõhiseid ligipääsusi. Lisaks meeskonnapõhisele kuluarvestusele pakutakse Kubecosti kuluarvestuse algoritmide abil potentsiaalseid kokkuhoiuvõimalusi [87], mis ei olnud küll antud lõputöö skoobis, kuid mille võiks tulevikus üheks kulude optimeerimise sisendiks võtta.

FinOps'i teises etapis saaks juurutada valminud töövahendi automaatsete hoiatusteadete edastamise funktsiooni, mis teavitaks kliente anomaaliatest ning potentsiaalsetest eelarve ületamistest [37].

7 Kokkuvõte

Antud lõputöö eesmärk oli leida ning realiseerida tehniline lahendus meeskonnapõhise infrastruktuuri kuluarvestuseks, mille abil muuta pilveteenuste kulud arendusmeeskondadele läbipaistvamaks. Lõputöö raames toodi välja taustainfo ning kirjeldati üksikasjalikult olemasolevat probleemi ning lõputöö eesmärki ja selle saavutamiseks seatud nõudeid.

Täpse ülevaate saamiseks tutvustati ning analüüsiti juba eksisteerivaid vaba- ja omandtarkvaralisi töövahendeid ning võrreldi nende iseärasusi, võimalusi, puudusi ja võimalikke kulusid. Selle käigus selgitati välja, kas neid oli võimalik antud lõputöö eesmärgi saavutamiseks kasutada.

Analüüsi käigus võrreldi ka erinevate lahenduste paigaldusvõimalusi, mille abil saadi detailne ülevaade koos teadmistega nende keerukuse ja võimekuse kohta. Analüüsi tulemusi võrreldi lõputöös seatud nõuetega, mille abil valiti välja tehniliseks lahenduseks sobiv töövahend.

Eesmärgini jõudmist kirjeldati erinevate etappidena, milles olid loetletud vajalikud ettevalmistused AWSi ja Kubernetese objektide märgiste (*tag*) lisamiseks, arhitektuuri kirjeldamine ning tehnilise lahenduse realiseerimine, mis oli võimeline loodud märgiste abil eristama meeskonnapõhiseid kulusid. Töö tulemuse osas kirjeldati ka testimist koos saadud tulemustega ning lõputöö skoobist välja jäävaid võimalusi ja edasisi samme tulevikus.

Lõputöö tulemusena valmis Entigo OÜ kliendile tehniline lahendus Kubecost töövahendi näol, millega võimaldati meeskonnapõhine infrastruktuuri kuluarvestus, mida oleks võimalik ka teiste ettevõtete jaoks kasutada.

Kasutatud kirjandus

- [1] Kubernetes, „Kubernetes Documentation,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://kubernetes.io/docs/home/>. [Kasutatud 14.04.2022].
- [2] Eurostat, „Cloud computing - statistics on the use by enterprises,“ 2021. [Võrgumaterjal]. Kättesaadav: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cloud_computing_-_statistics_on_the_use_by_enterprises#Use_of_cloud_computing:_highlights. [Kasutatud 12.03.2022].
- [3] Cloud Native Computing Foundation, „Annual Survey 2021,“ 2021. [Võrgumaterjal]. Kättesaadav: https://www.cncf.io/wp-content/uploads/2022/02/CNCF-AR_FINAL-edits-15.2.21.pdf. [Kasutatud 12.03.2022].
- [4] C. R. Jayachander Surbiryala, „Cloud Computing: History and Overview,“ 2019. [Võrgumaterjal]. Kättesaadav: <https://ieeexplore.ieee.org/document/9045506>. [Kasutatud 18.03.2022].
- [5] S. L. Garfinkel, „The Cloud Imperative,“ *Technology Review*, kd. 114, nr 6, pp. 74-76, 2011.
- [6] V. Rajaraman, „Cloud computing,“ *Resonance*, kd. 19, nr 3, pp. 242-258, 2014.
- [7] Amazon Web Services, „About AWS,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/about-aws/>. [Kasutatud 18.03.2022].
- [8] Sparkfabrik, „Microservices and Cloud Native Applications vs. Monolithic Applications,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://blog.sparkfabrik.com/en/microservices-and-cloud-native-applications-vs-monolithic-applications>. [Kasutatud 29.03.2022].
- [9] CNCF, „CNCF Kubernetes Project Journey Report,“ 2020. [Võrgumaterjal]. Kättesaadav: https://www.cncf.io/wp-content/uploads/2020/08/CNCF_Kubernetes_Project_Journey_Report.pdf. [Kasutatud 29.03.2022].
- [10] Datadog, „2021 Container Report,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://www.datadoghq.com/container-report/#1>. [Kasutatud 19.03.2022].
- [11] Pepperdata, Inc., „Enterprises Should Consider Optimization to Contain Costs and Improve Performance,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://www.pepperdata.com/pepperdata-new-survey-cloud-budgets>. [Kasutatud 20.03.2022].
- [12] M. F. J.R. Storment, „Cloud FinOps,“ 2019. [Võrgumaterjal]. Kättesaadav: https://learning.oreilly.com/library/view/cloud-finops/9781492054610/ch01.html#the_definition. [Kasutatud 20.03.2022].
- [13] M. F. J.R. Storment, „Chapter 6. The FinOps Lifecycle,“ 2019. [Võrgumaterjal]. Kättesaadav: <https://learning.oreilly.com/library/view/cloud-finops/9781492054610/ch06.html#inform>. [Kasutatud 12.04.2022].

- [14] FinOps Foundation, „FinOps Phases,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.finops.org/framework/phases/>. [Kasutatud 20.03.2022].
- [15] M. F. J.R. Storment, „The FinOps Lifecycle,“ 2019. [Võrgumaterjal]. Kättesaadav: https://learning.oreilly.com/library/view/cloud-finops/9781492054610/ch06.html#where_do_you_startquestion_mark. [Kasutatud 20.03.2022].
- [16] M. F. J.R. Storment, „FinOps Reports Should Be Accessible and Timely,“ 2019. [Võrgumaterjal]. Kättesaadav: https://learning.oreilly.com/library/view/cloud-finops/9781492054610/ch06.html#finops_reports_should_be_accessible_and. [Kasutatud 20.03.2022].
- [17] AWS, „AWS Cost Management User Guide,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cost-management/latest/userguide/cost-management-guide.pdf#ce-what-is>. [Kasutatud 26.03.2022].
- [18] AWS, „Analyzing your costs with AWS Cost Explorer,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cost-management/latest/userguide/ce-what-is.html>. [Kasutatud 26.03.2022].
- [19] AWS, „Using Cost Allocation Tags,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/cost-alloc-tags.html>. [Kasutatud 16.04.2022].
- [20] AWS, „Security best practices in IAM,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html#lock-away-credentials>. [Kasutatud 14.04.2022].
- [21] AWS, „AWS Cost & Usage Report,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/aws-cost-management/aws-cost-and-usage-reporting/>. [Kasutatud 26.03.2022].
- [22] AWS, „What are AWS Cost and Usage Reports?,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cur/latest/userguide/what-is-cur.html>. [Kasutatud 26.03.2022].
- [23] AWS, „How do I ingest and visualize the AWS Cost and Usage Report (CUR) into Amazon QuickSight?,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/premiumsupport/knowledge-center/quicksight-cost-usage-report/>. [Kasutatud 16.04.2022].
- [24] AWS, „AWS Pricing Calculator,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://calculator.aws/#/createCalculator/QuickSight>. [Kasutatud 14.04.2022].
- [25] AWS, „Cost & Usage Report Now Available to Member (Linked) Accounts,“ 2020. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/about-aws/whats-new/2020/12/cost-and-usage-report-now-available-to-member-linked-accounts/>. [Kasutatud 14.04.2022].
- [26] L. Gil, „CAST AI vs. Kubecost: A Perfect Match for End-to-End Cloud Cost Optimization,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://cast.ai/blog/kubecost-vs-cast-ai-a-perfect-match-for-end-to-end-cloud-cost-optimization/>. [Kasutatud 26.03.2022].
- [27] Crunchbase, „CAST AI,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.crunchbase.com/organization/cast-ai>. [Kasutatud 26.03.2022].
- [28] A. T. T. K. Žilvinas Urbonas, „How CAST AI uses KEDA for Kubernetes autoscaling,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://keda.sh/blog/2021-08-04-keda-cast-ai/#keda--horizontal-pod-autoscaler>. [Kasutatud 26.03.2022].

- [29] CAST AI, „Pricing,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://cast.ai/pricing/>. [Kasutatud 26.03.2022].
- [30] CAST AI, „Connect cluster,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.cast.ai/getting-started/external-cluster/eks/eks/#connect-cluster>. [Kasutatud 26.03.2022].
- [31] CAST AI, „How to Reduce Your Amazon EKS Costs by Half in 15 Minutes — CAST AI,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://aws.plainenglish.io/how-to-reduce-your-amazon-eks-costs-by-half-in-15-minutes-cast-ai-1172b794a388?gi=16eaf9515e0f>. [Kasutatud 09.04.2022].
- [32] CAST AI, „January 2022 Release notes,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://cast.ai/release-notes/>. [Kasutatud 14.04.2022].
- [33] W. Brown, „Effectively Managing Kubernetes Resources with Cost Monitoring,“ 2018. [Võrgumaterjal]. Kättesaadav: <https://medium.com/kubecost/effectively-managing-kubernetes-with-cost-monitoring-96b54464e419>. [Kasutatud 27.03.2022].
- [34] W. Brown, „Introducing Kubecost: a better approach to Kubernetes cost monitoring,“ 2019. [Võrgumaterjal]. Kättesaadav: <https://medium.com/kubecost/introducing-kubecost-a-better-approach-to-kubernetes-cost-monitoring-b5450c3ae940>. [Kasutatud 27.03.2022].
- [35] Stackwatch, Inc, „Network Traffic Cost Allocation,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407595973527#kubernetes-network-traffic-metrics>. [Kasutatud 28.03.2022].
- [36] Stackwatch, Inc, „Kubecost Product Features,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.kubecost.com/features/#unified-cost-monitoring>. [Kasutatud 28.03.2022].
- [37] Stackwatch, Inc, „Kubecost Alerts,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407601796759-Alerts-Documentation>. [Kasutatud 28.03.2022].
- [38] M. Dresser, „Kubecost Cost-Model,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://github.com/kubecost/cost-model/blob/develop/kubecost-exporter.md#running-kubecost-as-a-prometheus-metric-exporter>. [Kasutatud 31.03.2022].
- [39] Stackwatch, Inc., „AWS Out of Cluster,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407596810519#step-2-tag-your-resources>. [Kasutatud 15.04.2022].
- [40] Stackwatch, Inc., „Kubernetes Cost Allocation,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407601807383-Kubernetes-Cost-Allocation#kubernetes-cost-allocation>. [Kasutatud 15.04.2022].
- [41] Stackwatch, Inc., „AWS Cloud Integration,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407595928087-AWS-Cloud-Integration#aws-cloud-integration>. [Kasutatud 15.04.2022].
- [42] Stackwatch, Inc., „Multi-cluster,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407595970711-Multi-cluster#multi-cluster>. [Kasutatud 15.04.2022].
- [43] Harness CCM, „The Modern Software Delivery Platform,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://harness.io/products/platform>. [Kasutatud 02.04.2022].

- [44] Harness CCM, „Pricing,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://harness.io/pricing?module=cd>. [Kasutatud 02.04.2022].
- [45] Harness CCM, „Set Up Cloud Cost Management for AWS,“ 2022. [Võrgumaterjal]. Kättesaadav: https://ngdocs.harness.io/article/80vbt5jv0q-set-up-cost-visibility-for-aws#review_cost_and_usage_reports_cur_and_ccm_requirements. [Kasutatud 02.04.2022].
- [46] H. CCM, „Set Up Cost Visibility for Kubernetes Using an Existing Delegate,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.harness.io/article/kuiuc6x257-enable-continuous-efficiency-for-kubernetes#prerequisites>. [Kasutatud 02.04.2022].
- [47] Harness, „Set Up Cloud Cost Management for AWS,“ 2022. [Võrgumaterjal]. Kättesaadav: https://ngdocs.harness.io/article/80vbt5jv0q-set-up-cost-visibility-for-aws#step_4_create_cross_account_role. [Kasutatud 15.04.2022].
- [48] Harness, „View Cluster Cost Dashboard,“ 2022. [Võrgumaterjal]. Kättesaadav: https://ngdocs.harness.io/article/uai4ud1ibi-cluster-cost-dashboard#step_view_cluster_cost_dashboard. [Kasutatud 15.04.2022].
- [49] crane, „Crane: Cloud Resource Analytics and Economics,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://github.com/gocrane/crane>. [Kasutatud 09.04.2022].
- [50] Crane, „Getting Started,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.gocrane.io/dev/#introduction>. [Kasutatud 09.04.2022].
- [51] Kubernetes, „Overview,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://github.com/kubernetes/kube-state-metrics>. [Kasutatud 09.04.2022].
- [52] Prometheus, „Node Exporter,“ [Võrgumaterjal]. Kättesaadav: https://github.com/prometheus/node_exporter#node-exporter. [Kasutatud 09.04.2022].
- [53] Gocrane, „Fadvisor: FinOps Advisor,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://github.com/gocrane/fadvisor#estimated-cluster-costs>. [Kasutatud 09.04.2022].
- [54] Gocrane, „Crane Roadmap for H1 2022,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.gocrane.io/dev/roadmaps/roadmap-1h-2022/>. [Kasutatud 15.04.2022].
- [55] Vantage, „Vantage Documentation,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.vantage.sh/#overview>. [Kasutatud 09.04.2022].
- [56] Vantage, „Setup AWS,“ 2022. [Võrgumaterjal]. Kättesaadav: https://docs.vantage.sh/connecting_aws/#connecting-your-aws-account. [Kasutatud 09.04.2022].
- [57] Vantage, „Compare plans,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.vantage.sh/pricing>. [Kasutatud 15.04.2022].
- [58] Vantage, „Kubernetes Costs,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.vantage.sh/kubernetes/#how-it-works>. [Kasutatud 15.04.2022].
- [59] Vantage, „Cost Reports,“ 2022. [Võrgumaterjal]. Kättesaadav: https://docs.vantage.sh/cost_reports/. [Kasutatud 10.04.2022].
- [60] Gocrane, „Deploy on cloud,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://github.com/gocrane/fadvisor#deploy-on-cloud>. [Kasutatud 16.04.2022].
- [61] R. Alvarez-Parmar, „How to track costs in multi-tenant Amazon EKS clusters using Kubecost,“ 2020. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/blogs/containers/how-to-track-costs-in-multi-tenant-amazon-eks-clusters-using-kubecost/>. [Kasutatud 16.04.2022].

- [62] HashiCorp, „About the Terraform Language,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.terraform.io/language#about-the-terraform-language>. [Kasutatud 16.04.2022].
- [63] RedHat, „What is YAML?,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://www.redhat.com/en/topics/automation/what-is-yaml>. [Kasutatud 25.04.2022].
- [64] AWS, „AWS CodePipeline FAQs,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/codepipeline/faqs/?nc=sn&loc=5>. [Kasutatud 16.04.2022].
- [65] Argo CD, „Argo CD - Declarative GitOps CD for Kubernetes,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://argo-cd.readthedocs.io/en/stable/#what-is-argo-cd>. [Kasutatud 16.04.2022].
- [66] AWS, „Creating Cost and Usage Reports,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cur/latest/userguide/cur-create.html>. [Kasutatud 17.04.2022].
- [67] AWS, „Setting up Athena using AWS CloudFormation templates,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cur/latest/userguide/use-athena-cf.html>. [Kasutatud 17.04.2022].
- [68] AWS, „Creating an IAM role and policy for your service account,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/eks/latest/userguide/create-service-account-iam-policy-and-role.html>. [Kasutatud 17.04.2022].
- [69] AWS, „Create an IAM OIDC provider for your cluster,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/eks/latest/userguide/enable-iam-roles-for-service-accounts.html>. [Kasutatud 17.04.2022].
- [70] AWS, „Spot Instance data feed,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/spot-data-feeds.html>. [Kasutatud 17.04.2022].
- [71] AWS, „AssumeRole,“ 2022. [Võrgumaterjal]. Kättesaadav: https://docs.aws.amazon.com/STS/latest/APIReference/API_AssumeRole.html. [Kasutatud 17.04.2022].
- [72] A. Valdes, „The best Infrastructure as Code tools for 2022,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://www.clickittech.com/devops/infrastructure-as-code-tools/>. [Kasutatud 17.04.2022].
- [73] J. Ramos, „Kubernetes GitOps Tools,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://itnext.io/kubernetes-gitops-tools-cf0247eb5368?gi=8a1c7c10e482>. [Kasutatud 17.04.2022].
- [74] Stackwatch, Inc, „Documentation,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407595950359>. [Kasutatud 20.04.2022].
- [75] Helm, „Helm Template,“ 2022. [Võrgumaterjal]. Kättesaadav: https://helm.sh/docs/helm/helm_template/. [Kasutatud 20.04.2022].
- [76] Argo CD, „Declarative Setup,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://argo-cd.readthedocs.io/en/stable/operator-manual/declarative-setup/#applications>. [Kasutatud 20.04.2022].

- [77] Argo CD, „Automated Sync Policy,“ 2022. [Võrgumaterjal]. Kättesaadav: https://argo-cd.readthedocs.io/en/stable/user-guide/auto_sync/#automatic-pruning. [Kasutatud 20.04.2022].
- [78] Argo CD, „Sync Options,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://argo-cd.readthedocs.io/en/stable/user-guide/sync-options/#no-prune-resources>. [Kasutatud 20.04.2022].
- [79] AWS, „Setting up an Amazon S3 bucket for Cost and Usage Reports,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cur/latest/userguide/cur-s3.html>. [Kasutatud 20.04.2022].
- [80] AWS, „Spot Instances,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html>. [Kasutatud 21.04.2022].
- [81] M. F. J.R. Stormont, „Chapter 9. Tags, Labels, and Accounts, Oh My!,“ 2019. [Võrgumaterjal]. Kättesaadav: <https://learning.oreilly.com/library/view/cloud-finops/9781492054610/ch09.html#conclusion-id00013>. [Kasutatud 21.04.2022].
- [82] Kubernetes, „Recommended Labels,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://kubernetes.io/docs/concepts/overview/working-with-objects/common-labels/#labels>. [Kasutatud 21.04.2022].
- [83] AWS, „Tagging a private repository,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/AmazonECR/latest/userguide/ecr-using-tags.html>. [Kasutatud 21.04.2022].
- [84] AWS, „Activating User-Defined Cost Allocation Tags,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/activating-tags.html>. [Kasutatud 21.04.2022].
- [85] AWS, „The AWS Cost & Usage Report: The Next Step on Your Cost Management Journey,“ 2019. [Võrgumaterjal]. Kättesaadav: <https://aws.amazon.com/blogs/enterprise-strategy/the-aws-cost-usage-report-the-next-step-on-your-cost-management-journey/>. [Kasutatud 09.04.2022].
- [86] Stackwatch, Inc, „kubecost-masterpayer-account-permissions,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://github.com/kubecost/cloudformation/blob/master/kubecost-masterpayer-account-permissions.yaml>. [Kasutatud 21.04.2022].
- [87] Stackwatch, Inc, „Optimization Insights,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.kubecost.com/features#optimization-insights>. [Kasutatud 23.04.2022].
- [88] „How to Reduce Your Amazon EKS Costs by Half in 15 Minutes — CAST AI,“ 2021. [Võrgumaterjal]. Kättesaadav: <https://aws.plainenglish.io/how-to-reduce-your-amazon-eks-costs-by-half-in-15-minutes-cast-ai-1172b794a388?gi=16eaf9515e0f>. [Kasutatud 09.04.2022].
- [89] Inapps, „HARNESS.IO TRANSLATES CLOUD COSTS TO DEVELOPERS – INAPPS 2022,“ 2022. [Võrgumaterjal]. Kättesaadav: <https://www.inapps.net/harness-io-translates-cloud-costs-to-developers-inapps-2022/>. [Kasutatud 09.04.2022].
- [90] Crane, „Documentation,“ 2022. [Võrgumaterjal]. Kättesaadav: https://raw.githubusercontent.com/gocrane/crane/main/docs/images/first_cluster.png. [Kasutatud 09.04.2022].

- [91] Red Hat, „What is GitOps?“, 2021. [Võrgumaterjal]. Kättesaadav: <https://www.redhat.com/en/topics/devops/what-is-gitops>. [Kasutatud 16.04.2022].
- [92] Stackwatch, Inc, „Installing Kubecost“, 2022. [Võrgumaterjal]. Kättesaadav: <https://guide.kubecost.com/hc/en-us/articles/4407601821207-Installing-Kubecost#installing-kubecost>. [Kasutatud 17.04.2022].
- [93] Kubernetes, „Configure Service Accounts for Pods“, 2022. [Võrgumaterjal]. Kättesaadav: <https://kubernetes.io/docs/tasks/configure-pod-container/configure-service-account/>. [Kasutatud 21.04.2022].
- [94] RedHat, „What is YAML?“, 2021. [Võrgumaterjal]. Kättesaadav: <https://www.redhat.com/en/topics/automation/what-is-yaml>. [Kasutatud 25.04.2022].
- [95] AWS, „Cost and Usage Reports“, 2022. [Võrgumaterjal]. Kättesaadav: <https://docs.aws.amazon.com/cur/latest/userguide/cur-s3.html>. [Kasutatud 2022.26.03].

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Marko Lindeberg

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Meeskonnapõhine infrastruktuuri kuluarvestus Entigo OÜ kliendi näitel“, mille juhendaja on Kristiina Hakk ja kaasjuhendaja on Rein Remmel.
 - 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.

16.05.2022

✓

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.

Lisa 2 – Kubernetes juurutusressursi märgiste lisamise näide

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: event-database
  labels:
    app: event-database #rakenduse nimi
    team: Platform-IoT #meeskonna nimi
    product: event #tootenimetus
    env: dev-infra #keskkond
    component: database #komponendi täpsustus
spec:
  replicas: 1
  selector:
    matchLabels:
      app: event-database
  strategy:
    type: Recreate
  template:
    metadata:
      name: event-database
      labels:
        app: event-database #rakenduse nimi
        team: Platform-IoT #meeskonna nimi
        product: event #tootenimetus
        env: dev-infra #keskkond
        component: database #komponendi täpsustus
```

Lisa 3 – *Spot Data Feed* ja IAM pääsumandaadiks vajalike IAM rollide loomise TerraFormi skriptid

```
# IAM poliitika pääsumandaadi jaoks
resource "aws_iam_policy" "kubecost-iam-policy" {
  count = contains(["dev"], terraform.workspace) ? 1 : 0
  name   = "kubecost-assume-masterpayer-${local.cluster_name}"
  policy = <<EOF
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AssumeRoleInMasterPayer",
      "Effect": "Allow",
      "Action": "sts:AssumeRole",
      "Resource": "arn:aws:iam::[REDACTED]:role/KubecostRole-
${local.aws_account}"
    }
  ]
}
EOF
}

# IAM poliitika Spot Data Feed andmevoo ligipääsuks
resource "aws_iam_policy" "kubecost-spot-policy" {
  count = contains(["dev"], terraform.workspace) ? 1 : 0
  name   = "kubecost-spot-data-feed-access-${local.cluster_name}"
  policy = <<EOF
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "SpotDataAccess",
      "Effect": "Allow",
      "Action": [
        "s3:ListAllMyBuckets",
        "s3:ListBucket",
        "s3:HeadBucket",
        "s3:HeadObject",
        "s3:List*",
        "s3:Get*"
      ],
      "Resource": "${aws_s3_bucket.spot-data-feed[0].arn}"
    }
  ]
}
EOF
}
```

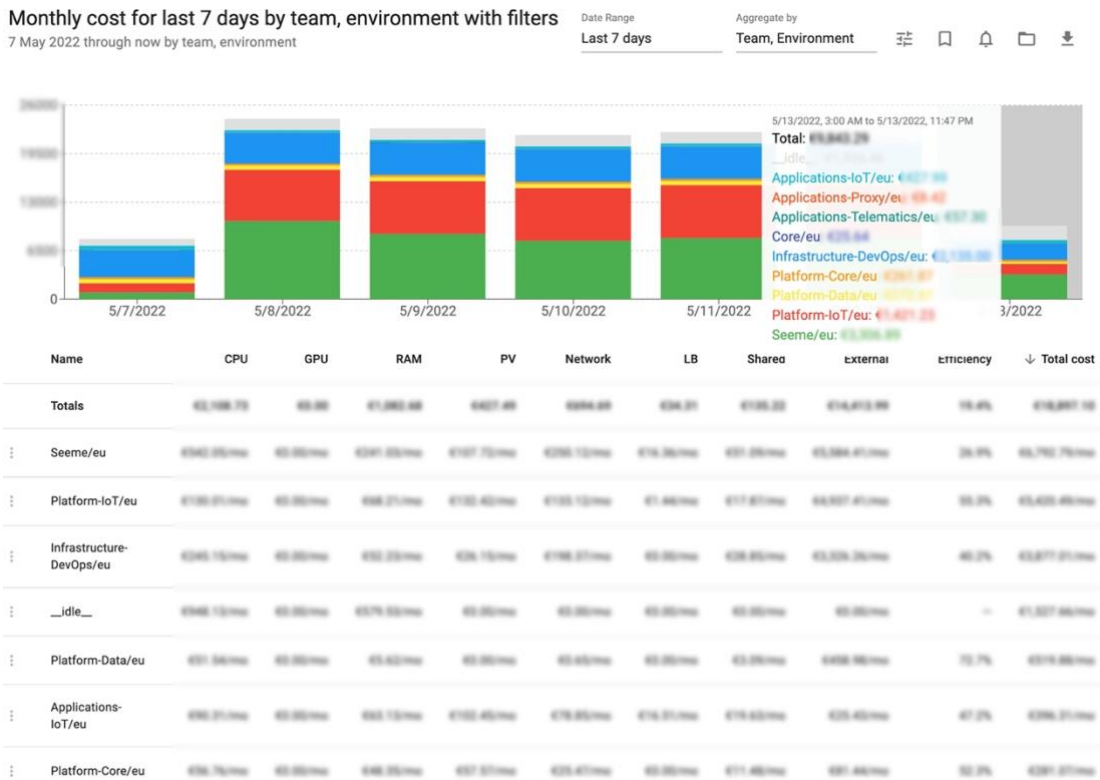
Lisa 4 – Teenuskonto sidumine IAM rolliga ja selle kontrollitud tulemus

```
# käsk teenuskonto sidumiseks IAM rolliga
kubectl annotate serviceaccount -n monitoring kubecost-cost-analyzer \
eks.amazonaws.com/role-arn=arn:aws:iam::[REDACTED]:role/KubecostRole-
[REDACTED]

# käsk teenuskonto kirjeldamiseks
kubectl describe serviceaccounts -n monitoring kubecost-cost-analyzer

# teenuskonto kirjelduse väljund, mille märgistest ja annotatsioonist
on näha, et konto on IAM rolliga seotud
Name: kubecost-cost-analyzer
Namespace: monitoring
Labels: app=cost-analyzer
        app.kubernetes.io/instance=kubecost
        app.kubernetes.io/managed-by=eksctl
        product=kubecost
        type=monitoring
Annotations: eks.amazonaws.com/role-arn:
arn:aws:iam::[REDACTED]:role/KubecostRole-[REDACTED]
Image pull secrets: <none>
Mountable secrets: kubecost-cost-analyzer-token-pfkkg
Tokens: kubecost-cost-analyzer-token-pfkkg
Events: <none>
```

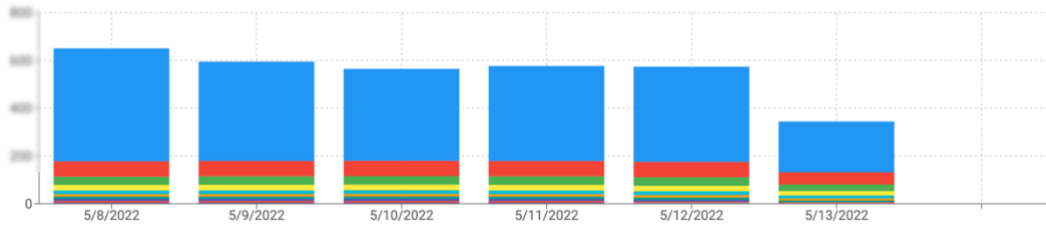
Lisa 5 – Meeskonnapõhine Kubernetese klastri ning summeritud AWSi teenuste (External veerg) kulujaotus ühe keskkonna kohta kuu lõikes (eemadatud kuluväärtustega)



Lisa 6 – Meeskonnapõhine detailne AWSi teenuste kulujaotus seitsme päeva lõikes (eemaldatud kuluväärtustega)

Last 7 days by service daily with filters
8 May 2022 through now by Service

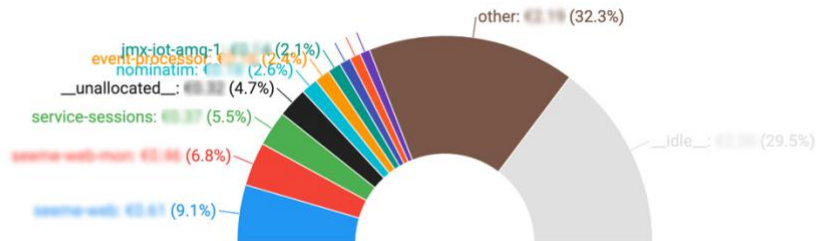
Date Range: Last 7 days | Breakdown: Service



Name	Credits	Adjusted	Total cost
Totals	€0.00	€0.00	€3,298.71
AmazonEC2 <code>kubernetes_cluster=...</code> <code>kubernetes_label_env=eu</code>	€0.00	€0.00	€3,298.71
AmazonES <code>kubernetes_cluster=...</code> <code>kubernetes_label_env=eu</code> <code>kubernetes_label_team=Infrastructure-DevOps</code>	€0.00	€0.00	€375.87
AmazonMSK <code>kubernetes_cluster=...</code> <code>kubernetes_label_env=eu</code>	€0.00	€0.00	€228.87
AmazonEFS <code>kubernetes_cluster=...</code> <code>kubernetes_label_env=eu</code> <code>kubernetes_label_team=Infrastructure-DevOps</code> <code>map_migrated=d...</code>	€0.00	€0.00	€134.18
AmazonRedshift <code>kubernetes_cluster=...</code> <code>kubernetes_label_env=eu</code> <code>kubernetes_label_team=Platform-Data</code>	€0.00	€0.00	€96.81
AmazonElastiCache <code>kubernetes_cluster=...</code> <code>kubernetes_label_env=eu</code>	€0.00	€0.00	€60.87

Lisa 7 – Rakenduspõhine kulujaotus ühe tunni lõikes (eemaldatud kuluväärtustega)

8 May 2022 through now by service



Name	CPU	GPU	RAM	PV	Network	LB	Shared	External	Efficiency	Total cost
Totals	€1.01	€0.00	€1.96	€0.75	€1.17	€0.00	€0.00	€0.00	21.2%	€6.78
idle	€1.24/yr	€0.00/yr	€0.76/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	–	€2.00/yr
service-web	€0.26/yr	€0.00/yr	€0.76/yr	€0.00/yr	€0.07/yr	€0.00/yr	€0.00/yr	€0.00/yr	30.5%	€0.61/yr
service-web-main	€0.26/yr	€0.00/yr	€0.12/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	29.9%	€0.46/yr
service-sessions	€0.24/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.26/yr	€0.00/yr	€0.00/yr	€0.00/yr	41.3%	€0.07/yr
unallocated	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	€0.00/yr	19.4%	€0.00/yr

Lisa 8 – Rakenduse- ja meeskonnapõhine prognoositud kuluarvestus kuu lõikes, sorteeritud protsessori ressursikasutuse järgi (eemaldatud kuluväärtustega)

