



AWS QuickSight ja datajärvet: Tietojärvidatan visualisointi Quicksightissa

Juho Leppäkangas

Opinnäytetyö, AMK

5 2025

Tietojenkäsittelyn tutkinto-ohjelma

Leppäkangas, Juho

AWS Quicksight ja datajärvet: Tietojärvidatan visualisointi AWS Quicksightissa

Jyväskylä: Jyväskylän ammattikorkeakoulu. Toukokuu 2025, 35 sivua

Tietojenkäsittelyn tutkinto-ohjelma. Opinnäytetyö AMK.

Julkaisun kieli: suomi

Julkaisulupa avoimessa verkossa: kyllä

Tiivistelmä

Maailmanlaajuisesti dataa tuotetaan kokoajan yhä enemmän ja enemmän. Yritykset siis keräävät ja käsittelevät valtavasti massadataa, joten datajävrien hyödyntäminen massadatan säilömiseen on noussut keskeiseksi ratkaisuksi. Jotta pystytään ymmärtämään datan merkitystä ja tehdä päätöksiä sen perusteella, täytyy raakadata saada ymmärrettävään muotoon.

Opinnäytetyön tavoitteena oli luoda kolme interaktiivista koontinäyttöä Quicksight palvelulla datajärveen kerätystä datasta. Työ toteutettiin kahdella eri tavalla: Ensin hyödynnettiin valmista datajoukkoa, jonka jälkeen syvennettiin osaamista hakemalla reaaliaikaista dataa ulkoisista lähteistä eri AWS-palveluiden avulla.

Työn tuloksina syntyi visuaalisia ja interaktiivisia koontinäyttöjä tutkimuksessa käytetyistä NHL Draft- sekä Sportradarista haetusta datasetistä.

Opinnäytetyön tulosten perusteella onnistuttiin luomaan monipuolisia visualisointeja datan perusteella sekä syventämään tietoa Quicksightin ominaisuuksista sekä sen kyvyistä datan analysoimiseen.

Avainsanat (asiasanat)

AWS, Quicksight, S3, Business intelligence, datan analysointi, visualisointi

Muut tiedot (salassa pidettävät liitteet)

Leppäkangas, Juho

AWS Quicksight and datalakes: Visualizing data lake data in AWS Quicksight

Jyväskylä: JAMK University of Applied Sciences, May 2025, 35 pages

Business administration. Degree programme in business information systems. Bachelor's thesis.

Permission for open access publication: Yes

Language of publication: Finnish

Abstract

Globally, the amount of data being generated is constantly increasing. Companies collect and process vast amount of big data, making data lakes a key solution for storing this data. To understand the significance of data and make decisions based on it, raw data must be transformed into an understandable format.

The objective of the thesis was to create three different dashboards using the Quicksight service, based on data stored in a data lake. The work was carried out in two ways: First, by utilizing an existing dataset, and then by deepening expertise through retrieving real-time data from external sources using various AWS services.

As a result of the thesis, visual and interactive dashboards were created from the NHL Draft dataset used in the study.

Based on the results, the thesis successfully produced versatile visualizations from the data and enhanced knowledge of Quicksight's features and its capabilities for data analysis.

Keywords/tags (subjects)

AWS, Quicksight, S3, Business intelligence, data analysis, visualization

Miscellaneous (Confidential information)

Sisältö

1	Johdanto	8
2	Tutkimusasetelma	9
2.1	Tutkimusmenetelmät	10
2.2	Tavoitteet ja aiheen rajaus	10
2.3	Toimeksiantaja	11
2.4	Tutkimuskysymykset	11
2.5	Tiedonhaku	11
3	AWS	12
4	Käytetyt AWS palvelut	13
4.1	S3 (Simple Storage Service)	13
4.2	Amazon Q palvelu	15
4.3	Quicksight palvelu	16
4.4	AWS Lambda	22
5	Toteutus	23
5.1	Datasetin hakeminen ja tallennus	23
5.2	Visualisointi	24
5.3	Muuttuvan datan visualisointi	27
5.4	Valmiit koontinäytöt	31
6	Opinnäytetyön tulokset	33
7	Pohdinta	34
7.1	Eettisyys ja luotettavuus	35
7.2	Kehittämissuhteet	36
	Lähteet	37

Kuviot

Kuva 1	Tuotetun datan määrä (Duarte 2025)	13
Kuva 2.	SPICE-tilan ja Direct Queryn ero (Yung, 2024)	19
Kuva 3	Manifest-tiedosto JSON	24
Kuva 4	Lasketun kentän lyhenteet	25
Kuva 5	Funktio kuvaamaan menetystä	26
Kuva 6	Epäonnistujat kuvattuna	26
Kuva 7	Muuttuvan datan infrastruktuurikuva	28
Kuva 8	EventBridge ajoitus	30

Kuva 9 Muuttuvan datan visualisointi	31
Kuva 10 Koontinäyttö 1.....	32
Kuva 11 Koontinäyttö 2.....	32
Kuva 12 Amazon Q kysely	34

Lyhenteet ja käsitteet

Datajärvet

Datajärvi (data lake) on pilvipohjainen tallennusratkaisu, joka mahdollistaa tietojen tallentamisen alkuperäisessä muodossaan. Datajärven pohjana toimii S3-palvelu.

S3

Amazon S3 (Simple Storage Service) on objektien tallennuspalvelu. S3-palvelu tarjoaa korkean saatavuuden ja kestävyuden (99,999999999%), johon voidaan tallentaa käytännössä rajattomasti erilaisia objekteja, kuten kuvia, videoita, tiedostoja.

Amazon Q

Generatiivinen tekoälypohjainen avustaja, joka mahdollistaa luonnollisen kielen kyselyt tiedon hakemiseen, koodin generointiin sekä datan analysointiin.

Lambda

Lambda on palvelimeton laskentapalvelu, joka mahdollistaa koodin suorittamisen ilman palvelimien hallintaa. Lambdan avulla voidaan suorittaa koodia erilaisiin tapahtumiin, kuten tiedostojen tallentamiseen, API-kutsuihin sekä sillä voidaan ajastaa erilaisia tehtäviä.

Application Programming Interface = API

API on rajapinta, joka mahdollistaa eri ohjelmistojen ja palveluiden välisen tiedonvaihdon. API-kutsujen avulla voidaan esimerkiksi hakea dataa ulkoisista lähteistä.

Datajoukko

myös tietojoukko on kokoelma dataa, joka sisältää sarakkeita ja rivejä taulukkomuodossa. Datajoukkoja käytetään muun muassa datan analysoinnissa ja visualisoinneissa.

S3-bucket

S3-bucket toimii yksilöllisenä kansiona, johon voidaan tallentaa erilaisia objekteja.

Quicksight

Pilvipalvelu datan analysointiin ja visualisointiin.

SPICE

SPICE (Super-Fast Parallel In-memory Calculation Engine) on Quicksight-palvelun sisäinen muistinvarainen laskentamoottori, joka tallentaa tuodun datajoukon Quicksight:n sisäiseen muistiin, jolloin kyselyjen tekeminen on nopeaa.

1 Johdanto

Maailmanlaajuisesti dataa tuotetaan yhä enemmän ja enemmän. Statistan tekemän tutkimuksen mukaan vuonna 2010 tuotetun datan kokonaismäärä oli noin 2 zettatavua, kun sen ennustetaan nousevan yli 182 zettatavuun vuoden 2025 aikana. (Singh 2025.) Yritykset keräävät ja käsittelevät siis valtavia määriä dataa, jota hyödynnetään päätöksien tekemisessä sekä analytiikassa. Datajärvet ovat nousseet keskeiseksi ratkaisuksi suurten datamäärien hallintaan. Amazon Web Services tarjoaa tähän Simple Storage Service lyhyemmin S3-palvelua, joka on erittäin skaalautuva sekä kustannustehokas objektien tallennusratkaisu. Datajärviä hyödynnetään tänä päivänä paljon, koska ne mahdollistavat raakadatan tallentamisen ja säilyttämisen alkuperäisessä muodossaan.

Datan visualisointi on noussut myös tärkeäksi tekijäksi organisaatioiden viestinnässä ja päätöksien tekemisessä. AWS:n tarjoama työkalu Quicksight on pilvipohjainen business intelligence (BI) työkalu, joka mahdollistaa datan visualisoinnin selkeiksi kuviksi, jolloin mahdollisimman monet organisaation jäsenet pystyvät keräämään tietoa ja tekemään päätöksiä datan pohjalta. Quicksightilla pystytään luomaan interaktiivisia koontinäyttöjä sekä sillä voidaan hakea analysoitavaa dataa useista eri lähteistä.

Työn toimeksiantaja Skillwell Oy on Jyväskyläläinen vuonna 2018 perustettu IT-alan yritys, joka on erikoistunut Amazon Web Services- pilvipalveluihin sekä migraatioihin. Toimeksiantajan tavoitteena oli selvittää Quicksightin ominaisuudet sekä millä tavoin yritys voisi hyötyä tästä palvelusta.

Tässä opinnäytetyössä keskitytään yksinkertaisen datajärven toimintaan sekä Quicksight palvelun toimintoihin ja ominaisuuksiin. Työssä pyritään vastaamaan kahteen tutkimuskysymykseen Quicksightin ominaisuuksista, miten Amazon Q:ta voidaan hyödyntää koontinäyttöjen rakentamisessa sekä miten muuttuvan datan visualisointi onnistuu Quicksightilla.

Työn tavoitteena on selvittää S3-datajärven kyvykkyydet sekä Quicksightin ominaisuuksien testaamisen NHL Draft -datasetin avulla. Tavoitteena on rakentaa toimiva arkkitehtuuri, joka yhdistää S3-palvelun sekä Quicksightin visualisointityökalun. Työssä keskitytään ensin valmiin datasetin tallennusprosessiin sekä datasetin visualisointiin luomalla kolme interaktiivista koontinäyttöä NHL Draft datasetin pohjalta. Tämän jälkeen luodaan infrastruktuuri, jolla pystytään tarkastelemaan visualisointeja datasetin muuttuessa päivittäin. Työ valittiin sen ajankohtaisuuden vuoksi, jolloin erilaista

dataa syntyy valtavia määriä. Kerätyn datan avulla voidaan tehostaa liiketoimintaa ja tukea päätöksentekoa visualisointien avulla entistä tehokkaammin.

2 Tutkimusasetelma

Tutkimuksen tarkoituksena oli selvittää, miten Amazon Q:ta voidaan hyödyntää koontinäyttöjen rakentamisessa ja millä tavoilla interaktiivisuutta voidaan lisätä koontinäyttöihin erilaisilla parametreilla ja suodattimilla sekä kuinka muuttuvan datan visualisointi onnistuu Quicksight-palvelulla

ja mitä se vaatii. Tutkimus toteutettiin kahdella erilaisella datasetillä. Aluksi valmiin datasetin avulla kehitettiin prototyyppi NHL Draft tietojoukosta, joka tallennettiin S3-bucketiin yksinkertaisena datajärvenä ja testaamalla Quicksightin interaktiivisia ominaisuuksia, kuten erilaisia suodattimia sekä parametreja ja Amazon Q:ta. Tämän jälkeen haluttiin luoda infrastruktuuri, jonka avulla pystytään visualisoimaan muuttuvaa dataa.

2.1 Tutkimusmenetelmät

Opinnäytetyö tehdään tutkimuksellisenä kehittämistyönä, jossa tutkitaan laadullisia ominaisuuksia. Työssä pyritään ymmärtämään kokonaisvaltaisesti käytettyjen palveluiden hyödyt sekä ominaisuudet. Tämä tutkimusmenetelmä on tapa, jonka avulla ilmiötä pyritään ymmärtämään syvällisesti ja kokonaisvaltaisesti. Kehittämistutkimuksessa voidaan luoda aivan uutta tai parantaa olemassa olevia palveluita, joista voidaan saada liiketoiminnallisia hyötyjä toimeksiantajalle.

2.2 Tavoitteet ja aiheen rajaus

Työn tavoitteena on selvittää Quicksight palvelun ominaisuudet ja kyvyn kommunikoida yksinkertaisen datajärven kanssa NHL Draft -datasetin avulla. Tavoitteena on rakentaa toimiva arkkitehtuuri, joka yhdistää S3-palvelun sekä Quicksightin visualisointityökalun. Työssä keskitytään ensin valmiin datasetin tallennusprosessiin sekä datasetin visualisointiin luomalla kolme interaktiivista koontinäyttöä NHL Draft datasetin pohjalta käyttäen myös Amazon Q tekoälypalvelua koontinäytöjen luontiin. Tämän jälkeen luodaan infrastruktuuri, jolla pystytään tarkastelemaan visualisointeja datasetin muuttuessa. Työ valittiin sen ajankohtaisuuden vuoksi, jolloin dataa syntyy enenevässä määrin sekä halua oppia data analytiikasta enemmän.

Aihe on rajattu tietojoukon tallentamiseen S3-palveluun sekä Quicksightin yhdistäminen S3-palvelun kanssa, sisältäen myös infrastruktuurin muuttuvan datan visualisointiin Quicksightissa. Työssä ei keskitytä datan esikäsittelyyn, koska tietojoukot, joita käytettiin oli valmiiksi hyvin jäsennettyä, joten tälle ei ollut erillistä tarvetta.

Työssä käytetyt palvelut edistävät kestävästä kehitystä sekä ekologisesta että sosiaalisesta näkökulmasta tukemalla samalla taloudellista kustannustehokkuutta. Ekologisesti nämä serverless-palvelut vähentävät fyysisen infrastruktuurin tarpeita, mikä parantaa muunmuassa energiatehokkuutta ja tukee näin AWS:n hiilineutraaleja tavoitteita vuoteen 2040 mennessä (AWS Sustainability 2025). Sosiaalisesti Quicksight-palvelu parantaa tiedolla johtamista tarjoamalla helposti ymmärrettäviä koontinäyttöjä käyttäjilleen.

2.3 Toimeksiantaja

Työn toimeksiantajana on Jyväskyläläinen IT-alan yritys Skillwell Oy, joka on perustettu vuonna 2018. Skillwell on Amazon Web Services (AWS) – pilvipalveluihin keskittyvä yritys, joka tarjoaa asiakkailleen digitaalisia palveluita ja niiden kehittämistä. Yhtiön Skillwell Oy liikevaihto oli 1,46 miljoonaa ja tilikauden tulos 47 tuhatta vuonna 2023 (Finder 2025).

2.4 Tutkimuskysymykset

Opinnäytetyö tehdään laadullisena kehittämistutkimuksena. Opinnäytetyössä pyritään vastaamaan kahteen tutkimuskysymykseen, jotka ovat:

1. Miten Amazon Q:ta voidaan hyödyntää Quicksightissa?
2. Miten lähes reaaliaikaisen datan visualisointi toimii quicksightissa?

2.5 Tiedonhaku

Tiedonhakua on toteutettu käyttämällä pääsääntöisesti kolmea erilaista lähdettä. AWS:n omat dokumentaatiot verkkosivuilla, AWS sertifioidujen ammattilaisten kirjoittamia artikkeleita, jotka tuovat omia näkökulmia palveluiden käyttöön. Myös JanetFinna -tietokannassa olevaa kirjallisuutta. Lähteitä on arvioitu kriittisesti ja pyritty etsimään luotettavia lähteitä käyttämällä erilaisia hakusanoja liittyen tietojärviin, massadataan sekä Quicksight palveluun. Lähteitä valittaessa on vertailtu useita eri lähteitä ja pyritty valitsemaan ajankohtaisia ja uusimpia lähteitä tietoperustaan, sillä tiedot ja palveluiden ominaisuudet päivittyvät ja kehittyvät jatkuvasti. Näin ollen lähteiden on hyvä olla ajantasaisia työn laadun varmistamiseksi.

3 AWS

Amazon Web Services lyhyemmin AWS julkaisi heidän ensimmäisen verkkopalvelunsa vuonna 2002 hoitamaan verkkokaupan toimintaansa. Vuonna 2006 AWS alkoi tarjoamaan ensimmäiseksi IaaS-palveluita asiakkailleen. AWS tarjoaa virtuaalikoneita, tallennusratkaisuja ja verkkopalveluita. AWS tarjoaa yhteensä yli 200 erilaista palvelua ja on näin pilvipalvelun tarjoajien markkinajohtaja noin 32 prosentin markkinaosuudella muihin verrattuna (Reno 2024). AWS oli ensimmäinen pilvipohjaisten palveluiden kehittäjä, joka tarjosi asiakkailleen ensimmäisenä hinnoittelumallin, jossa yritys maksaa vain käytetyistä resursseista (Kirvan 2024).

AWS tarjoaa kolmea erilaista pilvipalvelu tyyppiä, jotka lueteltu alla.

IaaS = Infrastructure as a Service on palvelutyyppi, jossa asiakkaat voivat vuokrata it-infrastruktuuria, kuten virtuaalikoneita, tallennustilaa sekä muita verkkoresursseja omien tarpeidensa mukaan (Hurja 2024).

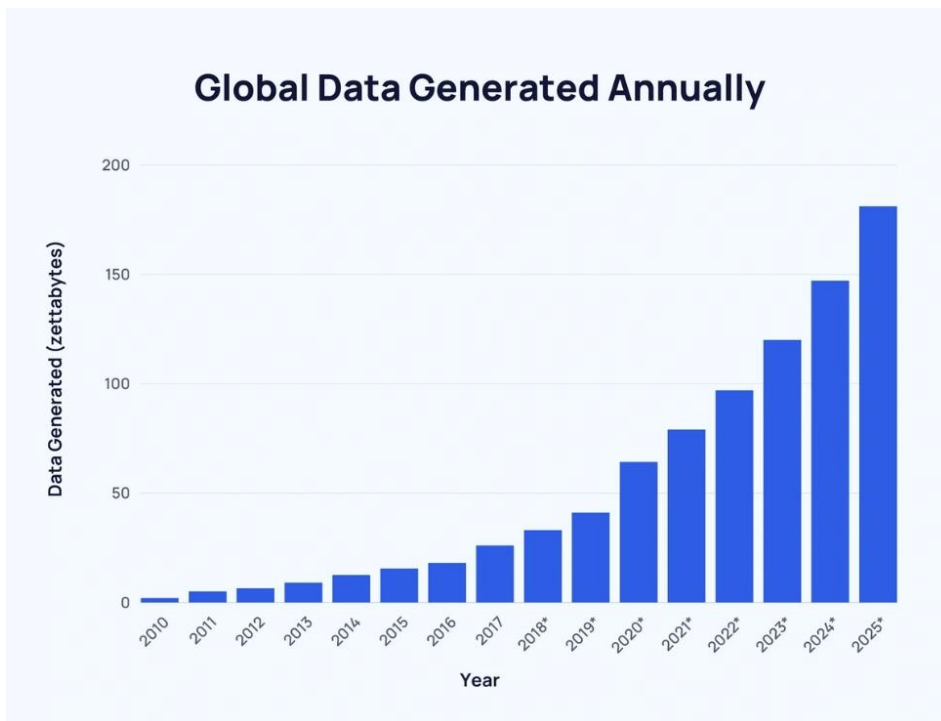
PaaS = Platform as a Service on palvelutyyppi, jossa kehittäjät voivat rakentaa, testata ja ajaa sovelluksia ilman, että heidän tarvitsee hallinnoida taustalla olevaa infrastruktuuria (Hurja 2024).

SaaS = Software as a Service on palvelumalli, jossa ohjelmistot toimitetaan ja hallinnoidaan pilvipalveluna. Asiakkaiden ei tarvitse asentaa eikä ylläpitää niitä omilla laitteillaan vaan AWS vastaa kaikesta ylläpidosta, tietoturvasta ja päivityksistä (Hurja 2024).

4 Käytetyt AWS palvelut

4.1 S3 (Simple Storage Service)

Maailmanlaajuisesti dataa tuotetaan yhä enemmän ja enemmän. Statistan tekemän tutkimuksen mukaan vuonna 2010 tuotetun datan kokonaismäärä vastasi noin 2 zettatavu, kun sen ennustetaan nousevan noin 182 zettatavuun vuoden 2025 mennessä. Yksi zettatavu vastaa noin miljardia teratavu (Singh 2025). Datan kokonaismäärän tuottaminen maailmanlaajuisesti on suurta ja nopeasti kasvavaa, tällöin datan hallitsemiseen tarvitaan luotettavia, skaalautuvia sekä kustannustehokkaita ratkaisuja. Tähän tarpeeseen Amazon S3 on luotettava valinta.



Kuva 1 Tuotetun datan määrä (Duarte 2025)

Simple Storage Service tunnetummin S3, julkaistiin Amazonin toimesta vuonna 2006. Palvelu oli ensin saatavilla Yhdysvalloissa mutta vuotta myöhemmin S3-palvelu tuli saataville myös Eurooppaan (Cloudsee 2024). S3 on objektien tallennuspalvelu, joka sisältää niin kutsuttuja ämpäreitä (Bucket), johon voidaan ladata esimerkiksi kuvia, tiedostoja eri formaateissa, videoita ynnä muuta. Tiedostoja voidaan käytännössä ladata rajattomasti S3:een, S3 konsolin kautta ladatun tiedoston

enimmäiskoko on rajattu 160 gigatavuun. AWS CLI:llä PUT komennolla tiedoston enimmäiskoko on viisi gigatavua.

Amazon S3-palvelu on saatavilla kaikilla AWS:n alueilla ympäri maailman. S3 tarjoaa 99,9999999 % kestävyden, mikä käytännössä tarkoittaa sitä, että jokaista kymmentä miljoonaa tallennettua objektia kohden, voidaan menettää yksi objekti kerran 10 000 vuodessa (Yigit Ogun 2022). S3 palvelun lähes rajaton skaalautuvuus mahdollistaa datan tallennuksen ja hakemisen ilman suorituskykykompromisseja, mikä tekee siitä hyvän alustan datajärville.

Hinnoittelu

Amazon S3 tarjoaa useita ominaisuuksia ja tallennusluokkia. Saatavilla on monia erilaisia tallennusluokkia erilaisten käyttötapauksien ja halutun suorituskyvyn mukaan. Eri tallennusluokilla voidaan optimoida objektien suorituskykyä ja saatavuutta sekä minimoida tallennuskustannuksia. (Amazon Simple Storage 2025, 2077-2078) S3 standard versiossa ensimmäiset 50 terabittiä per kuukausi maksaa 0.023 dollaria per gigabitti ja yli 500 terabittiä 0.021 dollaria per gigabitti. Hinnat vaihtelevat S3 alueen mukaan.

Datajärvet

Digitalisoitumisen alussa relaatiotietokannat olivat ensisijainen menetelmä tietojen hallintaan, jota yritykset käyttivät tietojen keräämiseen, tallentamiseen ja analysointiin. Relaatiotietokannat tarjosivat yrityksille tavan tallentaa ja analysoida tietoja SQL-kielen avulla. Tänä päivänäkin relaatiotietokannat ovat päteviä erittäin strukturoidun datan hallintaan. Datan määrä ja laajuus ovat kuitenkin kasvaneet merkittävästi, jolloin yrityksillä voi olla useita hajautettuja tietokantoja, joissa voi olla erilaisia käyttäjiä ja käyttötarkoituksia. Tämä johti tietosiiloihin eli hajautettuihin tietovarastoihin koko organisaatiossa (Amazon Web Services, n.d.).

Tietojärvet ovat syntyneet yhdistämään organisaation eri tietokannat samaan paikkaan. Datajärvet ovat keskitettyjä varastoja, joihin voidaan tallentaa strukturoitua ja strukturoimattomia tietoja missä tahansa mittakaavassa. Dataa voidaan tallentaa sellaisenaan ilman, että sitä täytyy ensin muokata. AWS:n tietojärviin voidaan ladata dataa lähes mistä tahansa tietolähteestä mukaan lukien ERP- ja toimitusketjunhallinta järjestelmät.

AWS käyttää tietojärviä suurten datamassojen varastointiin. Datajärven perustana toimii S3-talennuspalvelu (Simple Storage Service). Amazon S3 on paras paikka datajärvien rakentamiseen sen kestävyden, skaalautuvuuden ja turvallisuuden vuoksi. Datan yhdistämiseen yhtenäiseksi tietomalleiksi käytetään erilaisia koneoppimismalleja. Datajärvet mahdollistavat reaaliaikaisen näkyvyyden ja analytiikan toimitusketjussa.

4.2 Amazon Q palvelu

AWS julkaisi Amazon Q palvelunsa marraskuussa vuonna 2023. Amazon Q on generatiivisella tekoälyllä toimiva avustaja, jonka avulla käyttäjät voivat ratkaista erilaisia ongelmia, luoda sisältöä sekä saada oivalluksia ja neuvoja tehtävien nopeuttamiseksi (Barth 2023). Amazon Q auttaa käyttäjiään lisäämään kehittäjiensä tuottavuutta sekä päätöksentekoa. Amazon Q on rakennettu Amazon Bedrockille, joten käyttäjät voivat hyödyntää näitä Bedrockin tuottamia toimintoja turvallisuuden, suojaamisen sekä tekoälyn vastuullisen käytön varmistamiseksi (Amazon Q Business 2025).

Generatiivinen tekoäly

Generatiivinen tekoäly on yksi tekoälyn osa-alueista, jolla pystytään luomaan uutta sisältöä, kuten kuvia, videoita, tekstiä ja ääntä. Generatiiviset mallit ovat keskeinen osa generatiivista tekoälyä, joilla tarkoitetaan syväoppimisverkostoja, jotka oppivat luomaan uutta sisältöä tietyn aineiston perusteella (Numminen 2023).

Generatiivista tekoälyä käytetäänkin tänä päivänä paljon muun muassa kuvien, videoiden sekä musiikin ja muun taiteen tuottamisessa. Myös pelimaailmassa generatiivista tekoälyä käytetään luomaan esimerkiksi pelihahmoja, erilaisia maailmoja sekä tarinoita. (Buser 2025).

Amazon Q:sta on saatavilla kaksi erilaista versiota, Amazon Q Business sekä Amazon Q Developer. Business versio tuo generatiivisen tekoälyn turvallisesti saataville kaikille organisaation jäsenille hyödyntäen yrityksen dataa, sisältöä sekä järjestelmiä. Amazon Q Business on rakennettu ongelmien ratkaisuihin, sisällön luomiseen sekä muiden toimintojen toteuttamiseen käyttäjänsä puolesta. Developer versio sen sijaan on suunniteltu avustamaan ohjelmistojen rakentamisessa, näiden käyttöönotossa sekä muokkaamisessa. Developer voidaan myös ladata koodieditoreihin, kuten Visual studio codeen.

Amazon Q in Quicksight:n hinnoitellaan käyttäjän roolin mukaan, mikä tarjoaa joustavuutta sekä helppoutta valita oikea hinnoittelumalli. Noin kolmen dollarin kuukausimaksulla voidaan toimittaa datanäkemyksiä skaalautuvasti koko organisaatiolle. Quicksight tarjoaa Author sekä Author Pro hinnoitteluvärsiota käyttäjille. Author versio maksaa 24 dollaria kuukaudessa ja sillä voidaan käyttää Amazon Q:n Q/A ominaisuuksia dataan luonnollisella kielellä. Author Pro versio sisältää kaikki Author version ominaisuuksien lisäksi generatiiviset BI-ominaisuudet, kuten aiheiden- ja datatarinoiden luonti sekä edistyneet analyysit skenaarioiden avulla. (User Guide 2025)

Quicksight:n lukijat voivat selata koontinäyttöjä, vastaanottaa sähköpostiraportteja ja ladata näitä tietoja. Lukijoille Quicksight tarjoaa kaksi hinnoittelumallia Reader sekä Reader Pro. Reader maksaa käyttäjää kohden kolme dollaria kuukaudessa ja Reader Pro 20 dollaria kuukaudessa. Reader Pro versio tarjoaa generatiivisten datatarinoiden luomisen ja jakamisen sekä edistyneet analyysit skenaario- ominaisuuden avulla. (Amazon Quicksight 2025)

4.3 Quicksight palvelu

Datan visualisointi on tapa näyttää tietoa visuaalisina elementteinä, kuten kaavioina tai diagrammeina. Datan visualisoinnilla muutetaan monimutkaiset tai isot tietojoukot graafiseksi kuvaksi, jolloin datan tulkitseminen ja päätöksenteko on helpompaa.

Datan visualisoinnin tärkeimpiä hyötyjä ovat liiketoiminnan parantaminen, auttaa parantamaan päätöksentekoa ja havaitsemaan mahdollisia poikkeamia. Visualisoimalla isompi kuva datasta voidaan tunnistaa asiakkaiden käyttäytymismalleja, löytää erilaisia trendejä ja saada oivalluksia liiketoiminnan parantamiseksi. Datan visualisointi korostaa asiakkaiden tarpeita, sillä voidaan parantaa tuotteita tai palveluita strategisesti ja vähentää toiminnan tehottomuutta. Datan visualisointiin Quicksight-palvelussa on monia erilaisia tapoja. Siihen voidaan käyttää pylväs- ja viivakaavioita,

esimerkiksi myynnin seuraamiseen, piirakkakaavioita esittämään erilaisia osuuksia kokonaisuudesta tai lämpökarttoja (heat maps) arvojen vaihteluun.

Massadata

Yritykset tuottavat ja käsittelevät paljon massadataa erilaisista tietolähteistä, joten on tärkeää saada data kuvattua erilaisina kuvioina tai diagrammeina sen ymmärtävyyden tueksi. Massadatalta tarkoitetaan suurien ja monimutkaisten datakokonaisuuksien keräämistä, säilyttämistä ja analysointia. Big datasta puhuttaessa viitataan usein sen kolmeen V:hen (Velocity, Variety ja Volume). Volyyymi tarkoittaa tiedon määrää, Velocity viittaa tiedon käsittelyn nopeuteen ja Variety viittaa eri tietotyyppien määrään (Wigmore 2023).

AWS julkaisi Quicksight palvelunsa vuonna 2016. Quicksight oli aluksi saatavilla Yhdysvalloissa Oregonin ja Pohjois-Virginian alueilla sekä Euroopassa Irlannissa mutta tänä päivänä se on laajentunut käyttöön ympäri maailmaa (AWS 2016). AWS Quicksight on hyvin skaalautuva, palvelimeton ja koneoppimiseen perustuva Business Intelligence analytiikkapalvelu, jonka avulla massadatasta voidaan luoda erilaisia kaavioita ja kuvioita datan lukemisen helpottamiseksi.

AWS Quicksight tarjoaa datan visualisointia useista eri lähteistä, kuten paikallisista tietokannoista, AWS:n omista palveluista kuten S3 tallennuspalvelusta ja RDS tai vaihtoehtoisesti tiedostoja voidaan ladata suoraan CSV, JSON, ELF tai XLSX muodossa. Quicksightilla voidaan muodostaa yhteys myös useisiin SaaS-tietolähteisiin käyttämällä avointa valtuutusta (OAuth) tai muodostamalla näihin suora yhteys. Projektinhallintaohjelmisto Jira tukee suoraa yhteyttä Quicksightiin. Jotta voidaan muodostaa yhteys Github tai Adobe Analytics tietolähteisiin, Quicksightin on pystyttävä käyttämään tietolähdettä verkon kautta. Tällöin tarvitaan avoin valtuutus Quicksightille, jotta pystytään ottamaan yhteys näihin tietolähteisiin. Wigmore (2023) korostaa, että tällaiset työkalut ovat välttämättömiä massadatan analysoinnissa, kun taas AWS (2016) täsmentää, että Quicksight integroituu esimerkiksi S3-tallennuspalveluun, RDS-tietokantoihin ja SaaS-lähteisiin, kuten Jiraan, hyödyntäen OAuth-valtuutusta tai suoria yhteyksiä.

Turvallisuus

Jaettu vastuumalli (Shared responsibility model) tarkoittaa sitä, että AWS sekä heidän asiakkaansa ottavat yhdessä vastuun pilvipalveluiden käytön turvallisuudesta ja ylläpidosta. Tämä tarkoittaa sitä, että AWS huolehtii omien palveluidensa, kuten tallennustilan, virtuaalikoneiden, tietokantojen sekä verkon turvallisuudesta ja asiakkaat huolehtivat käyttämiensä resurssien turvallisuudesta, esimerkiksi valvomalla käyttöoikeuksia sekä pääsynhallintaa resursseihin (Patel 2023).

Yksi parhaista turvallisuuskäytänteistä Quicksightissa on tietojen salaus. Se on todella tärkeää niin yritykselle, kuin heidän asiakkailleen. Tietojen salaus voidaan jakaa kahteen eri ryhmään, kun tiedot liikkuvat paikasta toiseen sekä, kun tietoa ei siirretä aktiivisesti, esimerkiksi Quicksightissa, jossa data on tallennettuna SPICE:n välimuistiin.

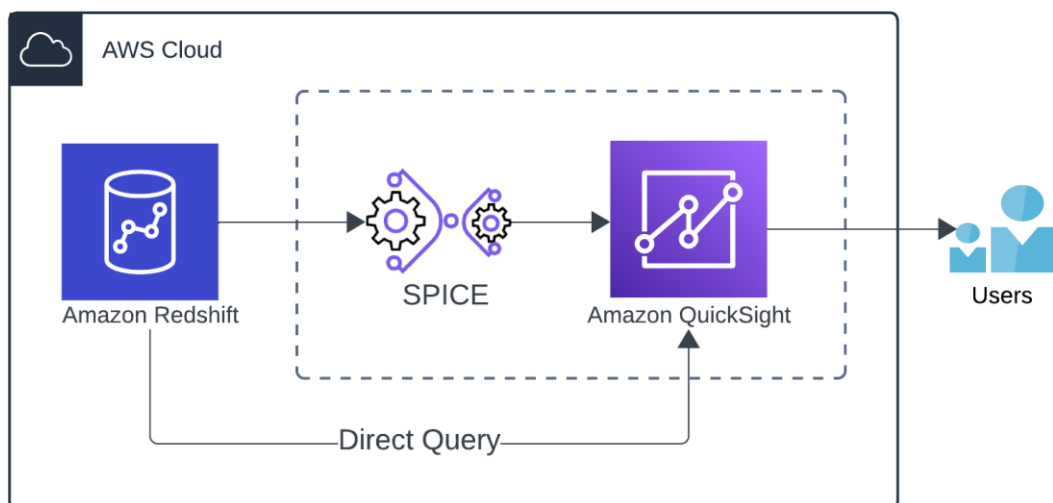
Quicksight tukee sekä rivi- että saraketason suojausta hallita tietoihin pääsyä. Saraketason suojaus on tärkeää organisaatioille, jotka käsittelevät paljon esimerkiksi henkilö- tai liiketoimintatietoja ja, kun halutaan hallita organisaation käyttäjien pääsyä tiettyihin tietojoukon sarakkeiden tietoihin käyttöoikeuksien tai käyttäjän tunnistetiedon perusteella (Vardhini 2024). Tällöin voidaan määrittää tiettyjen käyttäjien pääsy tietojoukkoihin sekä varmistaa että vain valtuutetut käyttäjät voivat tarkastella tietojoukkoja. Vardhini, 2024 esimerkin mukaan voidaan määrittää myyntitiimi näkemään vain myynteihin liittyviä tietoja, kun taas henkilöstötiimit voivat nähdä vain työntekijöihin liittyviä tietoja.

Quicksight Enterprise versio tarjoaa käyttäjilleen myös rivitason suojauksen tietojoukkoihin. Se rajoittaa pääsyä tietojoukon riveihin käyttäjien tai ryhmien tunnistetietojen perusteella. Tämä voidaan tehdä Quicksightin hallintapaneelissa tai vaihtoehtoisesti ladata datasettiin CSV-tiedosto, jossa annetut käyttöoikeudet ovat.

SPICE-analytiikkamoottori

Quicksight käyttää suorakyselytilaa tai sitten SPICE-tilaa (Kts. Kuvio 3.), joka on Quicksightin sisällä toimiva muistinvarainen laskentamoottori. Se on suunniteltu suorittamaan nopeasti monimutkaisia laskelmia ja toimittamaan tietoja joko ajoitettujen päivitysten, tarpeen mukaan tapahtuvien pyyntöjen (on-demand) tai API-pohjaisten päivityksien perusteella. SPICE laskentamoottori tallentaa tuodun datan omaan välimuistiin. SPICE siis nopeuttaa analyttisiä kyselyjä tuodulle datalle käytettävissä olevan tallennus- ja käsittelykapasiteetin myötä, eikä kuormita taustalla olevaa data-lähdettä. Dataa ei siis tarvitse hakea joka kerta, kun analyysiä tai visualisointia päivitetään. SPICE:n hinnoittelu perustuu ostettuun tallennuskapasiteettiin, mikä mahdollistaa tiedonkulutuksen useiden kojetaulujen ja käyttäjien välillä ilman lisäkustannuksia (Walavalkar 2024). SPICE-kapasiteetti on jaettu erikseen AWS-alueiden mukaan. Jokaisen AWS-tilin SPICE-kapasiteetti jaetaan kaikkien Quicksight käyttäjien kesken yhdellä AWS-alueella. SPICE-kapasiteettia on käytettävissä 20 gigatavua, jolloin ylittävistä kapasiteetista palvelu veloittaa noin 0.38 \$ per GB per kuukausi.

SPICE rajoittaa tuotuun tietojoukkoon lisättävien rivien määrää. Yhdessä tiedostossa voi olla enintään 2000 saraketta ja jokaisen sarakkeen nimessä enintään 127 merkkiä. Tietojoukkojen tuonti S3 (Simple Storage Service) palvelusta luettelo voi olla enintään 1 000 tiedostoa. Standardiversiossa käytettävissä on 25 miljoonaa riviä tai 25 gigatavua tietojoukkoa kohden, kun taas Enterprise versiossa käytettävissä on miljardi riviä tai yksi teratavu jokaista tietojoukkoa kohden (AWS 2024).



Kuva 2. SPICE-tilan ja Direct Queryn ero (Yung, 2024)

Direct Query

Quicksightissa voidaan käyttää myös suorakyselytilaa. Se mahdollistaa datalähteen kyselyt lähes reaaliajassa ja koontinäytöt päivittyvät kyselyjen mukaan reaaliajassa. Kun Direct Query tilaa käytetään, koontinäytöt ja analyysit päivittyvät aina kun ne avataan tai sitten automaattisesti 24 tunnin välein.

Toisin kuin SPICE, suorakyselytila ei aseta rajoituksia rivien määrälle tai datajoukon koolle, joten se sopii sellaisiin tilanteisiin, joissa datamäärä ylittää SPICE:n kapasiteetin sekä, kun halutaan saada reaaliaikaista dataa. Pidemmät ja monimutkaisemmat kyselyt voivat hidastaa kojetaulun päivittämistä. Toisinkuin SPICE-tilassa, Direct Querya käytettäessä kyselyt tehdään suoraan tietolähteeseen, esimerkiksi RDS tietokantaan. Kun yhteyttä muodostetaan JDBC/OBDC tuettuun tietokantaan, voidaan valita Direct Query -tila. Kun koontinäyttöön tehdään muutoksia suodattamalla tai päivittämällä kaavioita, Direct Query lähettää SQL kyselyt suoraan datalähteeseen. Datalähde käsittelee SQL-kyselyn ja päivittää koontinäytön reaaliajassa. Direct Query tilan käyttäminen on hyödyllistä etenkin silloin, kun data on sellaista, joka muuttuu. Esimerkiksi myynnin seuraaminen tai varastonhallintaan. (Yung & Ganapathy 2024)

Direct query tila käyttäminen mahdollistaa lähes reaaliaikaiset kyselyt datalähteeseen, joten kun käyttäjä avaa tai päivittää koontinäytön, Quicksight lähettää kyselyt suoraan datalähteeseen. Valittu datalähde on tällöin vastuussa kyselyjen suorittamisesta sekä tulosten palauttamisesta. Vastausaika riippuu tietolähteen nopeudesta, jolloin direct query tilan käyttäminen voi yhdistää ajantasaisista ja jatkuvasti päivittyvistä tiedoista saatavat hyödyt tehokkaasti koontinäyttöihin (Yung & Ganapathy 2024).

Amazon Q in Quicksight

Quicksightissa Amazon Q:ta voidaan käyttää muun muassa koontinäyttöjen rakentamiseen, visualisointiin ja tehdä sen avulla kyselyjä dataan luonnollisen kielen avulla.

Amazon Q:lla voidaan luoda datatarinoita, joiden avulla pystytään kertomaan yritykselle ja muille sidosryhmille tehokkaasti datasta sekä tehdä dataan perustuvia liiketoimintapäätöksiä. Data tarinoilla pystytään tekemään raakadatasta selkeä ja ytimekäs kokonaisuus, joka sisältää tekstiä, kuvia ja kaavioita. Näin kaikki pystyvät ymmärtämään dataa helposti ja tekemään sitä kautta liiketoiminnallisia johtopäätöksiä. Luonnollisella kielellä kysyttyihin kysymyksiin, Amazon Q antaa kattavia vastauksia ja tiedot selitetään yksityiskohtaisesti erilaisilla visuaaleilla ja tarinoilla. Tämä auttaa jokaista organisaation jäsentä tutkimaan ja syventämään ymmärrystään dataan. Koontinäyttöjen tekijät voivat lisätä visuaaleihin Verified-merkinnän, jolloin lukijat voivat varmistua visuaalin oikeellisuudesta. Verified merkinnät näkyvät myös ehdotetuissa kysymyksissä.

Jotta Amazon Q osaa auttaa käyttäjänsä paremmin antamaan odotettuja vastauksia, voidaan datajoukkoon tehdä muutoksia kuten antamalla synonyymeja tietyille joukoille, kuten esimerkiksi AnyCompany Bank:lle voidaan antaa synonyymi ACB. Näin Amazon Q tunnistaa myös ACB kyselyt oikein.

Interaktiivisuus

Quicksightilla luodut interaktiiviset toiminnot auttavat käyttäjiä sekä muita sidosryhmiä tutkimaan ja tarkastelemaan dataa paremmin. Koontinäytöt ovat ikään kuin sanomalehti, joita voidaan lukea mutta myös olla vuorovaikutuksissa. Käyttäjät voivat olla vuorovaikutuksessa koontinäytön kanssa ja he pystyvät suodattamaan sekä myös hakemaan erilaisia tietoja (AWS Quicksight 2025, 665-667). Lukijat voivat antaa myös palautetta ja kommentteja koontinäytöistä peukku ylös tai alas. Näin käyttäjä pystyy tarkistamaan omien visualisointien selkeyttä ja muuttaa niitä tarvittaessa. Verified merkinnällä olevat visualisoinnit ovat varmistettu tekijän osalta oikeiksi, joten koontinäytön lukijoille varmistuu visualisoinnin tietojen oikeellisuus. Koontinäytöt voidaan myös upottaa jo olemassa olevaan sovellukseen tai verkkosivuille.

Koontinäytön aikataulut on hyvä tapa kerätä tietoja muuttuvista tiedoista. Aikataulutuksen avulla voidaan ajoittaa ja jakaa raportteja PDF, CSV tai Excel muodossa käyttäjän sähköpostiin. Raportti voidaan aikatauluttaa jaettavaksi päivittäin, viikottain, kuukausittain jne. Koontinäyttöihin voidaan lisätä myös hälytyksiä, jolloin tietyn arvon ylittäessä tai muuttuessa annettujen ehtojen raja-arvoista, Quicksight lähettää hälytysviestin annettuun sähköposti osoitteeseen.

4.4 AWS Lambda

Lambda on tapahtumapohjainen palvelimeton laskentapalvelu, jolla voidaan suorittaa erilaisia koodeja ilman tarvetta palvelimien hallintaan. Palvelimeton palvelu tarkoittaa sitä, että AWS hallinnoi kaikesta taustalla olevasta infrastruktuurista, kuten palvelimista, käyttöjärjestelmän ylläpidosta, verkosta sekä kapasiteetista. Tämä on nopeaa ja kustannustehokasta, joten kehittäjät voivat keskittyä vain koodin kirjoittamiseen (Awati 2024).

AWS Lambda skaalautuu myös automaattisesti ylös tai alas kysynnän mukaan. Se lisää laskentaresursseja, kuten suorittimen ja muistin käyttöä kysynnän kasvaessa. Lambda käyttää pyyntökoh- taista hinnoittelua, mikä tarkoittaa, että käyttäjät maksavat vain Lambdan käsittelemistä pyyn- nöistä ja ajasta, jonka AWS on käyttänyt näiden pyyntöjen palvelemiseen (Slingerland 2024). Lambda koodi on myös tilaton, mikä tarkoittaa, että voidaan käynnistää instansseja ilman pitkiä käyttöönottoaikoja tai muita konfiguraatioviiveitä.

Lambda voidaan integroida muiden AWS palveluiden kanssa, kuten API Gateway, DynamoDB, S3, Step Functions, SNS ja SQS. Tässä työssä Lambda integroidaan S3-palvelun kanssa. Lambda tukee monia eri ohjelmointikieliä, kuten Node.js, Pythonia, Javaa, .NET (C#) ja Rubyä. Lisäksi AWS Lambda tarjoaa mahdollisuuden käyttää mitä tahansa ohjelmointikieltä. Tämä onnistuu joko Lambda Layers:ien tai konttikuvien avulla. Layers tarjoaa joustavan tavan hallita riippuvuuksia ja näin ollen mahdollistaa myös muiden kuin tuettujen ohjelmointikielten käytön mukautetulla suori- tusympäristöllä (AWS Lambda 2025, s. 508-512). Lambda Layer on ZIP-arkisto, joka sisältää koodi- kirjastoja, riippuvuuksia sekä konfiguraatioita, jotka pystytään liittämään Lambdaan.

Monet eri tapahtumalähteet voivat laukaista Lambda funktioita. Näitä kutsutaan Triggereiksi, ku- ten muuttamalla objektia S3-palvelussa tai päivittämällä taulukoita DynamoDB:ssä. Lambdalla voi- daan suorittaa mitä tahansa käyttäjän määrittämiä toimintoja, kuten lähettää ilmoituksia tai aloit- taa jonkin työnkulun (Rehemägi 2021).

Amazon EventBridge Scheduler on palvelimeton ajoitustyökalu, jota voidaan käyttää muun muassa Lambda funktion käynnistämiseen aikataulun mukaisesti tai jonkin tapahtuman seurauksena. EventBridgen avulla voidaan ajoittaa miljoonia tehtäviä, jotka voivat käynnistää yli 270 AWS-palvelua ja yli 600 API-toimintoa(EventBridge Scheduler s. 1-2). EventBridgellä voidaan luoda kolmenlaisia ajoitettuja sääntöjä. Säännöllisellä ajoituksella EventBridge suorittaa näitä toimintoja säännöllisin väliajoin, esimerkiksi 20 minuutin välein kun taas ajoitetun säännön suoritusajaksi ja päivämääräksi määritetään cron-lauseke. Cron-lausekkeella voidaan määritellä suoritus tehtäväksi tiettyyn aikaan tiettyinä päivinä viikossa tai kuukaudessa. EventBridgellä voidaan määritellä myös kertaluontoinen aikataulu (EventBridge Scheduler s.19-22).

5 Toteutus

5.1 Datasetin hakeminen ja tallennus

Työssä käytetty NHL draft -datajoukko haettiin Kaggle.com verkkosivuilta. Datajoukko sisältää tietoja kaikista NHL:ään varatuista pelaajista vuosilta 1968 – 2022. Työ aloitettiin toimeksiantajan valmiiksi pystyttämässä ympäristössä, jossa palveluiden käyttöoikeudet olivat valmiiksi määritetty. AWS:n käyttöliittymässä luotiin aluksi S3-bucket tulevalle datalle sen tallennusta varten.

Tallennus

NHL-draft datasetti ladattiin S3-bucketiin, joka toimi tässä työssä yksinkertaisena datajärvenä. Tietojoukko oli valmiiksi hyvin jäsenneltyä csv-muotoista dataa, mikä mahdollisti sen käytön Quick-sightissa ilman tarvetta datan esikäsittelylle.

5.2 Visualisointi

Quicksightille annettiin käyttöoikeudet S3-palveluun, jotta Quicksight pystyy lukemaan tiedostoja S3-bucketista. Datasetti vietiin Quicksightiin suoraan S3-bucketista luomalla ensin uusi datasetti Quicksightin käyttöliittymässä. Jotta yhteys S3-buckettiin saatiin, tarvittiin JSON-muotoinen manifest-tiedosto, joka sisältää tiedostojen tarkat osoitteet S3-bucketissa. Tässä tapauksessa kaikki tarvittava tieto oli yhdessä csv-muotoisessa tiedostossa, joten manifest-tiedostosta luotiin todella yksinkertainen antamalla vain S3-bucketin tarkka osoite sekä tiedoston formaatti.

```
{
  "fileLocations": [
    {
      "URIs": [
        "https://quicksightnhldraft.s3.eu-west-1.amazonaws.com/draft/nhldraft1.csv"
      ]
    }
  ],
  "globalUploadSettings": {
    "format": "CSV"
  }
}
```

Kuva 3 Manifest-tiedosto JSON

Tietojoukon tuonnin yhteydessä valittiin käytettäväksi SPICE-analytiikkamoottori, joka valittiin sen nopean suorituskyvyn takia. SPICE mahdollisti datan tallentamisen välimuistiin, mikä vähensi kyse-lyjen latausaikoja. Datasetin sarakkeet ja rivit tarkistettiin vielä Quicksight editorissa mahdollisten virheiden ja Null-arvojen osalta.

Muutokset datasettiin

Tietojoukko oli valmiiksi puhdasta mutta analysoinnin helpottamiseksi siihen lisättiin muutamia laskettuja kenttiä, kuten draft_decade ja position_Name. Draft_decade kenttä ryhmitteli varatut pelaajat vuosikymmenten mukaan, mikä mahdollisti trendien analysoinnin vuosikymmenittäin. Tämä tehtiin lisäämällä laskettu kenttä ja antamalla kentälle nimi ja lisäämällä funktio $\text{floor}(\{\text{year}\} / 10) * 10$. Position_name laskettu kenttä muutti pelipaikkojen lyhenteet ('C', 'RW') kokonaisiksi nimiksi, jotta visualisointien luettavuus olisi lukijalle helpompaa ja selkeämpää. Tämä toteutettiin lisäämällä laskettuun kenttään funktio (kts. Kuva 4).

```
Muokkaa laskettua kenttää

PositionName ✎

1 ifelse(
2   {position} = 'G', 'Goaltender',|
3   {position} = 'D', 'Defence',
4   {position} = 'RW', 'Right Wing',
5   {position} = 'LW', 'Left Wing',
6   {position} = 'C', 'Center',
7   'Unknown'
8 )
```

Kuva 4 Lasketun kentän lyhenteet

Datasetistä huomattiin myös, että sarakkeet points, goals, assists sekä plusmiinus olivat merkkijono tyyppisiä, eikä numeerisia, jolloin huomattiin että lajittelemalla goals sarake laskevaksi, Quicksight antoi ensimmäiseksi arvot 99 ja sen jälkeen vasta 984. Tämä korjattiin muuttamalla Quicksight editorissa kaikille näille sarakkeille numeeriset arvot, jotta visualisoinnit näyttivät sarakkeiden tulokset oikein.

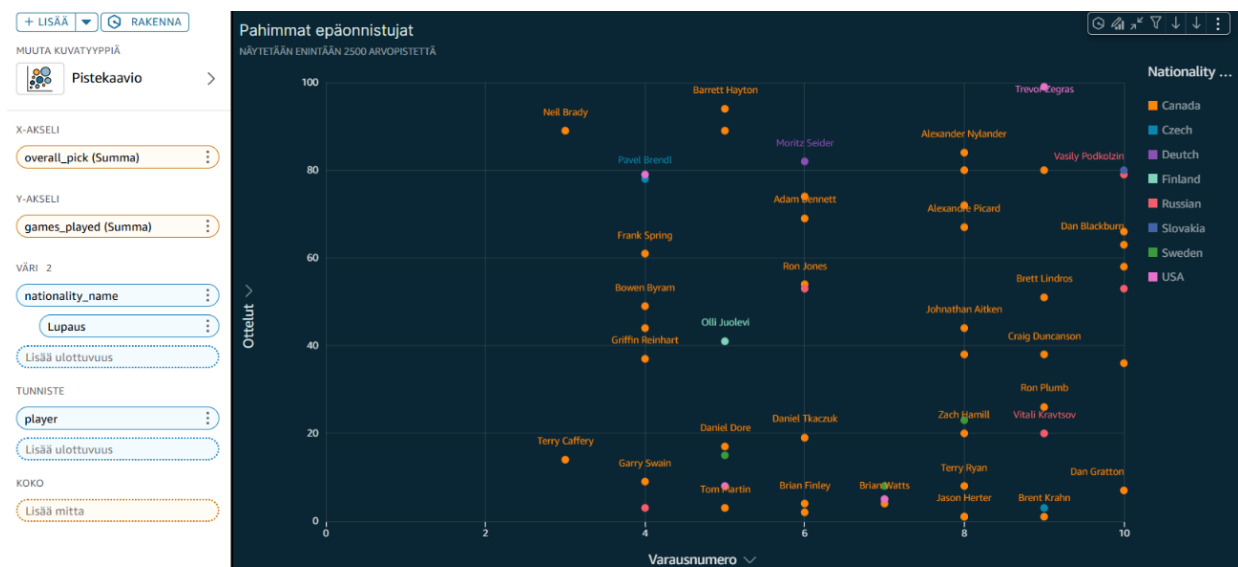
Lisäksi haluttiin kuvata odottamattomat menestyjät ja menetetyt lupaukset omissa pistekaavioissaan. Tämä toteutettiin luomalla laskettu kenttä lupaukset, joka näyttää pelaajat, joilla on pieni varausnumero (alle 9) mutta pelattuja pelejä alle 100. Pelaajat, joiden varausnumero oli yli 200 mutta otteluita pelattu yli 700, kuvattiin odottamattomiksi menestyjiksi. Haluttiin lisäksi lisätä näihin visuaaleihin kansalaisuudet värikoodattuna, jolloin koontinäytön lukijat pystyvät tarkastelemaan pelkästään haluamiaan pelaajiaan kansalaisuuksittain.

```

1 ifelse(
2   {overall_pick} > 200 AND {games_played} > 700, 'Odottamaton menestyjä',
3   {overall_pick} <= 8 AND {games_played} < 100 AND year < 2010, 'Menetetty lupaus',|
4   'Muut'
5 )

```

Kuva 5 Funktio kuvaamaan menetyä sekä lupauksia



Kuva 6 Epäonnistajat kuvattuna

Interaktiivisuus

Koontinäytöissä hyödynnettiin erilaisia interaktiivisia ominaisuuksia parantamaan koontinäyttöjen seuranta ja luettavuutta. Jokaiseen koontinäyttöön lisättiin erilaisia parametreja, jotta koontinäytön käyttäjä voi valita esimerkiksi alasvetovalikosta varausvuoden, kansalaisuuden tai joukkueen, joka sitten päivittää jokaisen visualisoinnin automaattisesti, mihin parametrit ovat liitetty.

Amazon Q hyödyntäminen

Yksi tutkimuskysymyksistä koski Amazon Q:n eli generatiivisen tekoälyn hyödyntämisestä Quicksightin koontinäyttöjen rakentamisessa ja muokkaamisessa.

Amazon Q:ta hyödynnettiin rakentamalla yksi koontinäyttö kokonaisuudessaan käyttäen pelkästään Amazon Q:ta. Amazon Q toteutti annettujen kyselyjen mukaan paikkaansapitviä kaavioita sekä kuvioita, kuten "show player, goals and points where nationality is FI". Yksinkertaiset kyselyt toimivat hyvin.

5.3 Muuttuvan datan visualisointi

Yksi tutkimuskysymyksistä koski reaali aikaisten dashboardien toimintaa ja miten tämän voisi toteuttaa käytännössä. Tässä esimerkissä käytettiin Lambda palvelua funktion suorittamiseen, jolla haettiin sportradar API:sta NHL otteluiden tuloksia, jonka jälkeen tulokset tallennettiin JSON-muodossa S3 bucketiin. Lambda ajastettiin käynnistymään päivittäin kello kuusi sekä Quicksight-palveluun lisättiin päivitysheito datasetin päivitykseen päivittäin kello 6:15. Infrastruktuuri (Kts. Kuva 7) hyödyntää AWS:n palvelimettomia palveluita Sportradar API-pyyntöjen tekemiseen, datan tallentamiseen ja visualisointiin.

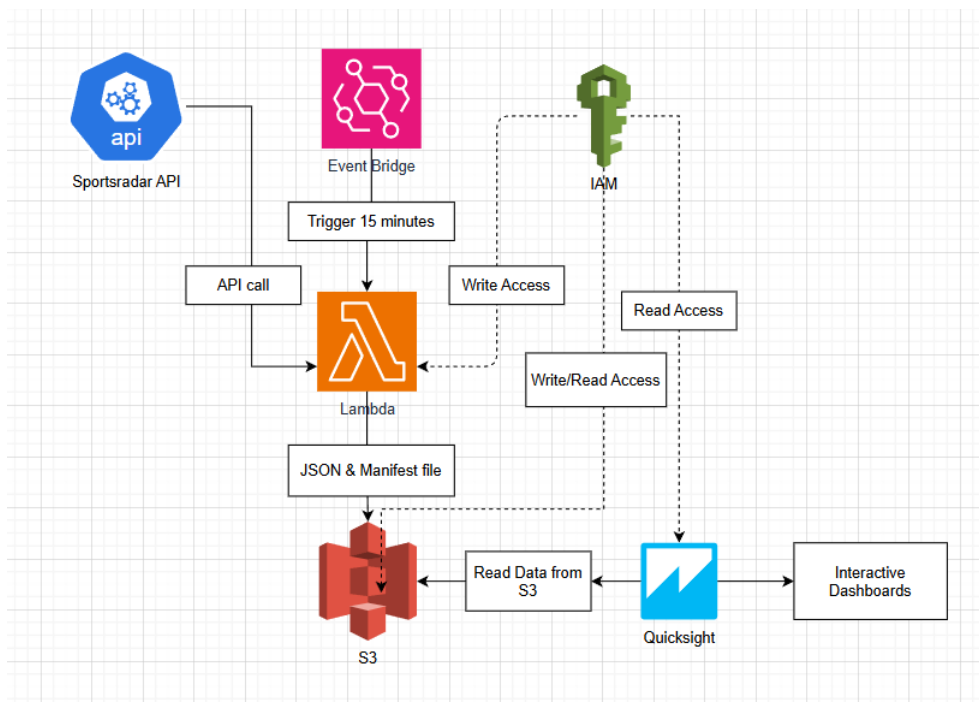
AWS Lambda: Lambda funktio tekee pyynnön Sportradar-API:lle, hakee dataa ja päivittää Quicksightin tarvitsemaan manifest-tiedostoon oikean polun. Funktio tallentaa datan S3-palveluun.

Amazon EventBridge: EventBridgeillä ajastetaan Lambda funktion suoritus päivittäin.

Amazon S3: S3 toimii tallennuspaikkana Sportradarista haetulle datalle ja manifest-tiedostolle. Data tallennetaan päivämäärän mukaiseen polkuun, josta Quicksight voi lukea uusimman sinne ladatun tiedoston.

Quicksight: Quicksight lukee S3:een ladatun manifest-tiedoston, jonka avulla se pystyy hakemaan uusimman sinne ladatun tiedoston ja visualisoida dataa.

IAM: Luku- ja kirjoitusoikeudet palveluille.



Kuva 7 Muuttuvan datan infrastruktuurikuva

Sportradar API

Sportradar verkkosivu tarjoaa moniin eri urheilulajeihin liittyvää dataa, kuten otteluiden tuloksia ja muita tilastoja. Tässä työssä käytetään NHL Apin Daily Schedule endpointia, jotta saadaan päättyneiden otteluiden tulokset. Api kutsujen tekemiseen tarvitaan API-avain, jotta voidaan kutsua Daily Schedule endpointia. Opinnäytetyötä varten tehtiin käyttäjätili, jolla saatiin ilmainen kokeiluversio ja api avain 30. päiväksi käyttöön.

Lambda ja EventBridge

Ensiksi luotiin lambda funktio, joka hakee sportradar Daily Schedule API:sta NHL otteluiden tulokset ja tallentaa ne S3-bucketiin sekä päivittää Manifest-tiedoston Quicksightia varten. Funktio ajastettiin suoritettavaksi päivittäin kello kuusi aamulla, jolloin pelien lopputulokset ovat selvillä. Tähän lambda funktioon käytettiin Grok-tekoälypalvelua auttamaan koodin kirjoituksessa.

Sportradarista saatu API avain tallennettiin lambdan ympäristömuuttujiin, jolloin API avaimen vaihtuessa ei koodia tarvitse muuttaa, vaan API avain haetaan ympäristömuuttujasta, jolloin avaimen säilytys on turvallista.

```
s3_client = boto3.client('s3')

def lambda_handler(event, context):
    try:
        # Haetaan API avain ympäristömuuttujasta
        try:
            api_key = os.environ['API_KEY']
        except KeyError:
            return {
                'statusCode': 500,
                'body': json.dumps('Error: API_KEY environment variable is missing')
            }

        # Määritetään edellinen päivä dynaamisesti
        yesterday = (datetime.datetime.now() - timedelta(days=1)).strftime('%Y/%m/%d')
        date_path = yesterday.replace("/", "-") # Esim. 2025-05-11
        schedule_url = f'https://api.sportradar.com/nhl/trial/v7/en/games/{yesterday}/schedule.json'

        # Haetaan Daily Schedule
        print(f"calling Daily Schedule API: {schedule_url}")
        params = {'api_key': api_key}
        headers = {'accept': 'application/json'}
        response = requests.get(schedule_url, headers=headers, params=params, timeout=10)
        response.raise_for_status()

        schedule_data = response.json()
        print(f"API response: {json.dumps(schedule_data, indent=2)[:500]}...")
        games = schedule_data.get('games', [])
        print(f"Found {len(games)} games in schedule")

        if not games:
            print(f"No games found for date {yesterday}")
            return {
                'statusCode': 400,
                'body': json.dumps(f'No games found for date {yesterday}')
            }
    }
```

Kuva 8 Lambda funktio

Lambda koodissa ensin haettiin API-avain Lambdan ympäristömuuttujasta, jonka jälkeen määritettiin päivämäärä dynaamisesti, jolla saadaan otteluiden edelliset tulokset. Tämän jälkeen tehdään kutsu Daily Schedule -endpointille. API vastauksen data litistetään Quicksight-analytiikkaa vasten, jonka jälkeen määritetään S3-bucketin osoite ja tallennetaan saatu data sinne. Tämän jälkeen päivitetään vielä Quicksight:n tarvitsema manifest-tiedosto, joka sisältää viittauksen uusimpaan S3-polkuun.

Lambda koodi testattiin käyttöliittymässä sekä tarkastettiin mahdollisten virheiden varalta. Tämän jälkeen tarkistettiin S3-bucketiin saapuneen datan oikeellisuus. Näiden jälkeen Lambda otettiin käyttöön.

Tämän jälkeen haluttiin automatisoida lambda funktion ajo, jottei tarvitse erikseen käyttöliittymän puolella käydä suorittamassa lambda funktiota. Lambdaan lisättiin Eventbridge triggeri ajoittamaan lambda funktion suoritus päivittäin kello 06:00, jolloin ottelut ovat pelattu sekä tulokset ovat tulleet. Tähän valittiin cron-tyyppinen ajastus cron(0 6 * * ? *), jonka rakenne tarkoittaa

vasemmalta oikealle minuutteja, tunteja, päivää kuukaudesta, kuukautta, päivää viikosta sekä vuotta. Tähti (*) merkillä tarkoitetaan että se suoritetaan joka päivä, kuukausi sekä vuosi. Kysymysmerkki (?) taas on vaatimus cron-ilmallisissa ja se varmistaa, että ajastus tapahtuu joka päivä ilman viikonpäivä rajoitusta.

Add trigger

Trigger configuration help

EventBridge (CloudWatch Events)
aws | architectures | schedules | management tools

Rule
 Pick an existing rule, or create a new one.
 Create a new rule
 Existing rules

Rule name
 Enter a name to uniquely identify your rule.

Rule description
 Provide an optional description for your rule.

Rule type
 Trigger your target based on an event pattern, or based on an automated schedule.
 Event pattern
 Schedule expression

Schedule expression
 Self-trigger your target on an automated schedule using [Cron](#) or [rate expressions](#). Cron expressions are in UTC.

e.g. rate(1 day), cron(0 17 * MON-FRI *)

Lambda will add the necessary permissions for Amazon EventBridge (CloudWatch Events) to invoke your Lambda function from this trigger. [Learn more](#) about the Lambda permissions model.

[Cancel](#) [Add](#)

Kuva 9 EventBridge ajoitus

Quicksight

Datasetin asetuksissa tietojoukkoon lisättiin päivitysaikataulu, jolloin tietojoukon tiedot päivittyvät päivittäin kello 7 aamulla, kun Lambda funktio on ajettu ja otteluiden uudet tulokset ovat tallennettuna S3 bucketiin.

Luo päivitysaikataulu ×

Aikavyöhyke

Aloitusaika

Tiheys

[LISÄÄ UUSI AIKATAULU](#) [PÄIVITÄ NYT](#)

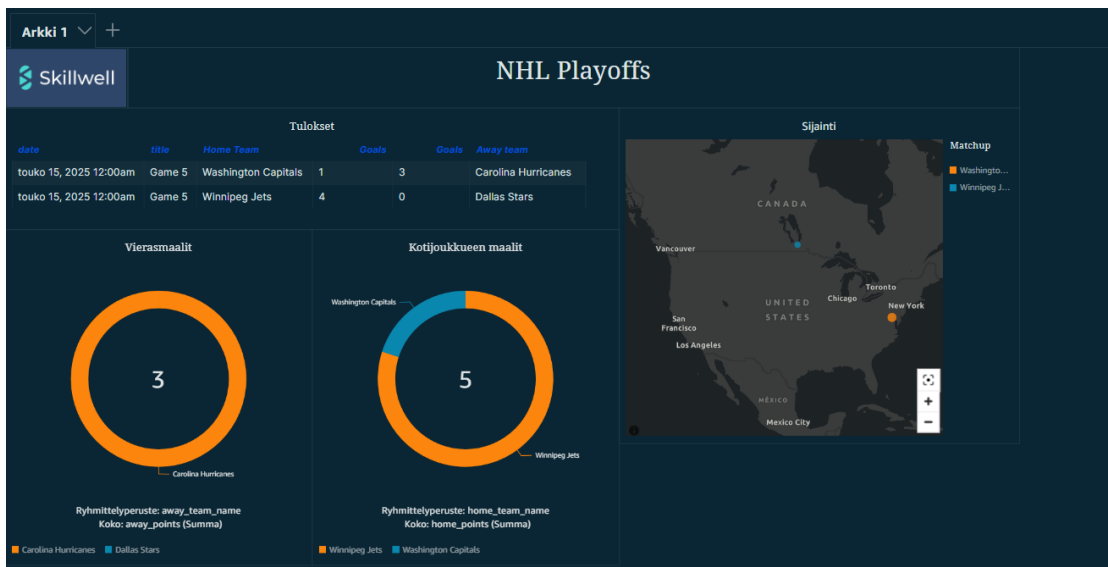
sähköpostiviesti omistajille, jos päivitys epäonnistuu

Toiminnot

seuraavan ajan kuluessa: Viimeiset 90 päivää

5.4 Valmiit koontinäytöt

Kuvassa 10 nähdään otteluparit ja pelien lopputulokset sekä kaupungit, joissa pelit ovat pelattu. Suppean datamäärän vuoksi visualisoinnit ovat hyvin yksinkertaisia mutta koontinäytössä kuitenkin pystyttiin demoamaan toimintaa, jossa tietojoukko muuttuu ja koontinäytöt päivittyvät muutoksien mukaan.

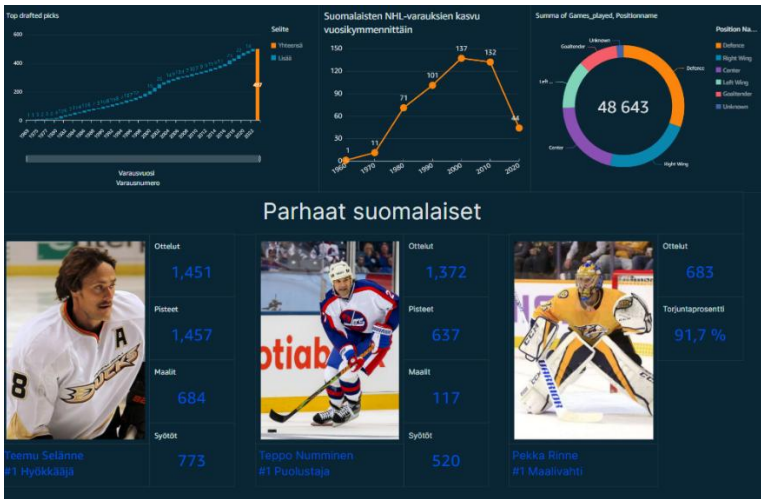


Kuva 10 Muuttuvan datan visualisointi

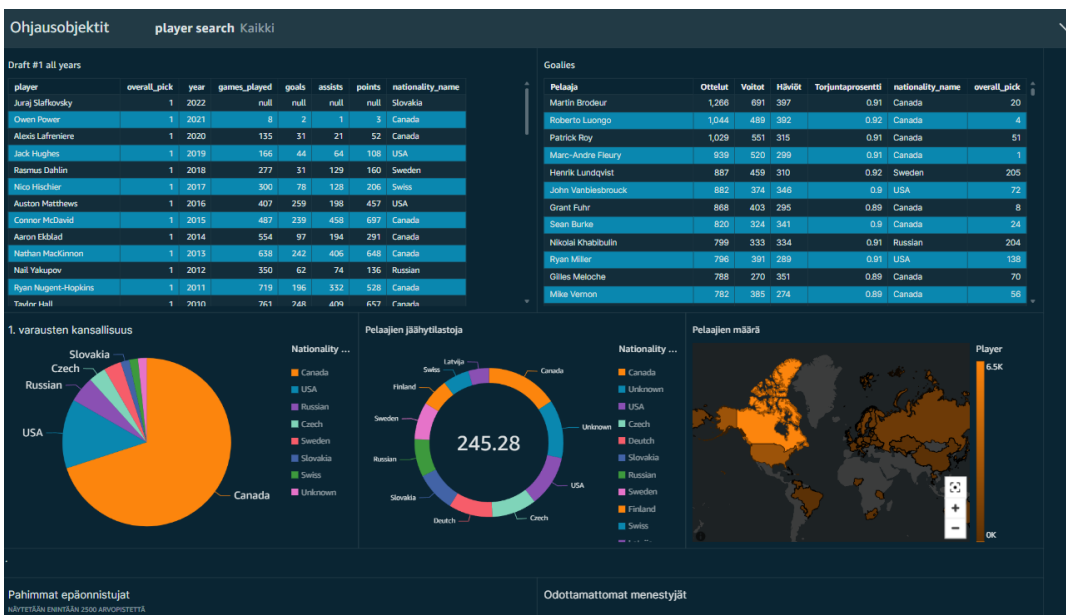
Muiden koontinäyttöjen rakentamisessa pyrittiin näyttämään visualisointimahdollisuuksia sekä muita ominaisuuksia kuten suodattimia sekä parametreja. Näiden koontinäyttöjen rakentamiseen käytettiin valmista csv-muotoista datajoukkoa, joka ladattiin Kaggle.comista. Koontinäytöissä haluttiin näyttää mahdollisimman erilaisia visualisointityyppejä kuten perinteisiä viiva- ja ympyräkaavioita sekä filled map tyyppisiä visualisointeja kuvaamaan tietyn maan dataa, kuten esimerkiksi pelaajien määriä kansalaisuuskittain. Jos käyttäjä haluaa nähdä tietoja pelkästään tietyn maan pelaajista, hän voi painaa ympyräkaaviosta kansalaisuuden ja Quicksight näyttää valitun kansalaisuuden pelaajat. Lisäksi koontinäyttöön lisättiin erilaisia hakuparametreja, joilla lukijat pystyvät suodattamaan sekä hakemaan nimen perusteella pelaajia tietojoukkojen kentistä.

Ensimmäisessä koontinäytössä näytettiin pelkästään suomalaisten pelaajien tilastoja. Koska datajoukko sisältää jokaisen varatun pelaajan vuosilta 1968 – 2022, koontinäytölle lisättiin suodatin, joka näytti pelkästään varatut suomalaiset pelaajat. Tämä toteutettiin suodattamalla kaikki muun

maan kansallisuudet pois paitsi suomalaiset ja lisäämällä suodatin koskemaan kaikkia arkin visuaaleja.



Kuva 11 Koontinäyttö 1



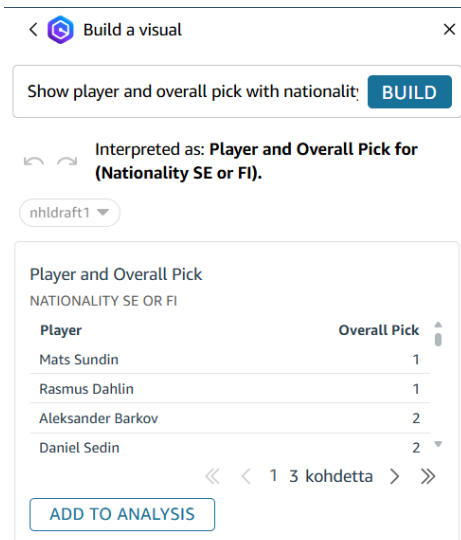
Kuva 12 Koontinäyttö 2

Koontinäytössä 2 (Kuva 11) haluttiin visualisoida koko tietojoukon tietoja ja lisätä mahdollisuus hakea pelaajien nimiä taulukosta. Tämä toteutettiin lisäämällä ensin parametri ”Search” ja antamalla tietotyyppi merkkijono. Tämän jälkeen parametriin lisättiin ohjausobjekti ja tyyliksi tekstikenttä, jolloin käyttäjä pystyy etsimään pelaajien nimiä taulukosta kirjoittamalla pelaajan nimen. Taulukoon lisättiin tämän jälkeen suodatin, johon liitettiin aikaisemmin tehty parametri ”Search”. Koontinäytöissä on pyritty käyttämään mahdollisimman monipuolisesti erilaisia kaavioita sekä kuvia datasetin ehdoilla.

6 Opinnäytetyön tulokset

Opinnäytetyön tärkeimpänä tavoitteena oli saada ymmärrystä Quicksight-palvelun hyödyistä sekä sen keskeisistä ominaisuuksista ja myös datajärvi datan yhdistämisestä Quicksightiin kahdella eri tavalla. Aluksi ladattiin yksittäinen csv-tiedosto S3-bucketiin ja lisättiin tämä tietojoukko Quicksightiin. Näin ollen pystyttiin käyttöliittymässä luomaan erilaisia kaavioita helposti. Toimeksiantajan puolesta haluttiin kuitenkin myös luoda koontinäyttö, joka päivittyy automaattisesti kun tietojoukko muuttuu. Tähän otettiin käyttöön Lambda palvelu, jolla haettiin Sportradar API:sta dataa ja se tallennettiin S3 bucketiin. Nämä tavoitteet saavutettiin onnistuneesti sekä Quicksightin ominaisuuksia, kuten suodattimia ja parametreja käytettiin koontinäytöissä monipuolisesti.

Ensimmäisenä tutkimuskysymyksenä oli ”miten Amazon Q:ta voidaan hyödyntää Quicksightissa”. Amazon Q:lla toteutettiin kokonaan yksi koontinäyttö. Antamalla luonnollisella kielellä ohjeita mitä visualisointiin haluttiin. Näiden kysymysten jälkeen ne lisättiin koontinäyttöön ja tehtiin muutoksia muunmuassa väreihin sekä visualisointityyppiin.



Kuva 13 Amazon Q kysely

Toisena tutkimuskysymyksenä oli “Miten muuttuvan datan visualisointi toimii Quicksightissa”. Tämä oli toimeksiantajan Skillwellin toive, joka ratkaistiin sillä, että etsittiin ensin sopiva verkkosivu, jossa tällaista dataa on saatavilla. Sportradar API osoittautui tähän sopivaksi, koska jo aiemmin käytetty yksittäinen CSV-tiedosto koostui jääkiekko tilastoista niin koontinäytöistä saatiin yhdenmukaisia käyttäen eri menetelmiä. Pyrittiin siis näyttämään Quicksightin koontinäytöissä aina uusimpien otteluiden tulokset, jotka päivittyivät otteluiden päätyttyä. Tässä onnistuttiin ja tuloksina saatiin visuaalinen koontinäyttö, joka päivittyy otteluiden mukaan.

7 Pohdinta

Muuttuvan datan visualisointi osoittautui osittain hankalaksi, koska tietojoukko mikä sportradarin kautta haettiin oli todella suppea eikä näin ollen pystytty kovin paljon luomaan erilaisia visuaaleja. Myös rajoitettujen api kutsujen takia ei pystytty visualisoimaan kovin monipuolisesti dataa, koska ensin täytyi kutsua daily schedule -apia, jotta saadaan otteluiden ID:t, jonka jälkeen olisi pitänyt kutsua Boxscore- apia, jossa olisi ollut yksityiskohtaisempia tilastoja peleistä. Näin ollen päädyttiin näyttämään vain pelien lopputulokset. Huomattiin myös että datan rakenteen merkitys oli tärkeää, sillä Quicksight ei pystynyt tunnistamaan sisäkkäisiä JSON-objekteja, jotka Sportradarin

API:sta haettiin. Sisäkkäiset JSON-objektit litistettiin Lambda-funktiossa, jolloin Quicksight näytti datan oikein.

Amazon Q:n avulla rakentamat visuaalit toteutettiin luonnollisella kielellä kysyen esimerkiksi pelaajat, joiden kansallisuus on Suomi tai Ruotsi ja näyttämällä pelatut ottelut sekä maalit ja syötöpisteet. Tämä onnistui hyvin ja Amazon Q osasi toteuttaa tämänkaltaiset yksinkertaiset pyynnöt hyvin.

Monimutkaisemmat kyselyt johti usein siihen ettei Q pystynyt tuottamaan oikeanlaista visuaalia tai muokkaamaan olemassa olevan taulukon tai kaavioiden solujen värejä. Kyselyiden tekeminen oli siis hyvin tarkkaa ja piti tunnistaa tietojoukon sarakkeet entuudestaan, jotta pystytään kysyä oikeanlaiset kysymykset Amazon Q:lta. Koska Amazon Q on hyvin riippuvainen tietojoukon sisällöstä voi olla vaikeaa rakentaa toimivia ja visuaalisia kuvakkeita, jos data jota käytetään on hyvin suppeaa. Esimerkkinä jos halusi muuttaa kaavion sarakkeen joksikin toiseksi, niin Amazon Q muutti vain sarakkeen nimen, eikä itse saraketta vaihtanut haluttuun.

Kestävän kehityksen kannalta tämän opinnäytetyön serverless-arkkitehtuuri palveluineen ovat hyvin skaalautuvia, eivätkä palvelut näin ollen kuluta virtaa ollessaan jatkuvasti päällä. Tämä on ekologisesti sekä taloudellisesti kestävä ratkaisu yrityksille.

7.1 Eettisyys ja luotettavuus

Työn luotettavuutta on pyritty arvioimaan koko opinnäytetyön aikana. Eettisyyden ja luotettavuuden kannalta opinnäytetyössä on pyritty noudattamaan Tutkimuseettisen neuvottelukunnan (TENK) ohjeistusta hyvästä tieteellisestä käytännöstä (HTK), jotka ovat luotettavuus, rehellisyys, arvostus sekä vastuunkanto (TENK 2023). Työssä käytetyt lähteet ovat pääosin Amazon Web Services:n kehitysdokumentaatioita sekä artikkeleita. Lähteinä on käytetty myös joitakin blogeja mutta näiden kirjoittajat ovat usein olleet AWS-sertifioituja ammattilaisia. Lähteitä valittaessa on vertailtu useita eri lähteitä ja pyritty valitsemaan ajankohtaisia ja uusimpia lähteitä tietoperustaan, sillä tiedot ja palveluiden ominaisuudet päivittyvät ja kehittyvät jatkuvasti. Grok tekoälyä hyödynnettiin Lambdakoodin toteutuksessa.

Tutkimus ei sisältänyt henkilötietoja tai muun arkaluontoisen materiaalien käsittelyä, joten salassapitosopimus tai tutkimuslupa ei ollut tarpeellinen.

7.2 Kehittämisehdotukset

Reaaliaikaisen analysoinnin ja visualisoinnin voisi toteuttaa esimerkiksi hyödyntämällä DynamoDB tietokantaa sekä kinesis data stream palvelua, jolloin tiedot päivittyisivät lähes reaaliaikaisesti ja Quicksightin koontinäytöt myös. Tämä mahdollistaisi myös datatarinoiden hyödyntämisen otteluiden aikana muun muassa luomalla ottelun aikaisia trendejä.

Useiden API-pyyntöjen tekeminen mahdollistaisi paremman ja monipuolisemman datan visualisoinnin. Kun datan määrä olisi tarpeeksi monipuolista ja kattavaa, Amazon Q:ta voisi hyödyntää näiden tietojen kyselyihin paremmin. Monipuolinen data mahdollistaisi kiinnostavia datatarinoita sekä se parantaisi Amazon Q:n kykyä tuottaa tarkempia ja parempia vastauksia kyselyihin.

Lähteet

Amazon Q Business - User Guide. 2025. S. 1–3. Viitattu 18.4.2025. <https://docs.aws.amazon.com/pdfs/amazonq/latest/qbusiness-ug/qbusiness-ug.pdf#page=2792&zoom=100,48,234>

Amazon Quicksight – User Guide. 2025. Amazon Quicksight Pricing. S. 665–667. Viitattu 18.4.2025.
Business Intelligence Service – Amazon QuickSight Pricing – AWS

Amazon Simple Storage – User Guide. 2025. Amazon Simple Storage Service. S. 2077–2078. Viitattu 18.5.2025. <https://docs.aws.amazon.com/pdfs/AmazonS3/latest/userguide/s3-userguide.pdf>

Amazon Web Services. (n.d.). What is a data lake? Viitattu 6.6.2025. <https://aws.amazon.com/what-is/data-lake/>

Awati, R. 2024. What is AWS Lambda. Viitattu 17.5.2025. <https://www.techtarget.com/searchaws/definition/AWS-Lambda-Amazon-Web-Services-Lambda>

AWS Lambda Developer Guide. 2025. S. 508–512. Viitattu 16.5.2025. <https://docs.aws.amazon.com/pdfs/lambda/latest/dg/lambda-dg.pdf#page=1150&zoom=100,48,716>

AWS Whitepaper. 2024. Overview of Amazon Web Services. Viitattu 18.2.2025. <https://docs.aws.amazon.com/whitepapers/latest/aws-overview/introduction.html>

Buser, J. 2025. Living games are here. How Gen AI is leveling up the games industry. Viitattu 6.6.2025. <https://blog.google/products/google-cloud/generative-ai-video-games/>

Cloudsee. 2024. The history of Amazon S3. Viitattu 10.4.2025. <https://www.cloudseedrive.com/history-of-amazon-s3/>

Developer Guide. 2025. S. 508–512. Viitattu 18.5.2025. <https://docs.aws.amazon.com/pdfs/lambda/latest/dg/lambda-dg.pdf#runtimes-walkthrough>

EventBridge Scheduler – User Guide. 2025. S. 1–2, 19–22. Viitattu 15.5.2025.

<https://docs.aws.amazon.com/pdfs/scheduler/latest/UserGuide/scheduler-user-guide.pdf.pdf#page=24&zoom=100,48,62>

HGinsight. 2025. AWS Market Share 2025. Viitattu 24.2.2025. <https://hginsights.com/blog/aws-market-report-buyer-landscape>

Hurja. 2024. Pilvipalveluiden palvelumallit: PaaS, IaaS, BaaS ja SaaS. Viitattu 24.2.2025.

<https://www.hurja.fi/blogi/pilvipalveluiden-palvelumallit-paas-iaas-baas-ja-saas/>

Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. 2023. Viitattu 15.5.2025. https://tenk.fi/sites/default/files/2023-03/HTK-ohje_2023.pdf

Kirvan, P. 2024. What is AWS. Viitattu 25.2.2025. <https://www.techtarget.com/searchaws/definition/Amazon-Web-Services>

Numminen, L. 2023. Mikä on generatiivinen tekoäly?. Viitattu 18.4.2025. <https://www.finnishup.com/mika-on-generatiivinen-ai/>

Op, M. 2022. NHL Draft player analysis (1968–2022). Viitattu 15.5.2025. <https://www.kaggle.com/code/mattop/nhl-draft-player-analysis-1963-2022>

Rajas, W. 2024. Best practises with Amazon Quicksight SPICE dataset and Direct Query Mode. Viitattu 15.4.2025. <https://rajaswalavalkar.medium.com/master-tips-for-amazon-quicksight-spice-dataset-direct-query-mode-to-boost-your-analytics-game-d7dea5ba51de>

Reno, N. 2024. Cloud Market Growth Stays Strong in Q2 While Amazon, Google and Oracle Nudge Higher. Viitattu 24.2.2025. <https://www.srgresearch.com/articles/cloud-market-growth-stays-strong-in-q2-while-amazon-google-and-oracle-nudge-higher>

Singh, V. 2025. What is Amazon S3. Viitattu 19.4.2025. https://www.datacamp.com/blog/what-is-s3?dc_referrer=https%3A%2F%2Fwww.google.com%2F

Slingerland, C. 2024. AWS Lambda Demystified: How it works, pros and cons and pricing. Viitattu 17.5.2025. <https://www.cloudzero.com/blog/aws-lambda/>

Sportradar. 2025. Viitattu 15.5.2025. <https://developer.sportradar.com/ice-hockey/reference/nhl-daily-schedule>

Tuomi, J. & Sarajärvi, A. 2018. Laadullinen tutkimus ja sisällönanalyysi. Uudistettu laitos. Helsinki: Kustannusosakeyhtiö Tammi. Viitattu 18.5.2025.

Yigit Ogun, A. 2022. How durable is AWS S3?. Viitattu 19.4.2025. <https://medium.com/@ayogun/how-much-durable-aws-s3-is-a3b2ce789c29>

Yung, R. & Ganapathy, K. 2024. Best practises for Amazon Quicksight SPICE and direct query mode. Viitattu 14.4.2025. <https://aws.amazon.com/blogs/business-intelligence/best-practices-for-amazon-quicksight-spice-and-direct-query-mode/>