

Juho Vikman

TIETOKUUTION HYÖDYNTÄMINEN OAMKIN RAPORTOINNISSA

TIETOKUUTION HYÖDYNTÄMINEN OAMKIN RAPORTOINNISSA

Juho Vikman
Opinnäytetyö
Kevät 2023
Tietojenkäsittelyn tutkinto-ohjelma
Oulun ammattikorkeakoulu

TIIVISTELMÄ

Oulun ammattikorkeakoulu
Tietojenkäsittelyn tutkinto-ohjelma

Tekijä(t): Juho Vikman

Opinnäytetyön nimi: Tietokuution hyödyntäminen Oamkin raportoinnissa

Työn ohjaaja(t): Minna Kamula

Työn valmistumislukukausi ja -vuosi: Kevät 2023

Sivumäärä: 43

Oulun yliopisto on siirtynyt raportoinneissaan käyttämään tietokuutioita ja Oulun ammattikorkeakoululla on tavoitteena siirtyä myös osittain käyttämään raportoinneissa tietokuutioita, jotka mahdollistavat tietojen monipuolisemman tarkastelun.

Tavoitteena on tutkia, mitä hyötyjä on tietokuutiorakenteen hyödyntämisessä Oulun ammattikorkeakoulun sisäisessä raportoinnissa. Tietokuutiota käytetään OLAP (Online Analytical Processing) -tekniikassa ja se on tietorakenne, jonka nimi kertoo sen muodostamasta rakenteesta.

Ajallisten rajoitteiden vuoksi, tietokuutiota ei päästy hyödyntämään raportin tekemisessä. Toiminnallisessa osuudessa luodaan Power BI -raportti hyödyntäen relaatiotietokantoja, sekä REST/JSON-rajapinnan kautta tulevaa dataa.

Asiakirjassa käydään läpi myös raporttien luomisessa tärkeimmät sovellukset ja tarvittavat teknikat, sekä laajempaa tarkastelua Power BI -raporttien luomisesta ja valmiina olevasta sisällöstä.

ABSTRACT

Oulu University of Applied Sciences
Degree Programme in Business Information Technology

Author(s): Juho Vikman

Title of thesis: Utilization of the data cube in Oulu University of Applied Sciences' reporting

Supervisor(s): Minna Kamula

Term and year when the thesis was submitted: Spring 2023

Number of pages: 43

The University of Oulu has switched to using data cubes in its reporting, and Oulu University of Applied Sciences aims to partially switch to using data cubes in its reporting, which enable a more versatile examination of the data.

The goal is to investigate the benefits of utilizing the data cube structure in Oulu University of Applied Sciences' internal reporting. Data cube used in the OLAP (Online Analytical Process) technology and is a data structure whose name indicates the structure it forms.

Due to time constraints, the data cube could not be utilized in making the report. In the operational part, a Power BI report is created using relational databases and data coming through the REST/JSON interface.

The document also reviews the most important applications and necessary techniques for creating reports, as well as a broader review of the creation of Power BI reports and the ready-made content.

SISÄLLYS

1	JOHDANTO	6
2	DATAN HAKEMISEEN KÄYTETTÄVÄT TEKNIIKAT	10
2.1	On-Line Analytical Processing.....	10
2.2	Relaatiotietokanta, SQL-kysely	12
3	RAPORTIT	14
3.1	Talousraportointi.....	14
3.2	Opiskelijatietoihin perustuva raportointi	14
3.2.1	Yksikkökohtaiset raportit	15
3.2.2	Muut raportit.....	20
4	RAPORTOINNISSA KÄYTETTÄVÄT TYÖKALUT	21
4.1	Microsoft Power BI	21
4.1.1	Power BI Service	21
4.1.2	Power BI Desktop	22
4.1.3	Käyttöoikeustyypit.....	22
4.2	Microsoft On-premises data gateway	25
4.3	Remote desktop connection	26
4.4	Tabular Editor 2.....	27
4.5	Bitwise SSH Server/Client.....	27
5	POWER BI -RAPORTTIEN LUOMINEN.....	29
5.1	Tietojen tuominen ja muokkaaminen relaatiotietokannasta	29
5.2	Tietojen tuominen ja muokkaaminen rajapinnan kautta.....	31
5.3	Tietomallin rakentaminen	34
5.4	Raportin ulkoasu	36
5.5	Laskentakaavojen määrittäminen.....	37
5.6	Testaaminen ja optimointi.....	39
5.7	Julkaiseminen.....	39
6	JOHTOPÄÄTÖKSET	41
	LÄHTEET.....	42

1 JOHDANTO

Opinnäytetyö tehdään Oulun ammattikorkeakoulun Data-analytiikka ja koulutuksen järjestelmät - tiimin pyynnöstä, sekä myös siellä työskennellessä oman työnkuvaan kuuluvien tehtävien tutkimista varten. Raportoinnissa tarvittavaa dataa on haettu tällä hetkellä MariaDB-tietokannasta siellä olevien koontitaulujen avulla. Datan hakemista on kuitenkin suunniteltu, että se haettaisiin ainakin joltain osin tietokuutioita hyödyntäen. Tietokuutioiden käyttö on uutta myös tiimillemme, joten tavoitteena on jakaa tutkimustietoa eteenpäin.

Opinnäytetyössä tutkitaan yleisesti relaatiotietokantojen ja tietokuutioiden eroavaisuuksia, niiden hyviä ja huonoja puolia ja sitä, kuinka paljon resursseja ajallisesti tietokuutioihin siirtyminen veisi. Palvelimet ovat Oulun yliopiston puolella ja olemmekin osaltaan riippuvaisia heidän resursseihinsa, koska meillä ei ole valmiuksia tehdä kaikkia tarvittavia toimenpiteitä tietokuutioiden käyttöönottoon.

Alkuperäisenä tavoitteena oli luoda Power BI -raportti, jossa hyödynnettäisiin tietokuutioita. Ajallisten rajoitteiden takia, tietokuutioiden käyttöönotto viivästyi, joten Power BI -raportin luomisessa on käytetty relaatiotietokantoja, sekä REST/JSON-rajapinnan kautta tullutta dataa. Luotu raportti on raporttipohja, joka pohjautuu Oulun ammattikorkeakoulun uuteen organisaatiomalliin. Raporttipohjaa pystytään muokkaamaan suodattimilla niin, että lukuja pystytään suodattamaan organisaatiomallin mukaisiin osaamisaloihin.

Tutkimusmetodeina käytetään tietoa Oulun Yliopiston puolelta tulevilta henkilöiltä, jotka ovat työskennelleet asian yhteydessä. Yliopisto on jo rakentanut taulukkoeditoria käyttämällä Oamkille mallin, joka antaa osittain ymmärryksen tietokuutioiden logiikasta. Saan myös heidän asiantuntijoiltaan tietoa tietokuutioiden toimintalogiikasta. Tietokuutioiden tarkastelua varten, käytän etätö-pöytäyhteyttä Oulun yliopiston ulkopuolisen käyttäjän tunnuksilla Tabular Editoria.

Käsitteet

Data

Tietojärjestelmiin tallennettua koneluettavaa ”raakatietoa”. Dataa rakenteistamalla saadaan sitä tietoa.

Rakenteeton data

Koneluettavaa tietoa, joka on metatiedotonta, mutta sen informaatiosta tulkitsemalla voidaan jalostaa tietoa.

Rakenteellinen data

Koneluettavaa tietoa, jolla on metatiedon mukainen semanttinen merkitys.

Relaatiotietokanta

Koostuu taulukoista, jossa tiedot ovat riveillä ja sarakkeissa. Tieto on käytettävissä heti tietokantaan lisäämisen jälkeen ja tauluja voi yhdistellä keskenään.

Tähtimalli (Star schema, dimensional modelling)

Dimensiomalli. Nimi tulee sen mallista, jossa keskellä on faktataulu ja ympärillä useita pienempiä dimensiotauluja, jotka yhdistyvät keskelle.

Faktataulu

Tähtimallissa käytetty termi. Dimensiotaulut yhdistyvät tähtimallin keskellä olevaan faktatauluun, joka on kooltaan isompi kuin dimensiotaulu. Tyypillinen faktataulu voisi esimerkiksi olla myynti.

Dimensiotaulu

Tähtimallissa käytetty termi. Kyselyä tehdessä dimensiotaulut yhdistetään faktatauluun. Dimensiotaulut ovat faktataulua pienempiä. Tyypillinen dimensiotaulu voisi olla esimerkiksi tuote.

OLAP (Online Analytical Processing)

Tekniikka, jolla järjestetään suuria tietokantoja yhteen tai useisiin kuutioihin. Tiedot noudetaan ja analysoidaan niin, että Pivot-tilukkoraportteja on helpompi luoda ja käyttää.

Tietokuutio

OLAP-tekniikassa käytetty termi. Tietorakenne, joka koostaa mitat analysoitavien dimensioiden tasojen ja hierarkioiden mukaan. Kuutiot eivät ole ”kuutioita” täysin matemaattisessa mielessä, koska niillä ei välttämättä ole yhtä paljon puolia. Ne ovat kuitenkin soveltuva metafora monimutkaiselle käsitteelle. (Microsoft Support 2021).

Power BI

Microsoftin Power BI -raportointi- ja analysointipalvelun avulla dataa voi visualisoida ja yhdistellä yritysten omista tietokannoista ja järjestelmistä, tiedostoista, intranetissä, netistä, erilaisista pilvipalveluista tai esimerkiksi avoimen datan palveluista. (Enho 2020).

Tunneling (Tunnelointi)

Verkkoyhteyksissä tunnelit ovat menetelmä tiedon siirtämiseksi verkon yli käyttämällä protokollia, joita kyseinen verkko ei tue. Tunnelointi toimii ”kapseloimalla” paketteja, eli paketteja kääritään muiden pakettien sisään. Paketit ovat pieniä tietokappaleita, jotka voidaan koota uudelleen niiden määränpäässä suuremmaksi tiedostoksi.

Tunnelointia käytetään usein virtuaalisissa yksityisverkoissa (VPN). Sillä luodaan turvallisia yhteyksiä verkkojen välille, se mahdollistaa ei-tuettujen verkkoprotokollien käytön ja joissakin tapauksissa antaa käyttäjien ohittaa palomuurit. (Cloudflare 2023).

The Open Systems Interconnection (OSI)

OSI-malli kuvaa seitsemää kerrosta, joita tietokonejärjestelmät käyttävät kommunikoidaan verkon yli (kuva 1). Se on ensimmäinen verkkoviestinnän standardimalli, jonka kaikki suuret tietokone- ja tietoliikenneyritykset omaksuivat. Kuvassa 1 OSI:n 7-kerroksinen malli on kuitenkin edelleen laajalti käytössä, koska se auttaa visualisoimaan ja viestimään verkkojen toiminnasta sekä eristämään ja ratkaisemaan verkko-ongelmia. (Impreva 2023).

7	Application Layer	Human-computer interaction layer, where applications can access the network services
6	Presentation Layer	Ensures that data is in a usable format and is where data encryption occurs
5	Session Layer	Maintains connections and is responsible for controlling ports and sessions
4	Transport Layer	Transmits data using transmission protocols including TCP and UDP
3	Network Layer	Decides which physical path the data will take
2	Data Link Layer	Defines the format of data on the network
1	Physical Layer	Transmits raw bit stream over the physical medium

Kuva 1 OSI-mallin seitsemän kerrosta (Oracle Netsuite 2022)

Solmu (node)

Tietorakenteen osa, johon on sisällytetty tietoa. Siihen sisältyy myös linkki muihin solmuihin, ja näin ollen niistä muodostuu lopulta jonkinlainen verkko.

Secure Shell (SSH) -protokolla

SSH määrittää salatut yhteydet käyttäjän (client) ja palvelimen välille ja sitä voidaan käyttää myös suojatun tunnelin luomiseen. SSH toimii OSI-mallin kerroksessa 7, eli sovelluskerroksessa (Cloudflare 2023).

UFO-tunnus

Oulun yliopistossa käytetään organisaation ulkopuolisen käyttäjän, joiden tietoja ei ole henkilöstöhallinnon järjestelmissä, mutta ovat oikeutettuja käyttämään yliopiston palveluja käyttöoikeuksien mukaan.

Mittari (Measure)

Mittari on Power BI:ssä oleva määritelty laskutoimitus, jotka suoritetaan kyselyn aikana, mutta ei tallenneta tietokantaan. Tämä määritetään DAX-lausekkeella tarvittavia funktioita käyttäen.

.bim -tiedostomuoto

Analysis Servicesin ja Power BI:n tietomallien tallennusmuoto.

2 DATAN HAKEMISEEN KÄYTETTÄVÄT TEKNIIKAT

Datan hakemisessa käytettäviä tekniikoita on erittäin paljon. Nämä tekniikat kuitenkin rajautuvat siihen, miten tietorakenne on rakennettu. Oamkin opiskelijadata on tällä hetkellä MariaDB-tietokannoissa Oulun Yliopiston hallinnoimilla palvelimilla. Aikaisemmin myös talousdata oli MariaDB-tietokannassa ja tietoa niistä haettiin koontitaulujen avulla, mutta taloushallintojärjestelmän muututtua, data tuotaisiin tietokuutioihin. Tässä osiossa keskitytään tällä hetkellä käytössä olevaan relaatio-tietokantoihin ja SQL-kyselyihin, sekä mahdollisesti tulevaisuudessa käytettyyn OLAP-tekniikkaan.

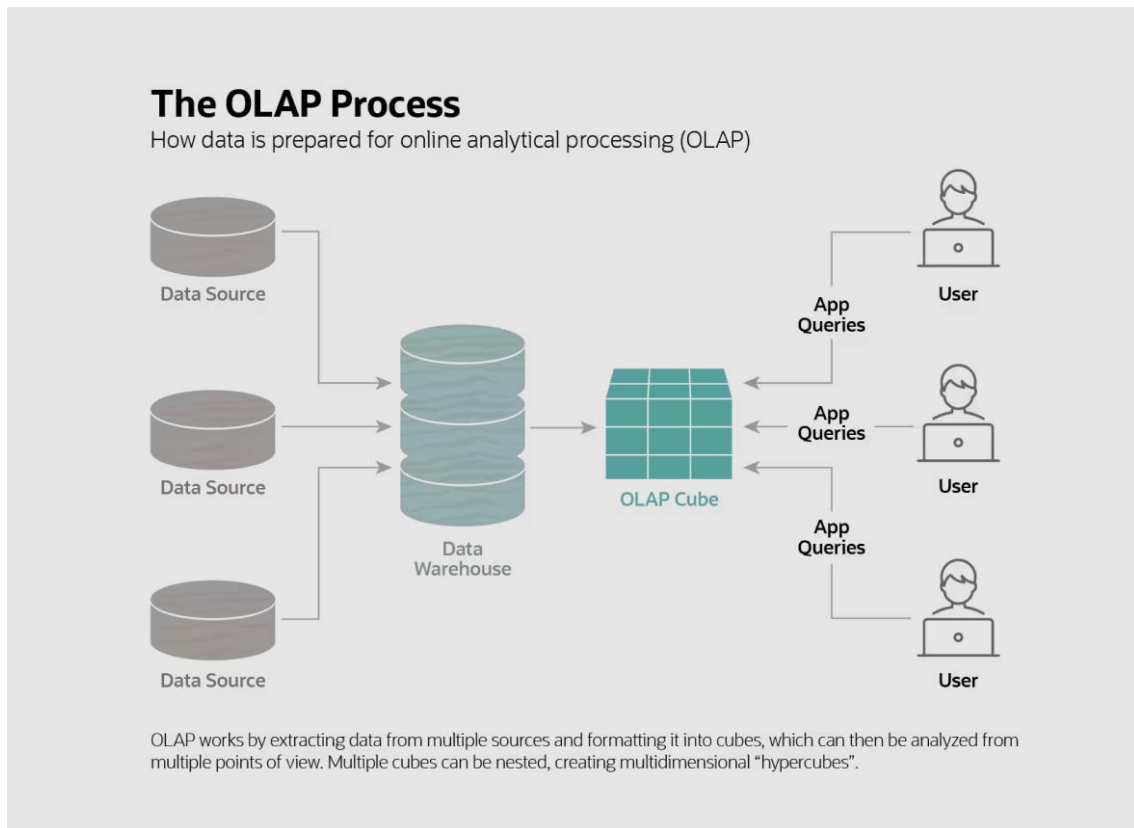
Tietokuutioissa, sekä relaatiotietokannoissa on molemmissa hyviä, sekä huonoja puolia. Erityisesti tietokuutioiden käyttöönotto on suurehko prosessi ja vaatii erilaisten tekniikoiden erikoisosaamista. Tämä vaatisi tässä tapauksessa myös aikaa, sekä resursseja tietokantojen kanssa työskentelevien henkilöiden perehdyttämiseen, sekä koulutukseen. Käytönoton kannattavuus näkyisi näin ollen kuutioiden luomisprosessissa vasta myöhemmin ja silloin, kun dataa on eri lähteistä paljon.

Relaatiotietokantojen hyöty näkyy tässä tapauksessa jo tietokantojen kanssa työskentelevien henkilöiden olemassa olevasta tiedosta ja osaamisesta. Varsinaisia ongelmia ei ole ollut relaatiotietokantojen kautta tulevista kyselyistä ja se on toiminut tehokkaasti. Relatiotietokannoista tulevien kyselyiden vaihtaminen tietokuutioista tuleviin kyselyihin onkin tullut tiimin ulkopuolelta.

2.1 On-Line Analytical Processing

On-Line Analytical Processing (OLAP) on menetelmä, jossa tiedot järjestetään luokkiin ja ennalta laskettuihin yhteenvetoarvoihin analyysia varten (kuva 2). OLAP-lähdetiedot ovat OLTP (Online Transactional Processing) -tietokantoja, jotka tallennetaan yleisesti tietovarastoihin. OLAP-tiedot perustuvat historiatietoihin ja koostettu rakenteisiin, jotka sallivat kehittyneen analyysin. OLAP-tiedot on myös järjestetty hierarkkisesti ja tallennetaan kuutioihin taulukoiden sijaan. Se on kehittynyt tekniikka, joka käyttää moniulotteisia rakenteita tietojen nopeaan käyttöön analysointia varten. Tämän organisaation avulla pivot-tilukkoraportin tai PivotChart-raportin avulla on helppo näyttää korkeatasoisia yhteenvetoja, kuten kokonaismyynnin loppusummat koko maassa tai koko alueella,

ja näyttää myös niiden sivustojen tiedot, joissa myynti on erityisen vahvaa tai heikkoa. (Microsoft Support 2021).



Kuva 2 OLAP-prosessissa poimitaan tietoja eri lähteistä ja "muotoillaan" ne kuutioiksi. Siten käyttäjät voi käyttää kuutiota analysointiaan varten.

Erityisesti tietokuutioiden luomisen jälkeen niiden käyttö tulisi tulevaisuudessa nopeuttamaan, sekä helpottamaan tietojen hakemista. Käyttäjänä taulujen yhdistämistä ei enää välttämättä tarvitse, kun tieto kuutiosta voidaan noutaa nopeammin. Tämä edistää samaan aikaan myös raporttien optimointia, eli järjestelmä pystyy tekemään kyselyt nopeammin.

Tietokuutioiden hyödyt

1. **Nopeat tiedonhaut:** Tietokuutiot mahdollistavat nopean tiedonhaun ja ovat tämän vuoksi hyödyllisiä yrityksille, jotka käsittelevät suuria tietomääriä.
2. **Parempi skaalautuvuus:** Tietokuutiot ovat skaalautuvia ja mahdollistavat helpon lisäyksen uusille tietolähteille ja tietueille. Tämä tekee niistä hyvän valinnan yrityksille, jotka tarvitsevat joustavaa tietojen varastointiratkaisua.

3. **Parempi suorituskky:** Tietokuutiot ovat suunniteltu suorittamaan nopeasti monimutkaisiakin tietojen analyysiprosesseja.
4. **Moniulotteinen analyysi:** Tietokuutiot mahdollistavat moniulotteisen analyysin, jolloin käyttäjät voivat tarkastella tietoja eri näkökulmista ja tasoista.

Tietokuutioiden haitat

1. **Kustannukset:** Tietokuutioiden käyttöönotto ja ylläpito voivat olla kalliita, eikä sen takia ne ole järkeviä pienille yrityksille.
2. **Haavoittuvuus:** Koska tietokuutiot ovat yleensä suuria ja monimutkaisia, ne voivat olla alttiita tietoturvaongelmille ja hakkeroinnille.
3. **Ammattitaidoin vaativuus:** Tietokuutioiden käyttö vaatii yleensä ammattitaitoisen henkilöstön, joka osaa käyttää niitä tehokkaasti ja hyödyntää niiden potentiaalia.
4. **Tietojen laatu:** Tietokuutioiden toimivuus riippuu suurelta osin tietojen laadusta ja sen varmistamisesta, että kaikki tiedot on oikein tallennettu ja järjestetty.
5. **Tietojen kokorajoitukset:** Tämän vuoksi voidaan joutua tekemään laajoja tietojen yhdistämisä tai yhteenvetoja.
6. **Joustamattomuus:** OLAP-järjestelmät eivät välttämättä pysty helposti mukautumaan muuttuviin liiketoiminnan tarpeisiin ja niiden muuttaminen tai laajentaminen voi vaatia huomattavia toimenpiteitä.

2.2 Relaatietietokanta, SQL-kysely

Relaatietietokanta koostuu useista taulukoista. Taulukoissa tiedot esitetään riveillä ja sarakkeissa. Tietokannassa tarvittavat tiedot pyritään jakamaan taulukkoihin siten, että yksi tieto tallennetaan vain yhteen paikkaan. Relaatietietokantaan tallennetaan myös tieto siitä, miten eri taulukot liittyvät toisiinsa.

Relaatietietokantojen etuna voidaan pitää tietojen helppoa saatavuutta, sillä tietojen käyttö on mahdollista heti tietokantaan tallentamisen jälkeen. Taulukkojen välille luodaan relaatio eli yhteys, jotka nopeuttavat ja tarkentavat tietojen päivittämistä, sillä muutos on tehtävä vain yhteen paikkaan. Relaatietietokanta on suunniteltava huolellisesti, sillä hyvin suunnitellusta tietokannasta voi helposti tyydyttää monimutkaiset tietotarpeet ja voi välttää paljon turhaa työtä, kun ei joudu luomaan uudelleen koko tietokantaa ja syöttämään kaikkia tietoja uudelleen. (Sarja 2006.)

SQL kyselykielen (Structured Query Language) avulla käsitellään tietokantoja. Siitä tuli vuonna 1986 ANSI:n (American National Standards Institute) standardi, sekä vuonna 1987 ISO:n (International Organization for Standardization) standardi. SQL:stä on useita eri versioita, mutta ne pohjautuvat suurilta osin ANSI/ISO-standardiin. Ollakseen ANSI-standardin mukaisia, versioiden tulisi tukea ainakin tärkeimpiä komentoja (kuten SELECT, UPDATE, DELETE, INSERT, WHERE) samalla tavalla. (W3Schools 2023.)

Oulun ammattikorkeakoulussa on käytössä avoimen lähdekoodin MariaDB SQL-relaatiotietokanta, joka pohjautuu MySQL-relaatiotietokantaohjelmistoon. Molempien kehityksessä on huomioitu niiden yhteensopivuus keskenään.

Relaatiotietokantojen suurimpana hyötynä on niiden helppokäyttöisyys. Niissä olevan tiedon muokkaaminen, lisääminen ja poistaminen on tehty helpoksi, mutta kuitenkin tietoturva edellä. Joissain tapauksissa kuitenkin relaatiotietokantojen rajoitukset, sekä suurien tietomäärien luoma monimutkaisuus saattaa olla haittana. On siis erittäin tärkeää, että tietomalli suunnitellaan tarkkaan.

Relaatiotietokantojen hyödyt

1. **Helppokäyttöisyys:** Relaatiotietokantojen käyttö on yleensä helppoa ja käyttäjäystävällistä.
2. **Joustavuus:** Relaatiotietokannat ovat erittäin joustavia. Tietueiden lisääminen ja rakenteen muuttaminen omien tarpeiden mukaan on helppoa.
3. **Tietoturva:** Relaatiotietokannat tarjoavat yleensä paremman tietoturvan kuin muut tietokantajärjestelmät.
4. **Tietojen eheys:** Relaatiotietokannat varmistavat tietojen eheyden ja yhdenmukaisuuden, mikä tarkoittaa sitä, että tietojen tallennus ja muokkaus on helpompaa ja virheiden riski on pienempi.

Relaatiotietokantojen haitat

1. **Suorituskyky:** Relaatiotietokannat voivat olla hidas ja raskas käsitellä, kun tietokantaan tallennetaan suuria määriä tietoja.
2. **Rajoitukset:** Relaatiotietokantojen käyttöön liittyy tietyt rajoitukset, kuten tietokantaan rakenne, taulukoiden koot ja tietotyyppien käyttö.
3. **Kustannukset:** Relaatiotietokantojen käyttöönotto ja ylläpito voi olla kallista.
4. **Monimutkaisuus:** Relaatiotietokannat voivat olla monimutkaisia ja vaativia ylläpitää.

3 RAPORTIT

Raporttien tekeminen niitä tarvitseville henkilöille helpottaa datan tarkastelua ja analysointia. Microsoftin Power BI -työkalua käyttäen saadaan tarvittava tieto suuristakin datamääristä eri visualisointeihin, joista voidaan havainnollistaa kyseisen raportin trendejä sekä merkityksellisiä tietoja.

Oamkissa on rakennettu erilaisia raportteja pääosin tiedolla johtamiseen johtoa, sekä esihenkilöasemassa työskenteleviä varten. Raportteja on kuitenkin myös muulle henkilökunnalle, kuten esimerkiksi raportit hiilijalanjäljestä, opiskelijapalautekyselyistä sekä opiskelijatilastoista. Jatkossa myös opiskelijadata halutaan tuoda paremmin esille koko henkilökunnalle.

3.1 Talousraportointi

Talousraportoinnissa on aikaisemmin käytetty relaatiotietokantaa, johon data on tullut taloushallintojärjestelmästä. Vuoden 2023 alussa taloushallintojärjestelmän vaihtumisesta toiseen, Oulun yliopistossa jo käytössä olleiden tietokuutioden käyttöä alettiin suunnitella myös Oulun ammattikorkeakoululle. Tämän takia erillistä rajapintaa ei alettu rakentamaan tiedon siirtymiseen relaatiotietokantoihin, koska tarkoituksena on ottaa tietokuutiot käyttöön. Tämänhetkistä talouden tilannetta ei pystytä näin ollen tuomaan Power BI -raportteihin, koska taloushallintojärjestelmästä tulevan datan määrittelyä tietokuutioksi ei ole saatu vielä tehtyä.

3.2 Opiskelijatietoihin perustuva raportointi

Opiskelijatietoihin perustuva raportointi on toteutettu Oamkissa yksikkökohtaisilla raporteilla, nämä koottuna koulutuksen johtoryhmälle, sekä joitain muita raportteja esimerkiksi yhteishakuun liittyen. Yksikkökohtaisissa raporteissa on pääosin samat mittarit, mutta on kuitenkin joitain yksittäisiä raportteja, jossa yksikön johtajat ovat halunneet omanlaisiaan tunnuslukuja. Yksikkökohtaisten raporttien esimerkki kuvissa tuodaan esille tekniikan-, sekä luonnonvara-alan yksikön raportti.

Oulun ammattikorkeakoulun organisaatiomuutoksen myötä, yksikkökohtaiset raportit tulevat päivittymään osaamisala -raportteihin vuoden 2023 kesän aikana. Raporttien ulkoasua tullaan muuttamaan, sekä yhdistetään linjaa mittareista, jotta raporteista saataisiin yhteneväisiä. Tarkoituksena olisi tuoda myös opiskelijadata näkyviin koko opetushenkilöstölle.

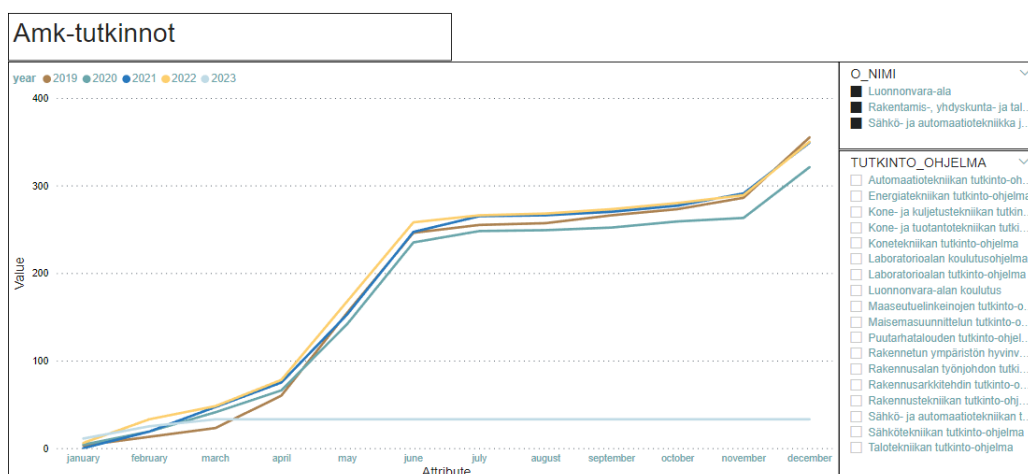
Muita raportteja opiskelijatietoihin perustuvasta datasta tehdään henkilöstön tarpeiden mukaan. Esimerkiksi Oamkin viestintä on tarvinnut yhteishakuun liittyvän raportin, josta pystyy tarkastelemaan mistä päin suomea hakemuksia on tullut, kuinka paljon paikkakunnittain ja niin edelleen.

3.2.1 Yksikkökohtaiset raportit

Yksikkökohtaisissa raporteissa on erilaisia mittareita tiedolla johtamisen tueksi. Mittarit ovat valittu raportteihin ammattikorkeakoulujen rahoitusmallin mukaisesti. Rahoitusmallissa suurin osuus tulee koulutuksesta, johon kuuluu suoritettut ammattikorkeakoulututkinnot, jatkuva oppiminen, työllistyminen, opiskelijapalaute, sekä ammatillisessa opettajakoulutuksessa suoritettut opinnot. Raporteissa olevat välilehdet koostuvat seuraavista tunnusluvuista:

AMK-, sekä YAMK-tutkinnot

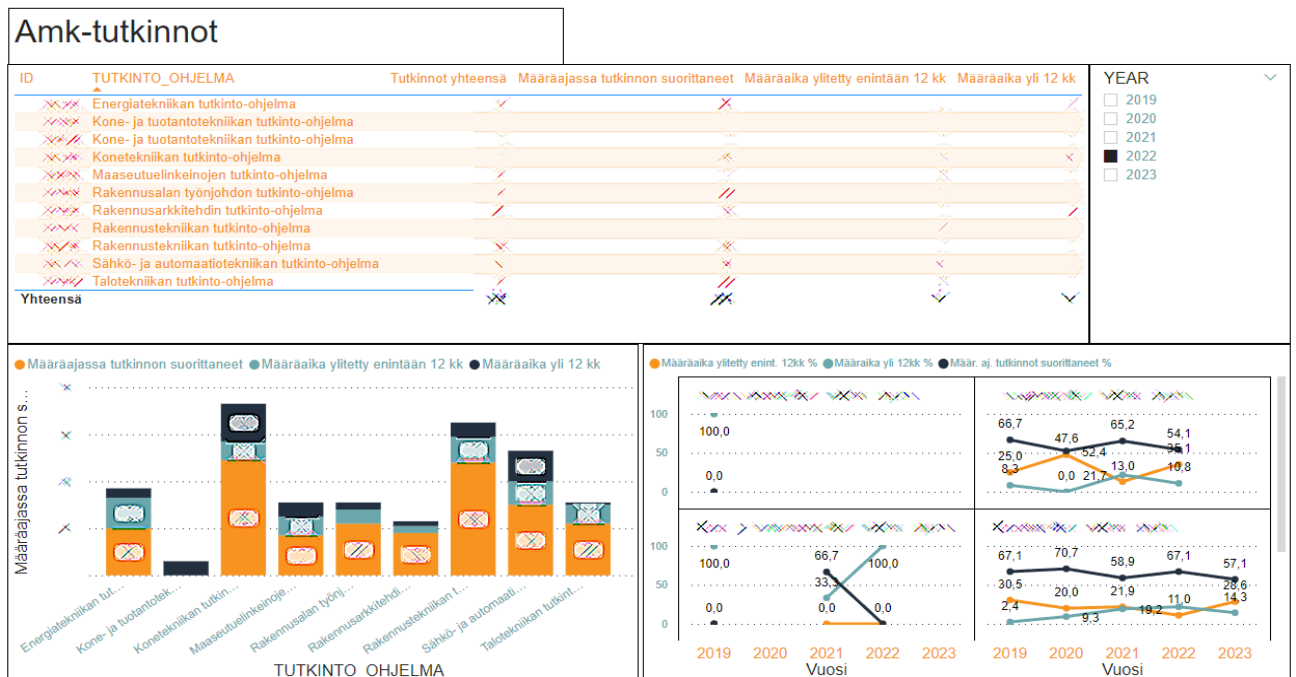
Kuvassa 3 olevassa graafissa AMK-, sekä YAMK-tutkinnot kuvataan viivakaaviolla. Viivakaaviossa eriväriset viivat kuvaavat tietyn vuoden trendiä. X-akseli kuvaa suoritettujen tutkintojen kuukautta ja Y-akselilla näkyy tutkintojen määrä. Graafia pystyy suodattamaan koulutusala, sekä tutkinto-ohjelma kohtaisesti.



Kuva 3 Viivakaavio -graafi AMK-tutkinnoista yksikköraportissa

Määräajassa tutkinnon suorittaneet

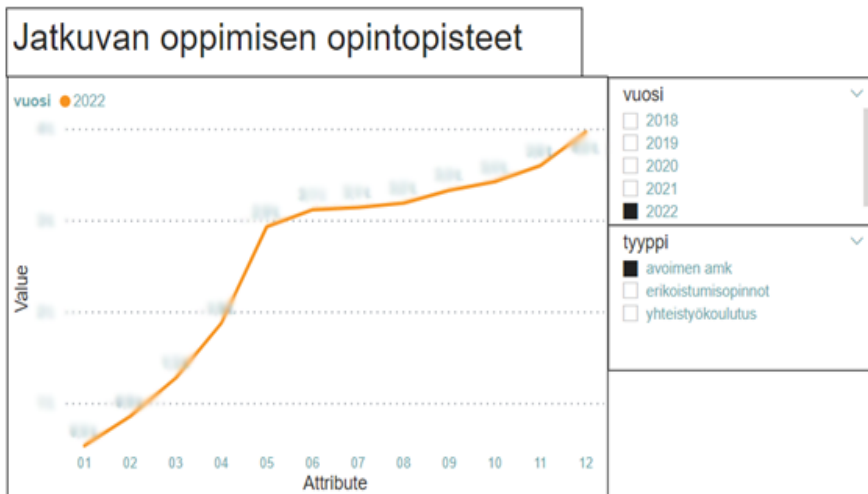
Kuvassa 4 olevassa määräajassa tutkinnon suorittaneiden graafissa tuodaan esille taulukkomuodossa tutkinto-ohjelmakohtaiset määrät vuositasona niin tutkinnoista yhteensä, määräajassa suorittaneista, määräaika ylitettynä enintään 12 kuukautta, sekä määräaika ylitettynä yli 12 kuukautta. Sama tieto tuodaan näkyville pylväsdiagrammina, jossa X-akselilla on tutkinto-ohjelmat ja Y-akselilla tutkintojen määrät. Kolmantena näkymänä on viivakaavio, jossa tutkinto-ohjelmat ovat niin sanotusti pieninä kerrannaisina omissa ruuduissaan, X-akselilla on vuosi ja Y-akselilla viivat eri värinä, onko tutkinto suoritettu määräajassa, ylitetty enintään 12 kuukautta vai ylitetty yli kahdella toista kuukaudella. Näissä määrät näkyvät prosentuaalisella osuudella kaikista tutkinnoista yhteensä.



Kuva 4 Määräajassa tutkinnon suorittaneiden välilehti yksikköraportissa

Jatkuva oppiminen ja opintopistekertymä

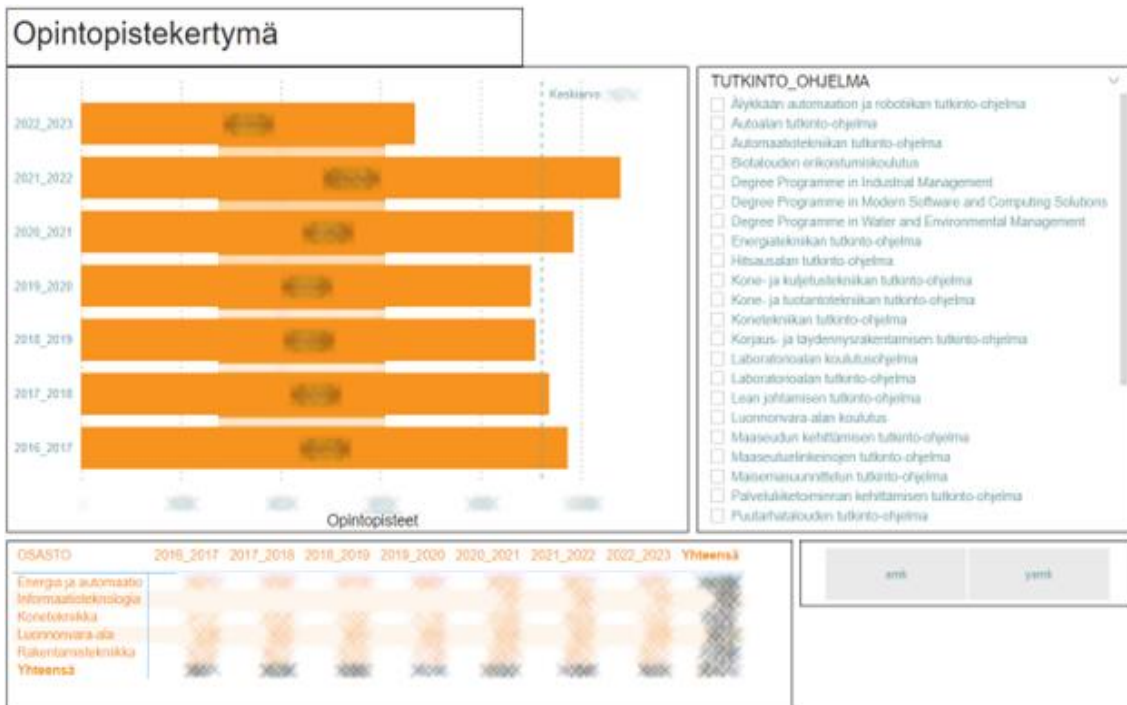
Kuvan 5 graafissa jatkuvan oppimisen opintopisteiden tiedot tuodaan viivakaavion avulla, kuinka paljon kyseisen yksikön jatkuvan oppimisen opintopisteitä on tullut kuukausitasolla. Graafissa X-akselilla on kuukaudet, Y-akselilla opintopisteiden määrä ja viivoja graafiin voi lisätä eri vuosilta. Tarkempaa tarkastelua pystyy tehdä suodattamalla esimerkiksi vain avoimesta ammattikorkeakoulusta tulevat opintopisteet.



Kuva 5 Jatkuvan oppimisen opintopisteiden välilehti yksikköraportissa

Opintopistekertymä

Kuvan 6 opintopistekertymä -graafissa vakionäkymänä on yksikön kaikkien tutkinto-ohjelmien ker-
rytetyt opintopisteet. Palkkikaaviossa Y-akselilla on lukuvuosi ja X-akselilla opintopisteiden määrä.
Lisäselitteenä näkyy kaikkien vuosien keskiarvo. Tarkempaa tarkastelua pystyy katsomaan palkki-
kaavion alla olevasta taulukosta, sekä opintopistekertymiä pystyy rajaamaan AMK, sekä YAMK
tasolla ja tutkinto-ohjelmittain.



Kuva 6 Opintopistekertymän välilehti yksikköraportissa

Julkaisut

Kuvan 7 graafi esittää julkaisujen määrän pylväsdiagrammimuodossa, jossa eri väriset pylväät ovat julkaisuvuoden mukaan ja Y-akselilla on julkaisujen määrä. Julkaisujen määrässä on viivettä, koska julkaisualusta Juulissa ei ole toimivaa rajapintaa, jota kautta tiedot voitaisiin päivittää. Tiedot julkaisusta kerätään Excel-tiedostoon, jota päivitetään säännöllisesti.



Kuva 7 Julkaisujen välilehti yksikköraportissa

TK-hankkeet

Tulevassa osaamisala -raporteissa TK-hankkeiden tilastointi jää pois ja TK-hanketta koskevat tunnusluvut tulevat omalle raportille uuden TKI-yksikön tullessa organisaatiomuutoksen myötä. Tällä hetkellä olevissa yksikkökohtaisissa raporteissa (kuva 8) on kuitenkin pylväsdiagrammit niin TK-hankkeiden lukumäärästä, kuin rahoituksesta. Näitä on tukemassa taulukko tarkempaa tarkastelua varten, sekä suodatusvalinnat vuodelle, projektin tyypille sekä projektin nykyiselle tilalle.

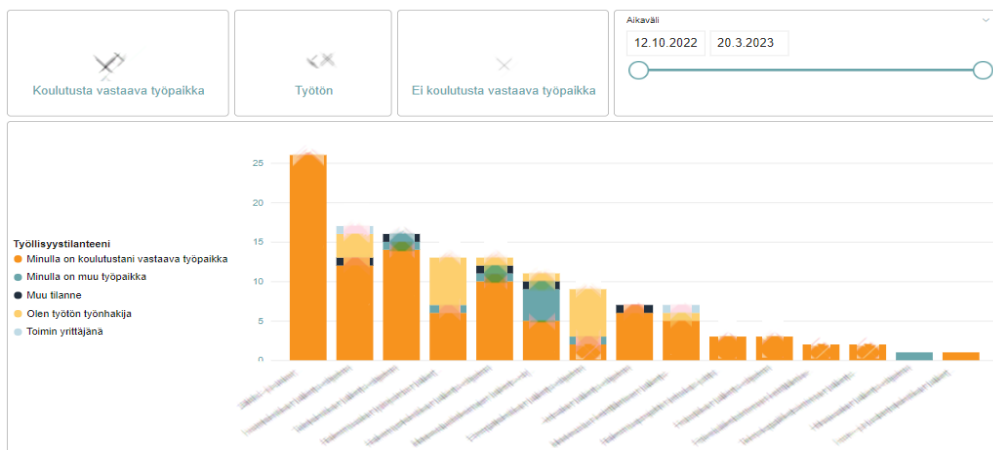


Kuva 8 TK-hankkeiden graafeja hankkeiden lukumäärästä, sekä rahoituksesta

Työllisyys

Kuvan 9 työllisyyteen keskittyvään graafiin tehdyn klusteroidun pylväsdiagrammin tieto tulee valmistumista edeltävän kyselyn kautta. X-akselilla on tutkinto-ohjelmat ja Y-akselilla on opiskelijoiden määrät. Selitteenä opiskelijamäärille on eri työllisyystilanteet omina väreinään pylväissä.

TYÖLLISYYS TUTKINTO-OHJELMITAIN 7.10.2022 ->



Kuva 9 Työllisyyttä mittaava graafi

3.2.2 Muut raportit

Muita rapotteja Oamkissa on useita, kuten esimerkiksi hiilijalanjälki-, opiskelijapalautekysely- ja yhteishakuraportti. Tieto eri raportteihin tulee usein kaikki eri tietolähteistä. Esimerkiksi hiilijalanjälki-raportissa näytettävä tieto tulee eri puolelta kerättyihin tietoihin ja ne ovat koostettu relaatiotietokannas olevaan tauluun.

Kuvassa 10 olevassa hiilijalanjälkiraportissa on haluttu tuoda Oulun ammattikorkeakoulun henkilöstölle näkyviin päästöt helposti tulkittavassa muodossa. Päästöt ovat luokiteltu kolmeen eri kategoriaan: kiinteistöt, matkustaminen ja hankinnat. Etusivulla näkyvät päästöt vuosittain, sekä selitteet eri luokista.



Kuva 10 Hiilijalanjälki-raportin etusivu

4 RAPORTOINNISSA KÄYTETTÄVÄT TYÖKALUT

Raportin tekemisessä joudutaan käyttämään useampia eri työkaluja eri tarkoituksia varten. Kaikien työkalujen tai ohjelmistojen käyttö ei kuitenkaan näy välttämättä itse raportin visualisointien tekemisessä, vaan niiden käyttäminen saattaa edellyttää esimerkiksi sitä, että data tietokannoista saadaan siirrettyä Power BI Desktop -sovellukseen. Eri työkalujen käyttö lisääntyy kuitenkin vielä entisestään, jos raportoinnissa aletaan käyttämään tietokuutioita. Tässä luvussa on lueteltu tällä hetkellä tarvittavat, sekä myöhemmin tulevaisuudessa tarvittavat työvälineet.

4.1 Microsoft Power BI

Power BI on kokoelma ohjelmistopalveluja, sovelluksia ja yhdistimiä, jotka yhdessä muuntavat toisiinsa liittymättömistä lähteistä peräisin olevan tiedon johdonmukaisiksi, visuaalisesti vaikuttaviksi ja vuorovaikutteisiksi näkemyksiksi. Tiedot voivat olla Excel-laskentataulukossa tai kokoelma pilvipohjaisia ja paikallisia hybriditietovarastoja. Power BI:n avulla voit helposti muodostaa yhteyden tietolähteisiin, visualisoida ja löytää tärkeitä asioita sekä jakaa niitä kenen kanssa haluat. (Microsoft Learn 2023a)

4.1.1 Power BI Service

Power BI Service on turvallinen Microsoftin isännöimä pilvipalvelu, joka antaa käyttäjien näyttää työpöytiä, raportteja ja Power BI sovelluksia ja niiden sisältöä, joka yhdistää toisiinsa liittyviä työpöytiä ja raportteja, käyttäen selainta tai mobiilisovellusta, jotka löytyvät kaikille käyttöjärjestelmille. (Sopanen, 2021).

Oamkissa Power BI Servicen kautta raportteja julkaistaan henkilökunnalle eri tasoittain. Työtiloja on useampia ja niiden näkyvyys on rajoitettu työtilan mukaan joko yksiköittäin, hallituksen johtoryhmälle sekä koko henkilöstölle. Power BI Servicen kautta työtilojen järjestelmänvalvoja voi jakaa raporteille käyttöoikeuksia ja asettaa esimerkiksi ajastetun päivityksen.

4.1.2 Power BI Desktop

Power BI Desktop on ilmainen sovellus, jonka asennat paikalliseen tietokoneeseen ja jonka avulla voit muodostaa yhteyden tietoihin sekä muuntaa ja visualisoida niitä. Power BI Desktopin avulla voit muodostaa yhteyden moniin eri tietolähteisiin ja yhdistää ne tietomalliin (tätä kutsutaan usein mallintamiseksi) (kuva 11). Tietomallin avulla voit luoda visualisointeja ja visualisointikokoelmia, joita voit jakaa raportteina muiden organisaatiossasi olevien henkilöiden kanssa. (Microsoft Learn 2023b).



Kuva 11 Esimerkki tietomallista ja tietojen yhdistämisestä

Power BI Desktop -työpöytäsovellus on raporttien luomisessa yksi tärkeimmistä ja eniten käytetyimmistä työkaluista. Datan hakeminen, sen muuntaminen ja yhdistäminen, taulukkojen ja graafien visualisointi ja kaikki muu näkyvä osuus raportissa tehdään tämän sovelluksen avulla.

4.1.3 Käyttöoikeustyytit

Taulukon 1 mukaisesti Power BI:n käyttäjäkohtaisia käyttöoikeustyyppiejä on kolmea erilaista: maksuton käyttöoikeus, Pro-käyttöoikeus, sekä käyttäjäkohtainen Premium (PPU). Eri käyttöoikeustyyppien tarve riippuu useasta eri asiasta. Tarvitseeko käyttäjä Premium-ominaisuuksia, minne sisältö tallennetaan sekä miten sitä käytetään. Käyttöoikeustyyppiejä on myös kapasiteettipohjaisina: kapasiteettipohjainen Premium-käyttöoikeus, PPU- ja Pro-käyttäjät, joilla on Premium kapasiteettipohjaisen käyttöoikeuden käyttöoikeudet, voi luoda sisältöä myös työtiloissa, jotka ovat Premium-kapasiteetti käyttöoikeuden alla.

Käyttöoikeuksia on siis yksittäisiä, eli käyttäjäkohtaisia käyttöoikeuksia ja käyttöoikeuksia, jotka ovat tilauksia, eli yrityksen tai organisaatioiden ostamia tallennuskapasiteetin tyyppisiä. Käyttäjäkohtaiset käyttöoikeudet ovat varsinkin isoissa organisaatioissa hyviä siitä, ettei kalliita tilauksia tarvitse olla useampia, koska kaikkien käyttäjien ei tarvitse luoda raportteja (taulukko 1).

Taulukko 1: Power BI:n käyttöoikeustyytit

Käyttöoikeustyyppi	Toiminnot, kun työtila on jaetussa kapasiteetissa	Lisätoiminnot, kun työtila on Premium-kapasiteetissa
Power BI (Ilmainen)	Käyttöoikeus sisältöön, jonka he luovat itse.	Pro- tai PPU-käyttäjien heille jakaman sisällön käyttäminen
Power BI Pro	Sisällön julkaiseminen toisiin työtiloihin, raporttinäyttöjen jakaminen, raporttien ja raporttinäkymien tilaaminen, jakaminen käyttäjille, joilla on Pro-käyttöoikeus	Sisällön jakaminen käyttäjille, joilla on maksuton tai PPU-käyttöoikeus
Power BI Premium per käyttäjä (PPU)	Sisällön julkaiseminen muihin työtiloihin, koontinäyttöjen jakaminen, koontinäyttöjen ja raporttien tilaaminen, jakaminen käyttäjille, joilla on PPU-käyttöoikeus	Sisällön jakaminen käyttäjille, joilla on maksuton ja Pro-käyttöoikeus

Power BI Pro

Power BI Pro on maksullinen pilvipohjainen ohjelmisto. Suurin ero Power BI Desktopin ja Pro:n välillä on kyky työskennellä yhdessä toisten Power BI käyttäjien kanssa ja jakaa raportteja sekä visualisointeja toisille Power BI -käyttäjille organisaatiossa (Sopanen, 2021).

Power BI Pro:n avulla käyttäjä voi esimerkiksi upottaa visualisointeja Power Bi sovelluksiin, integroitua esimerkiksi Azuren datapalveluun, jakaa raportteja toisille Power Bi käyttäjille, luoda työtiloja ja julkaista työpöytiä ja raportteja organisaation ulkopuolisille henkilöille, joilla on Power BI pro -lisenssi.

Power BI Premium

Power BI Premium -lisenssillä on Pro -lisenssiä parempi suorituskky ja skaalautuvuus. Premium -lisenssi on koko organisaation tasolla, kun taas Pro -lisenssi on henkilökohtainen. Kuvassa 12 olevassa taulukossa on laaja vertailu eri lisenssien välillä olevista ominaisuuksista.

Feature ¹	Power BI Pro	Power BI Premium Per user	Power BI Premium Per capacity
Collaboration and analytics			
Mobile app access	●	●	●
Publish reports to share and collaborate	●	●	
Paginated (RDL) reports	●	●	●
Consume content without a per-user licence			●
On-premises reporting with Power BI Report Server			●
Data prep, modelling, and visualisation			
Model size limit	1 GB	100 GB	400 GB
Refresh rate	8/day	48/day	48/day
Connect to more than 100 data sources	●	●	●
Create reports and visualisations with Power BI Desktop ²	●	●	●
Embed APIs and controls	●	●	●
AI visuals	●	●	●
Advanced AI (text analytics, image detection, automated machine learning)		●	●
XMLA endpoint read/write connectivity		●	●
Dataflows (direct query, linked and computed entities, enhanced compute engine)		●	●
Datamart creation		●	●
Governance and administration			
Data security and encryption	●	●	●
Metrics for content creation, consumption, and publishing	●	●	●
Application lifecycle management		●	●
Multi-geo deployment management			●
Bring your own key (BYOK)			●
Autoscale add-on availability			●
Maximum storage	10 GB/user	100 TB	100 TB
	Buy now >	Buy now >	Contact sales >

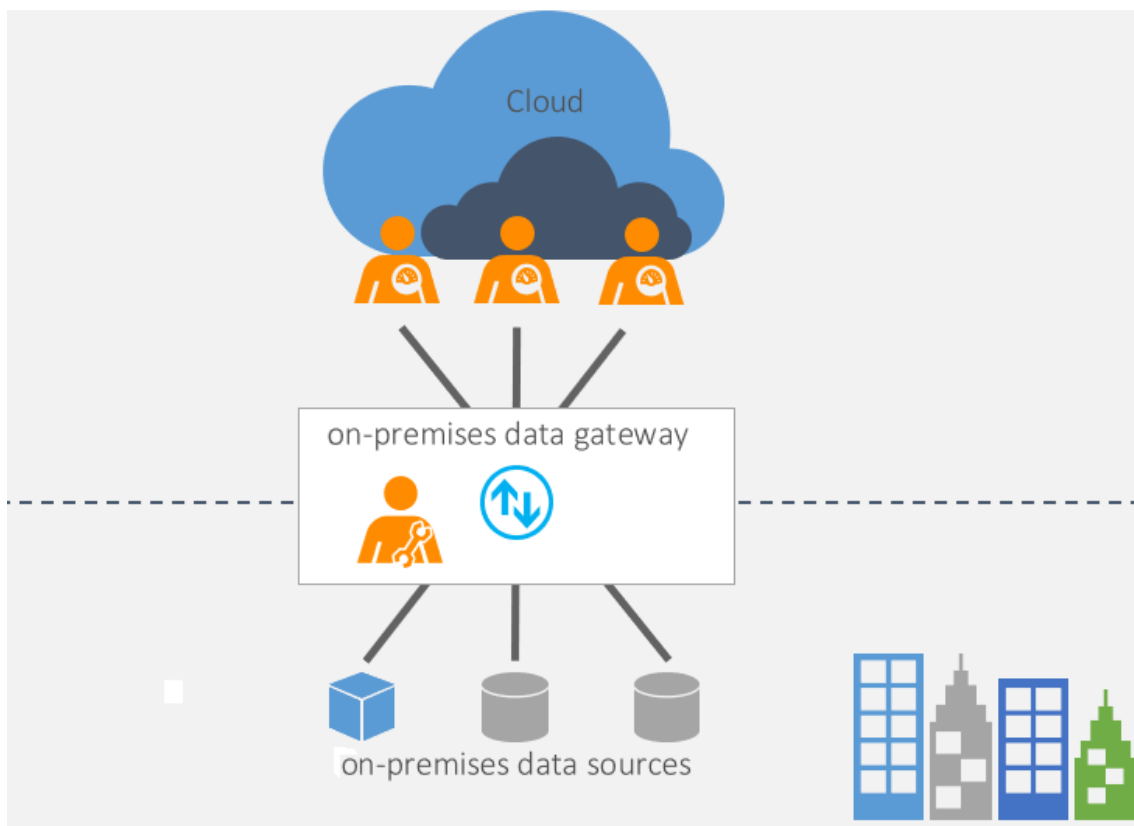
¹ Representation of features available in general availability (GA) status.

² Power BI Desktop is the data exploration and report authoring experience for Power BI, and it is available as a free download.

Kuva 12 Power BI -lisenssien välinen vertailu

4.2 Microsoft On-premises data gateway

Paikallinen tietoyhdyskäytävä tarjoaa turvallisen tiedonsiirron paikallisten tietojen ja useiden Microsoft-pilvipalvelujen, kuten Power BI:n, Power Appsin, Power Automaten, Azure Analysis Servicesin ja Azure Logic Appsin, välillä. Kuvassa 13 kuvataan tietoyhdyskäytävän toimintaperiaatetta. On-premises data gateway toimii näin ollen ”siltana” tiedonsiirrossa tietolähteestä pilvipalveluun. Paikallisen tietoyhdyskäytävän avulla raportteja ei tarvitse päivittää manuaalisesti, vaan asetuksien avulla voidaan määrittää tietojoukkojen päivitys.

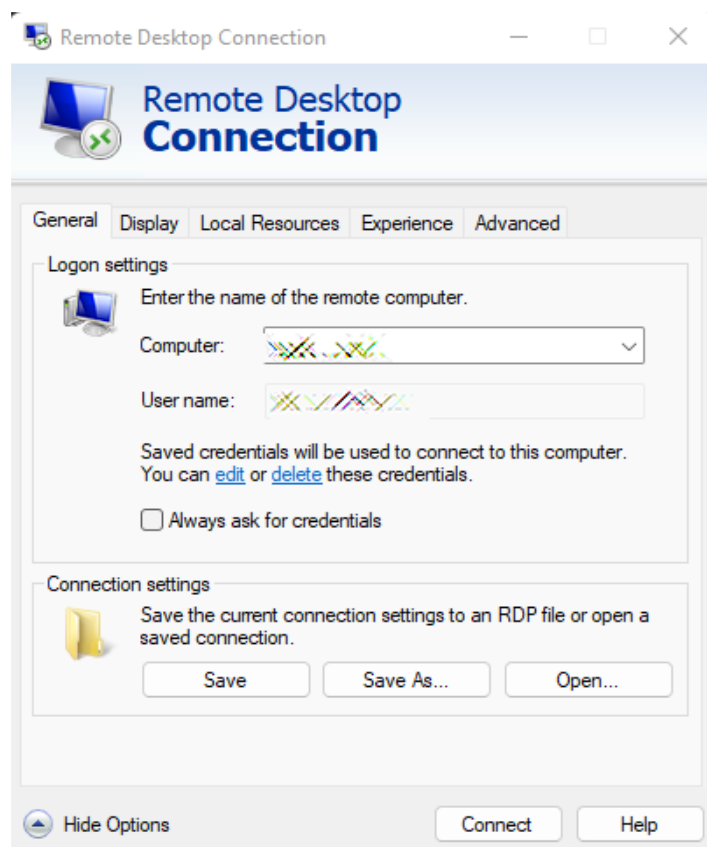


Kuva 13 On-premises data gatewayn toimintaperiaate

4.3 Remote desktop connection

Microsoftin julkaisemalla etätyöpöytä -sovelluksella voidaan ottaa yhteys etätietokonejärjestelmiin tai virtuaalisovelluksiin ja -työpöytiin. Tämän vuoksi voidaan käyttää joitain lisensoituja sovelluksia, sekä esimerkiksi yhteyksiä eri palvelimiin ei tarvitse luoda omalta tietokoneelta.

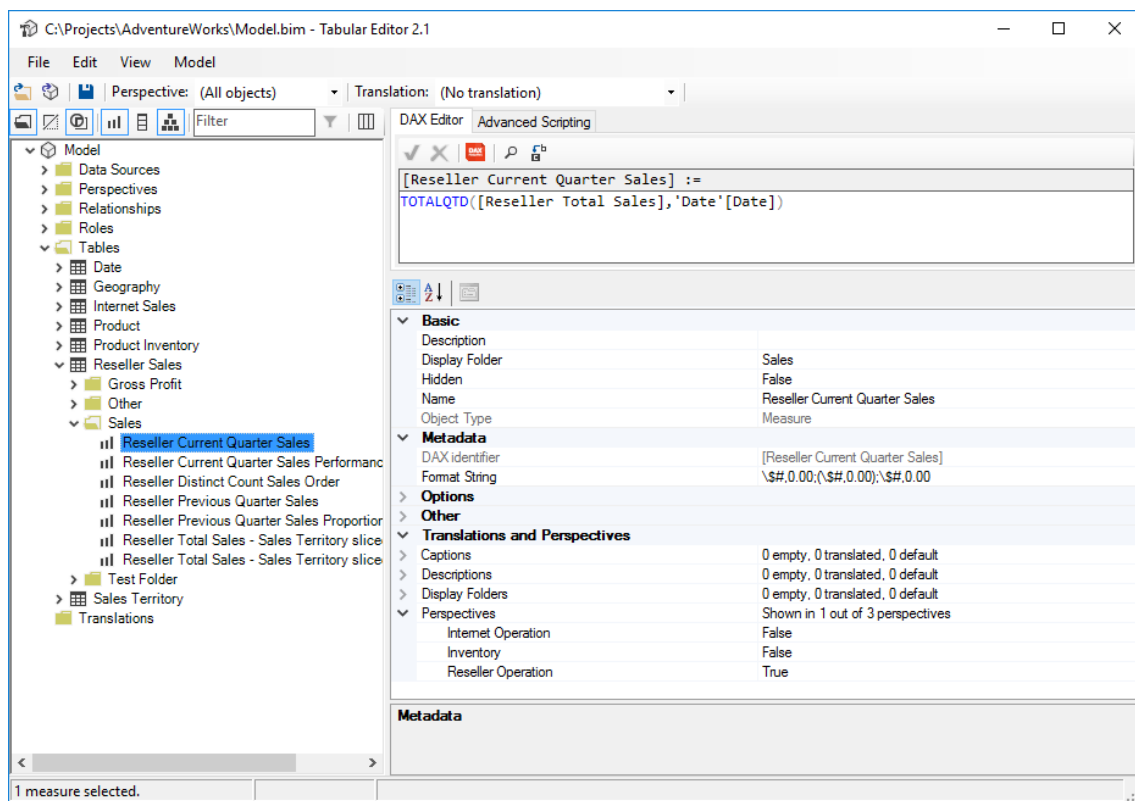
Tätä varten olemme saaneet Oulun yliopistolta vierailijatunnukset, eli niin sanotut UFO-tunnukset, joilla pystymme kirjautumaan yliopiston ylläpitämälle virtuaalityöpöydälle (kuva 14). Sovelluksen avulla pystymme käyttämään Tabular Editor 2 -sovellusta valmiiksi luoduilla tunnistetiedoilla.



Kuva 14 Kuvakaappaus Remote Desktop Connectionista

4.4 Tabular Editor 2

Tabular Editor (kuva 15) on vaihtoehtoinen editori SSDT:lle (SQL Server Data Tools) taulukkomallien luomiseen Analysis Servicesille (SSAS) ilman työtilapalvelinta. Sovellus on avoimen lähdekoodin sovellus, jolla muokataan .bim-tiedostoja, ilman että dataa käytetään suoraan mallista. Tämä offline-ominaisuus mahdollistaa nopeat muutokset suoraan .bim-tiedostoon, etenkin kun käsitellään mittareita (measure), laskettuja sarakkeita, kansioita, perspektiivejä ja käännöksiä. (SQLBI 2023). Sovelluksen avulla voidaan määritellä tietokuutiot, jota kautta jatkossa talousdata raporteille haetaan. Tietokuutioon tuleva data haetaan solmun (node) kautta ja sovelluksessa datasta luodaan lopulta tietokuutiomalli.

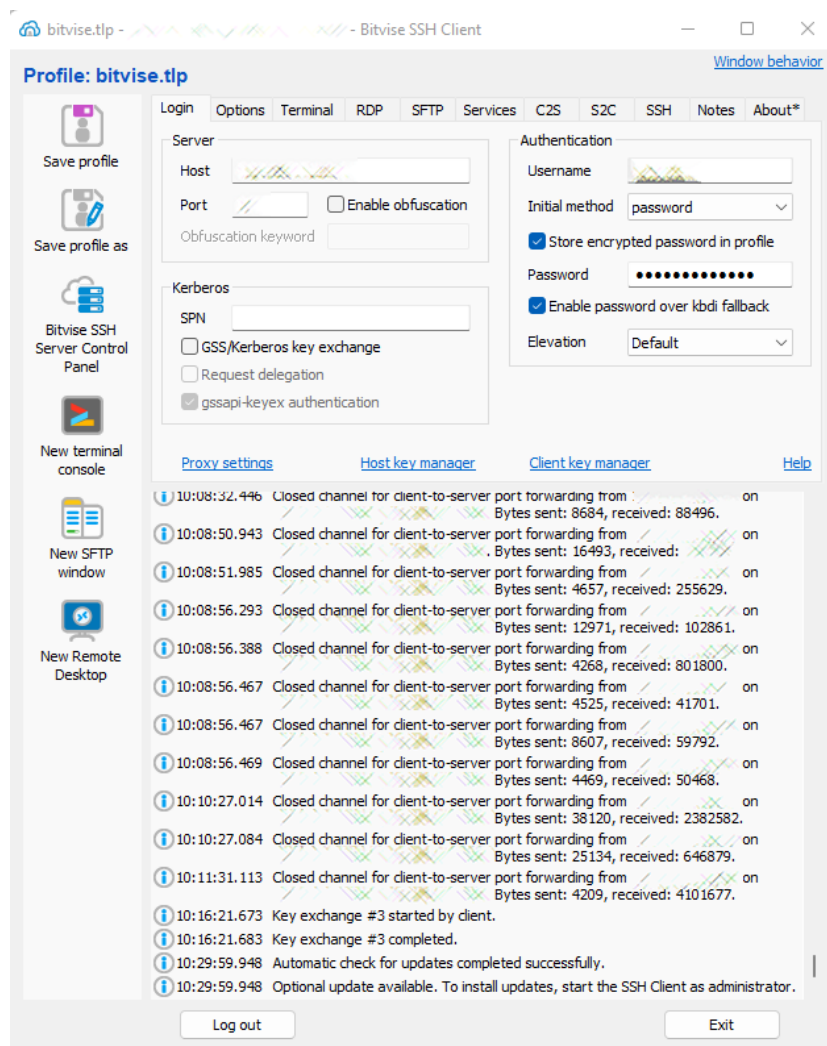


Kuva 15 Tabular Editor 2.1

4.5 Bitwise SSH Server/Client

Bitwise SSH (kuva 16) on tarkoitettu suojattujen tiedostojen siirtämiseen, Terminal Shelliin ja tunnelointiin. Tapauksessamme käytämme Bitwise SSH Clientia tunnelointiin. Palvelimessa käytetään

Bitvise SSH Serveriä, kun taas käyttäjän tietokoneella on SSH Client -sovellus. Tämä on niin sanottu client-to-server yhteys. Tapauksessani Oulun yliopiston ylläpitämiin palvelimiin otetaan yhteys, jotta voidaan hakea tietokannoista raportoinneissa käytettävää dataa.



Kuva 16 Kuvakaappaus Bitvise SSH Clientista, joka on yhteydessä palvelimeen

5 POWER BI -RAPORTTIEN LUOMINEN

Tässä luvussa käydään läpi, miten luodaan raportti käyttäen Power BI Desktop -työpöytäsovellusta. Tietokuutioiden käyttöönoton venyminen estää Power BI -raportin luomisen hyödyntäen tietokuutioiden kautta tulevaa dataa. Raporteissa käytettävä data tuodaan relaatiotietokannoista SQL-kyseilyillä, sekä verkosta tulevaa REST/JSON-rajapinnan kautta tulevaa koneluettavaa dataa. Ennen visualisointien luomista, dataa täytyy muokata, muuttaen esimerkiksi tietotyyppejä. Tietomalli täytyy rakentaa myös Power BI Desktopin sisällä, jotta tiedosta saadaan yhteneväistä.

Datan tuomisen, muokkaamisen ja tietomallin luomisen jälkeen, voidaan aloittaa visualisointien luominen, mittareiden tekeminen ja yleisen ilmeen rakentaminen raporteille. Tästä raportista tehdään raporttipohja, jota hyödynnetään osaamisalue-raporteissa, joissa tietojoukon sisältö pysyy käytännössä samana kaikissa ja suodattimia käyttämällä saadaan eri raporteissa näkymään tietoa haluamallaan tavalla.

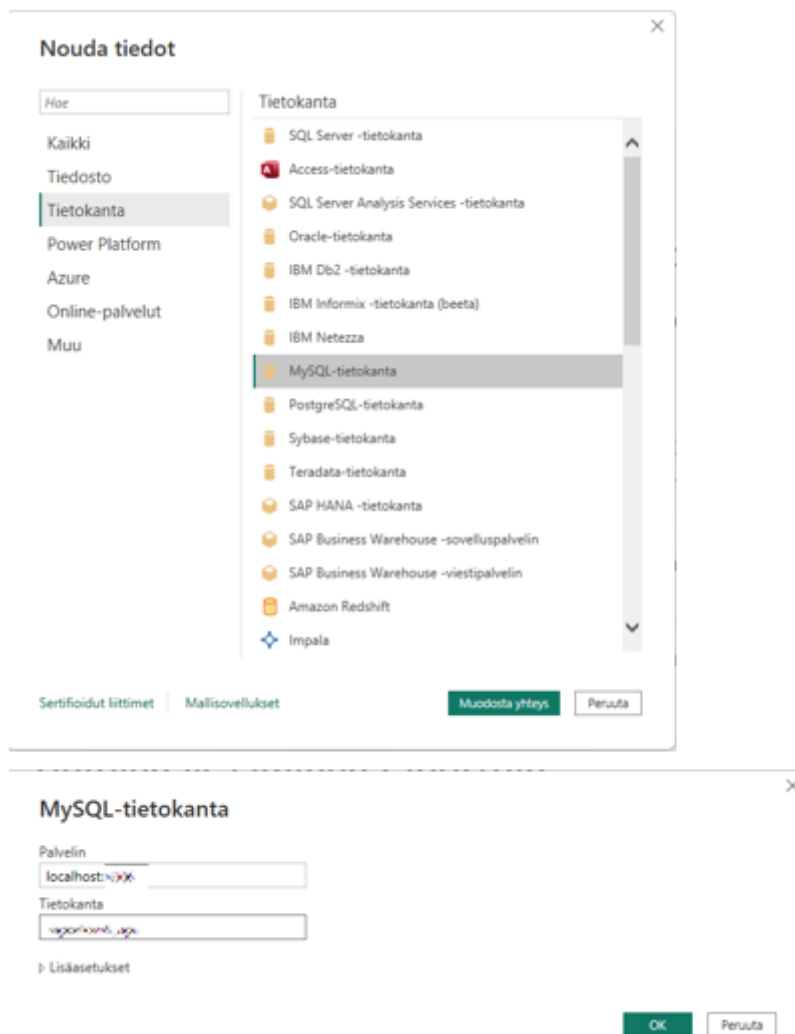
Raportin tulisi toimia sitä käyttävillä henkilöillä vaivattomasti, joten tarvittaessa raportin suorituskykyä tulisi testata. Suurten tietomallien ja datan määrä, sekä monimutkaiset mittarit hidastavat tiedon välittämistä graafeihin, joten tällaisissa tapauksissa tähän tulisi kiinnittää erityisesti huomiota.

Lopuksi itse raportti, sekä tietojoukko julkaistaan Power BI Service -pilvipalveluun, jotta muut käyttäjät pääsevät katsomaan raporttia. Pilvipalvelussa on mahdollista asettaa oikeuksia henkilöille, sekä ryhmille jotka pääsevät raporttia tarkastelemaan ja asettaa ajastettu päivitys tietojoukolle. Tapauksessani raportti sekä tietomalli julkaistaan testi -työtilaan.

5.1 Tietojen tuominen ja muokkaaminen relaatiotietokannasta

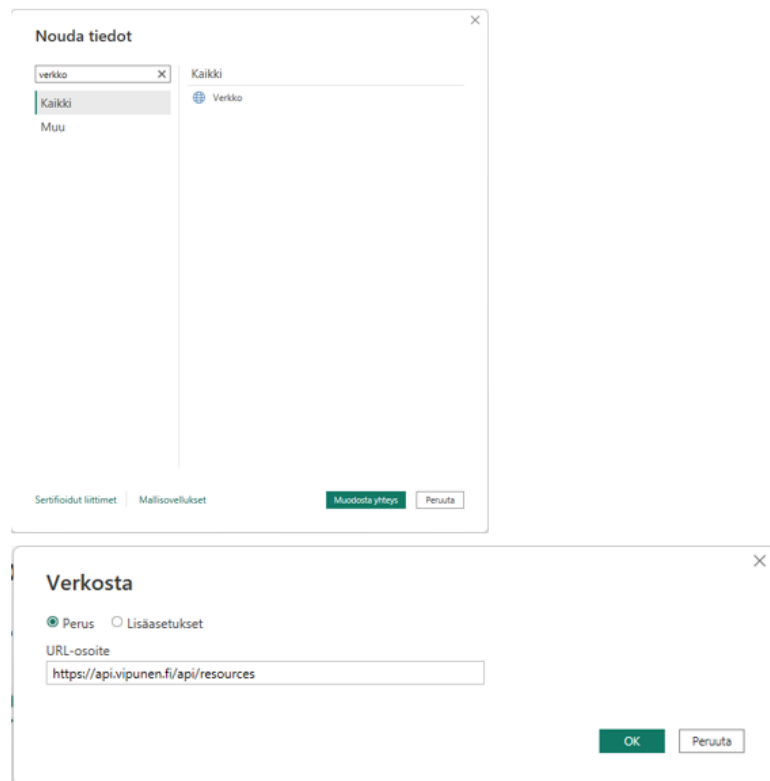
Tietojen tuominen valitsemasta lähteestä on raportin tekemisessä ensimmäinen asia, joka tulisi tehdä. Tarkoituksena on tuoda tietoja lähinnä MariaDB-tietokannasta, mutta lähteinä käytetään myös verkkoa, eli tietoja noudetaan opetushallituksen tarjoamasta Vipunen API -rajapinnan kautta. Tietojen noutaminen kyseisen rajapinnan kautta kerrotaan tarkemmin seuraavassa luvussa.

Aikaisemmassa luvussa mainitun Bitvisen avulla luodaan tunneliyhteys, jotta Power BI desktopin avulla pystytään luomaan yhteys palvelimelle, jossa tietokanta sijaitsee. Palvelimella toimivasta tietokannasta noudetaan tarvittavat taulut tietojoukkoon. Power BI desktop -työpöytäsovelluksessa avataan kuvan 17 tapaan Nouda tiedot -valikko ja etsitään tarvittava tapa tuoda tietoja. Tässä kyseisessä tapauksessa valitaan MySQL-tietokanta, syötetään palvelin ja tietokanta mihin halutaan ottaa yhteys ja valitaan OK. Tämän jälkeen ohjelma kysyy tarvittavat tunnukset, jos tietokanta on suojattu tunnuksilla. Sen jälkeen päästään valitsemaan tarvittavat taulut tietomalliin kuvan 18 siirtymistoinnolla sekä tarvittaessa myös valitsemaan taulut, joiden tietoja haluaa muuttaa.



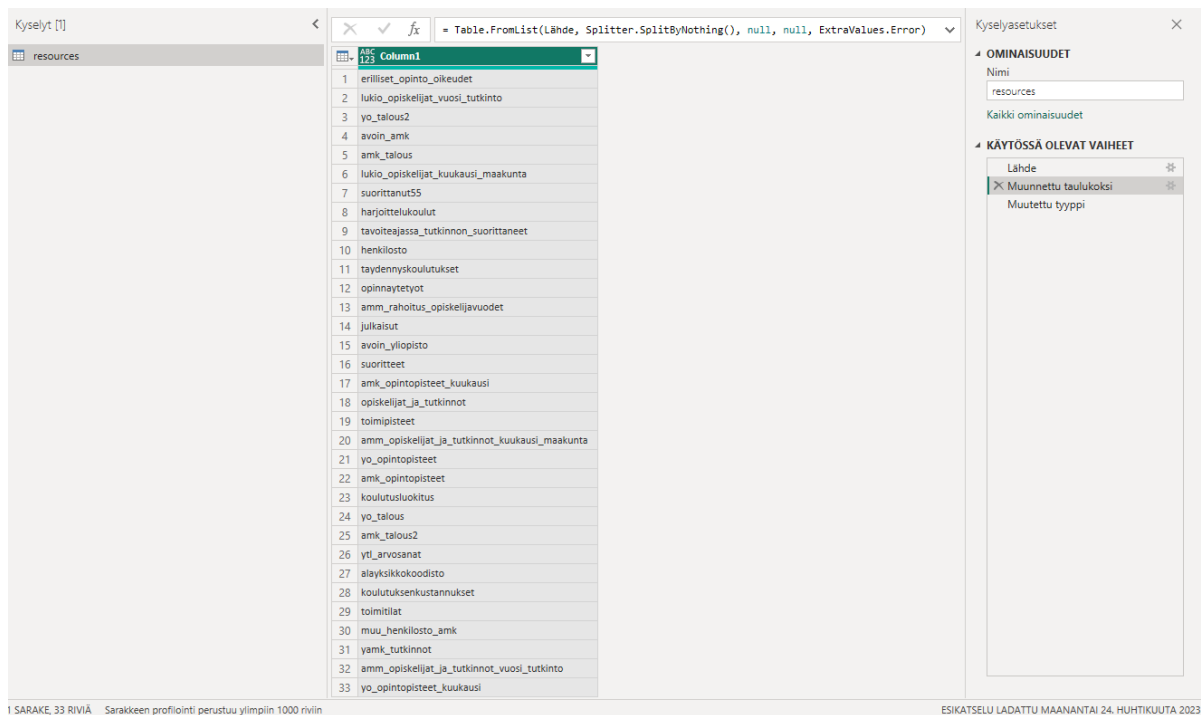
Kuva 17 Tietokantayhteyden luominen Power BI -desktop sovelluksella

Tietojen tuominen verkon kautta alkaa samalla tavalla, miten relaatiotietokantojen tuotiin aikaisemmassa luvussa. Nouda tiedot -ikkunassa valitaan verkko ja tämän jälkeen syötetään URL-osoite ja valitaan OK (kuva 19).



Kuva 19 Tietojen tuominen verkosta Power BI -työpöytäsovelluksessa

Tämän jälkeen aukeaa Power Query -editori ja tässä tapauksessa lisätty osoite (<https://api.vipunen.fi/api/resources>) palauttaa luettelon rajapinnan "resources" -listasta, eli listan saatavilla olevista tietojoukoista. Kuvassa 20 näkyy Power Query -editori, joka on automaattisesti luonut rajapinnasta tulleesta datasta taulukon. Taulukko itsessään sisältää vain tietojoukkojen otsakkeita, joiden sisällyttämästä datasta haluamme lopulta luoda tietokantataulun.



Kuva 20 Rajapinnan kautta palautunut lista

Tietojoukko -taulukon avulla voidaan luoda funktio, joka palauttaa tietojoukossa sisältämää dataa ja luo siitä taulun. Funktion voi luoda tietojoukko -taululle käyttämällä Power Queryn sisäistä Laajennettu editori -toimintoa (kuva 21).



Kuva 21 Laajennetulla editorilla kirjoitettu funktio

Kuvan 21 ensimmäisellä rivillä tietojoukko määritellään tekstimuotoon, rivi_0 numeeriseksi, sekä palautuva arvo tuodaan tauluna. Rivistä 4 eteenpäin, muuttujille annetaan erilaisia arvoja tarvittavissa muodoissa. Lopulta kun tarvittavien muuttujien arvot on annettu, määritellään muuttuja "output" ja palautetaan tämä kyseinen.

Kuvassa 22 näkyy funktion toiminnallisuus. Parametreihin lisätään tietojoukko -taulusta haluama tietojoukko, jonka funktio palauttaa taulukkomuodossa. rivi_0-kenttään syötetään arvo 0, tai haluama rivimäärä mitä funktio palauttaa. Arvo 0 palauttaa kaikki rivit.

X ✓ fx = (tietojoukko as text, rivi_0 as number) as table =>

Anna parametrit

tietojoukko
Esimerkki: abc

rivi_0
Esimerkki: 123

Käynnistä Tyhjennä

function (tietojoukko as text, rivi_0 as number) as table

Kuva 22 Luodun funktion toiminnallisuus

Datan määrä Vipusen rajapinnasta tulleista tietojoukoista on yleisesti valtava, koska dataa on jokaisesta korkeakoulusta, joissain tapauksissa toisesta asteesta yliopistoihin asti. Turhien sarakkeiden poistaminen ja tietojen suodattaminen on hyvin tärkeää, koska kaikkea saatavilla olevaa dataa ei tarvitse vertailla. Tietotyyppien vaihtaminen on myös tärkeää, koska haettu data taulukoissa on aina teksti muodossa.

5.3 Tietomallin rakentaminen

Tietomallin rakentamisessa tieto kerätään, muokataan ja yhdistetään eri lähteistä yhdeksi kokonaisuudeksi. Tietomallin rakentamisessa on hyvä ottaa huomioon se, miten tietoja raportoinnissa käytät ja onko sinulla kaikki tarvittava tieto tietokantojen tauluissa.

Yleisesti ottaen tehokkain malli on tähtimalli, jossa tietovarasto on rakennettu tähdenmuotoiseen malliin. Keskellä mallissa on faktataulu, joka sisältää keskeisen tiedon ja tämän ympärillä on niin sanottuja ulottuvuustauluja, jotka yhdistyvät faktatauluun tunnusavaimien kautta ja tekevät tietovarastosta yhtenäisen. Tähtimallin käyttämisellä helpotetaan suuren tietomäärän hallintaa ja käyttöä. Se on myös tehokas ja nopea tapa hakea tietoa, mikä taas parantaa suorituskykyä.

5.4 Raportin ulkoasu

Raportin ulkoasu on tärkeä osa raporttia, koska se on näkyvä osa raporttia tarkastelevalla henkilöllä. Power BI desktop -työpöytäsovelluksella voi valita teeman, jota muokkaamalla värimaailma, sekä esimerkiksi fontit pysyvät eri visualisoinneissa samanlaisena. Tässä tapauksessa raportti tulisi olla myös yhtenäinen, koska luodaan raporttipohja, jota muokataan pelkästään tietojen suodattimien avulla.

Kuvassa 24 näkyy valmiin raporttipohjan ilme. Valitsin teemaan harmaansävyisen taustan, sekä visualisointien pohjaan vaaleamman harmaan, jotta se erottuisi taustasta. Teeman värit, joita käytetään visualisoinneissa, kuten pylväsdiagrammien pylväiden väreissä, valitsin valmiista pohjasta, jossa on huomioitu saavutettavuus, sekä myös Oamkin värit. Pääväreinä toimivat musta, oranssi, punainen, turkoosi, violetti sekä sininen. Tekstin fontiksi valitui DIN-fonttiperhe, joka mielestäni erottuu hyvin taulukoissa, sekä on tyyliältään sopiva.



Kuva 24 Teemassa on tuotu Oamkin luvut näkyviin raportin eri välilehdillä oranssilla värillä

5.5 Laskentakaavojen määrittäminen

Laskentakaavat, eli mittarit (measures) suorittavat laskutoimituksia määriteltäviin arvoihin Data Analysis Expressions, eli DAX-kaavakielen avulla. DAX-kaavat käyttävät monia samoja funktioita, operaattoreita ja syntakseja kuin Excel-kaavat. DAX-funktiot on kuitenkin suunniteltu toimimaan relaatiotietojen kanssa ja suorittamaan dynaamisempia laskutoimituksia, kun teet toimia raporteissasi. DAX-funktioita on yli 200, ja ne tekevät kaikkea yksinkertaisista koosteista, kuten summa ja keskiarvo, monimutkaisiin tilasto- ja suodatusfunktioihin. Saatavilla on useita resursseja, joista saat lisätietoja DAX-kaavasta. (Microsoft Learn 2023c).

Raporttipohjassa käytetään yli viittäkymmentä eri mittaria, jota hyödynnetään kaikkia eri tarkoituksissa. Kaikkien mittareiden lukuja ei kuitenkaan näytetä visualisoinneissa, vaan jotkut mittarit toimivat vain laskemaan osa visualisoinneissa näkyvillä olevista mittareista.

Esimerkiksi Työllistymisen -välilehdellä tarvitaan useita erilaisia mittareita, jotta lukuja pystytään näyttämään tarpeeksi informatiivisesti. Tiedot tälle välilehdelle tulevat applications -taulukosta, joka on valmistumisjärjestelmästä tullut Excel-taulukko, jossa on tiedot valmistumishakemuksen tekneistä opiskelijoista. Tietosuoja syistä opiskelijat lasketaan valmistumispäivämäärän mukaan, koska taulukosta on poistettu opiskelijanumerot, jotta yksittäisiä henkilöitä ei pystytä tunnistamaan.

Ensimmäiseksi luotiin kuvan 25 mukainen mittari, jolla lasketaan kaikki taulukossa olevat opiskelijat. Kaavan rivillä 1 nimetään mittari ja laitetaan laskettava kaava IF-funktion sisälle. IF-funktio tarkistaa ja palauttaa arvon, jos se on TOSI. Muussa tapauksessa funktio palauttaa jälkimmäisen arvon. Rivin 2 ISBLANK -funktio tarkistaa, onko arvo tyhjä ja palauttaa joko arvon TOSI tai EPÄ-TOSI. Jos arvo on tosi, rivillä 3 palautetaan arvoksi 0. Jos rivillä ei ole tyhjiä rivejä, arvoksi palautetaan rivi 4. Tämä laskee opiskelijoiden määrän valmistumispäivämäärän mukaan.

```
1 count_opiskelijat = IF(  
2     ISBLANK(COUNT(applications[Valmistumispäivämäärä])),  
3     0,  
4     COUNT(applications[Valmistumispäivämäärä]))
```

Kuva 25 count_opiskelijat -mittarin kaava

Kun on luotu mittari, jolla pystytään laskemaan taulukossa olevat opiskelijat, vaikka suodattimia olisi käytössä, voidaan count_opiskelijat -mittaria hyödyntää muissa mittareissa. Jos halutaan esimerkiksi laskea koulutusta vastaavan työn työllisyysprosentti, luodaan ensiksi mittari, joka laskee kaikki opiskelijat, jotka ovat vastanneet työllisyystilanteeseen "Minulla on koulutustani vastaava työpaikka".

CALCULATE-funktioon syötetään laskettava lauseke, sekä määritetään suodatin tai suodattimia. Kuvan 26 rivillä 3 on laskettava lauseke, eli äskettäin luotu count_opiskelijat -mittari, jolla lasketaan opiskelijamäärä. Rivillä 4 ja 5 mittarille on annettu suodatin, jossa taulukossa "Työllisyystilanteeni" -sarakeessa palautetaan kaikki ne, joiden arvo on "Minulla on koulutustani vastaava työpaikka".

```
1 koulutustani vastaava työpaikka =  
2 CALCULATE(  
3     [count_opiskelijat],  
4     'applications'[Työllisyystilanteeni]  
5     IN { " Minulla on koulutustani vastaava työpaikka" }  
6 )
```

Kuva 26 koulutustani vastaava työpaikka -mittarin kaava

Kun on saatu opiskelijamäärä, joiden työllisyystilanne on koulutusta vastaava työpaikka, voidaan laskea työllistymisprosentti. Kuvan 27 luodussa mittarissa jaetaan henkilöt, joiden työllisyystilanne on koulutusta vastaava työpaikka, koko opiskelijamäärällä.

```
1 koulutustavastaava% = [koulutustani vastaava työpaikka]/  
    [count_opiskelijat]
```

Kuva 27 Koulutusta vastaavan työpaikan prosenttiosuus opiskelijoista

5.6 Testaaminen ja optimointi

Datan määrä tapauksessa ei ole niin valtavaa, että itse optimointiin tarvitsisi käyttää paljon resursseja. On kuitenkin tärkeä tiedostaa, että varsinkin suurten datamäärien käyttö ja monimutkaiset laskennat saattavat huonontaa käyttökokemusta raporttien käyttäjille. Power BI -raportteja pystyy optimoimaan tietolähteiden, tietomallien, visualisointien, sekä ympäristön, eli Power BI Servicen kapasiteetin, tietoyhdyskätävän ja verkon toiminnan kautta.

Power BI Desktop -työpöytäsovelluksessa on myös työkaluja raportin suorituskyvyn analysoimiseen, jolla pystyy tarkastelemaan aikoja, esimerkiksi kuinka kauan kullakin visualisoinnilla kuluu tietojen kyselyn tekemiseen.

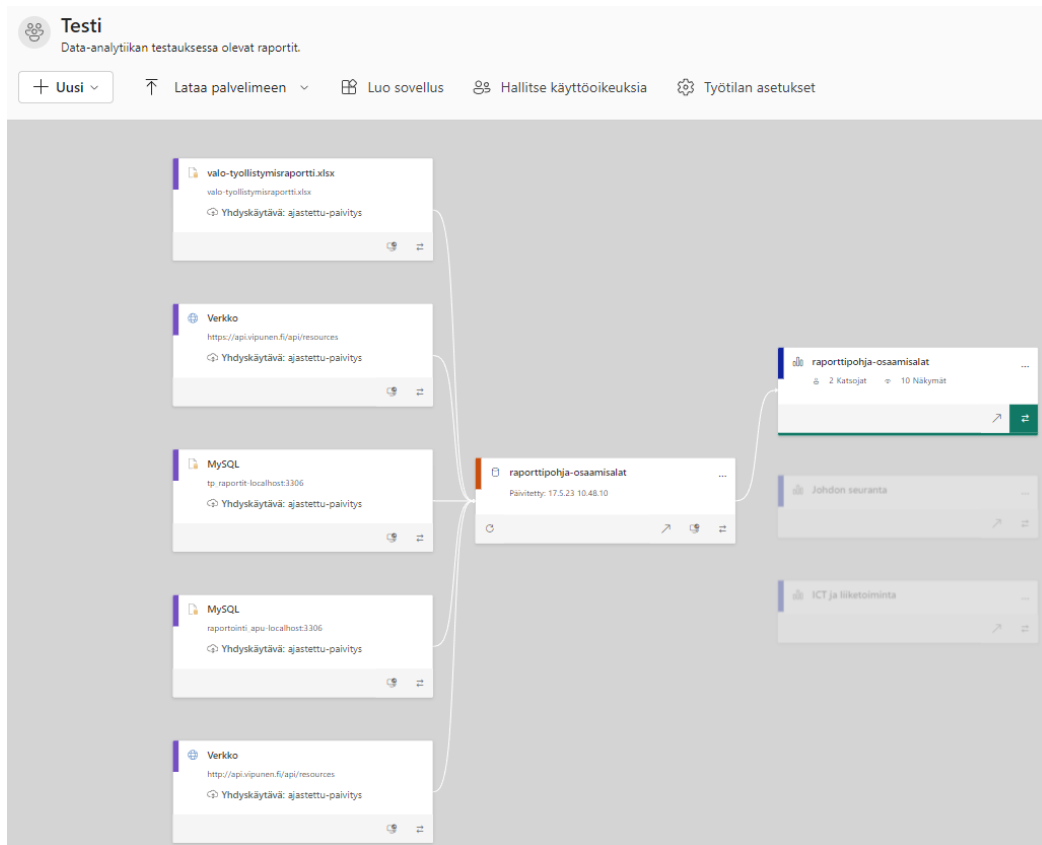
Testauksessa tulisi huomioida tietojen paikkansapitävyys, mittareiden laskukaavojen oikein toimivuus sekä raportin yleisilme, kuten oikeinkirjoitus, saavutettavuus ja muut vastaavat asiat.

5.7 Julkaiseminen

Julkaiseminen Power BI Service -pilvipalveluun on tietysti tärkeää siksi, että raportit saadaan muiden näkyville. Power BI Servicen puolella tulisi määritellä työtila, mihin tietojoukko ja itse raportti julkaistaan. Työtilaan voi määritellä, ketkä pääsevät raportteja katsomaan, niin henkilö-, kuin organisaatiotasolla.

Julkaiseminen tapahtuu Power BI Desktop -sovelluksen puolella painamalla ”Julkaise” -painiketta. Jos olet kirjautunut sovelluksessa, näet Power BI Servicen puolella olevat työtilat, joihin sinulla on kirjoitusoikeus. Valitsemalla työtilan johon raportin sekä tietojoukon haluat lisätä ja painamalla ”Valitse” -painiketta, sovellus alkaa siirtämään joukkoa Power BI Service -pilvipalveluun.

Power BI Service-pilvipalvelun työtilaan tulee kaksi tiedostoa, tietojoukko, sekä raportti. Tietojoukossa on raportilla olevat tietolähteet. Kuvan 28 Lineage-näkymässä pystyy tarkastelemaan, mitä tietolähteitä kyseisellä tietojoukolla on ja missä raporteissa kyseistä tietojoukkoa on käytetty.



Kuva 28 Kuvakaappaus Lineage-näkymästä Power BI Servicessä

Aikaisemmin Oulun ammattikorkeakoulussa on luotu jokaiselle yksikkökohtaiselle raportille omat tietojoukot. Tämä kuitenkin ei ole optimaalisin tapa käyttää tietojoukkoja, koska tietojen päivityksiä halutaan tehdä päivittäin ja tämä tarkoittaa, että jokaisen yksikön raportit täytyy päivittää yksitellen. Tietolähteet jokaisessa yksikkö -raportissa on kuitenkin lähestulkoon samat, joten optimaalisin tapa olisi käyttää yhtä tietojoukkoa ja jakaa tämä tietojoukko eri raporteille.

Power BI Servicen puolella pystytään myös asettamaan automaattinen päivitys tietojoukolle. Automaattisen päivityksen ottamisessa käyttöön tulee luoda yhdyskäytäväyhteys ja asettaa ”yhteyksien ja yhdyskäytävien hallinta” -välilehden kautta tietolähteiden tiedot, jotta yhdyskäytävä osaa luoda yhteydet päivitettäviin tietojoukkoihin. Tapauksessa On-premises data gateway -sovellus on asennettuna paikalliselle tietokoneelle, koska yhteyttä palvelimella toimivaan data gatewayhin ei ole saatu luotua.

6 JOHTOPÄÄTÖKSET

Opinnäytetyön tarkoituksena oli alun perin luoda Power BI -raportti hyödyntämällä tietokuutioita. Rajallisen ajan takia ja vastaan tulleiden esteiden takia, en kuitenkaan päässyt työssäni hyödyntämään tietokuutioita. Toiminnallisuudessa osuudessa luotiin kuitenkin uudenlainen raporttipohja Oulun ammattikorkeakoulun organisaatiomuutoksen myötä, hyödyntäen jo olemassa olevia relaatio-tietokantoja.

Tietokuutiot ovat erinomainen tapa käsitellä suuria määriä analysoitavaa dataa nopean tiedonhankunsa, sekä tietojen skaalautuvuuden takia. Niiden käyttöönotto on kuitenkin iso prosessi ja se saattaa viedä paljon aikaa ja resursseja. Uusien menetelmien käyttöönotossa täytyy huomioida myös ammattitaidon lisääminen ja tietokuutioiden ottamisessa käyttöön, ammattitaitoa vaatii paljon, jotta niitä pystytään käyttämään tehokkaasti ja hyödyntää niiden potentiaali. Se hyöty, mitä tietokuutioista saataisiin Oamkin raportoinneissa ei ole mielestäni työmäärään nähden tarpeellinen.

Raportoinneissa tällä hetkellä käytössä olevat jo valmiiksi rakennetut relaatiotietokannat, on täysin toimiva ratkaisu raporteille. Niiden ylläpito on tällä hetkellä Oulun ammattikorkeakoulun omien työntekijöiden saatavilla, eikä raporttien käytössä tai päivittämisessä ole huomattu minkäänlaisia suorituskyky ongelmia. Kaikki tarvittava tieto on saatu relaatiotietokantoihin, niiden käyttö on helppoa työntekijöille, sekä tietoturva yleisesti relaatiotietokannoissa on hyvällä tasolla. Verrattuna tietokuutioihin, relaatiotietokantojen tiedot ovat haasteellisia tulkita suoraan tauluista. Tietokuutioiden hyötynä on se, että erilaisia tauluja pystytään rakentamaan samasta datasta juuri sellaisia kuin halutaan. Relaatiotietokannoissa taulujen dataa joudutaan yhdistelemään keskenään, jotta joistain tauluista saadaan lukukelpoista.

Tietokuutioita tullaan hyödyntämään kuitenkin tulevaisuudessa Oamkin raportoinneissa, ainakin talousdataan liittyvissä raporteissa. Kuitenkin opiskelijadataan perustuvissa raporteissa olisi hyvä harkita, mitä hyötyjä tietokuutioista on raporttien luomisessa. Relaatiotietokantojen, sekä tietokuutioiden palvelin toimisivat molemmat Oulun yliopiston puolella, mutta oman käsitykseni mukaan, tietokuutioiden ylläpito olisi yliopiston puolella, kun taas relaatiotietokantojen ylläpito on täysin tai ainakin osittain Oamkin puolella. Kummankaan tavan kustannuksista en saanut tietoa, mutta tätä olisi hyvä myös pohtia, kun lopullista päätöstä tehdään.

LÄHTEET

Cloudflare 2023. What is tunneling? | Tunneling in networking. Hakupäivä 5.4.2023. <https://www.cloudflare.com/learning/network-layer/what-is-tunneling/>

Enho, Heidi 2020. Power BI – kaikki mitä sinun tulee tietää aloittaessasi. Blogipostaus 2020. Hakupäivä 30.11.2022. <https://hexcelligent.fi/2020/09/28/power-bi-kaikki-mita-sinun-tulee-tietaa-aloittaessasi-3/>

Github 2023. Kuvakaappaus. Artikkelissa otykier. Tabular Editor. Hakupäivä 13.4.2023. <https://github.com/TabularEditor/TabularEditor>

Imperva 2023. OSI Model. Hakupäivä 5.4.2023. <https://www.imperva.com/learn/application-security/osi-model/>

Imperva 2023. Valokuva. OSI Model. Hakupäivä 12.4.2023. <https://www.imperva.com/learn/wp-content/uploads/sites/13/2020/02/OSI-7-layers.jpg>

Microsoft Learn 2023a. Mikä Power BI on? Hakupäivä 25.5.2023. <https://learn.microsoft.com/fi-fi/power-bi/fundamentals/power-bi-overview>

Microsoft Learn 2023b. Mikä Power BI Desktop on? Hakupäivä 6.3.2023. <https://learn.microsoft.com/fi-fi/power-bi/fundamentals/desktop-what-is-desktop>

Microsoft Learn 2023c. Opetusohjelma: omien mittarien luominen Power BI Desktopissa. Hakupäivä 20.5.2023. <https://learn.microsoft.com/fi-fi/power-bi/transform-model/desktop-tutorial-create-measures>

Microsoft Learn 2022. What is an on-premises data gateway? Hakupäivä 23.5.2023. <https://learn.microsoft.com/en-us/power-bi/connect-data/service-gateway-onprem>

Microsoft Power BI 2023. Power BI pricing. Hakupäivä 23.5.2023. <https://powerbi.microsoft.com/en-au/pricing/>

Microsoft Support 2021. OLAP (Online Analytical Processing) – yleiskatsaus. Hakupäivä 23.5.2023. <https://support.microsoft.com/fi-fi/office/olap-online-analytical-processing-yleiskatsaus-15d2cdde-f70b-4277-b009-ed732b75fdd6>

Oracle Netsuite 2022. Valokuva. Artikkelissa Art Wittmann. What is OLAP? OLAP Defined. Hakupäivä 12.4.2023. <https://www.netsuite.com/portal/resource/articles/erp/online-analytical-processing-olap.shtml>

pradiptamukherjee 2023, Data Cube or OLAP approach in Data Mining. Hakupäivä 13.4.2023. <https://www.geeksforgeeks.org/data-cube-or-olap-approach-in-data-mining/>

Sarja, Jari 2006. Relaatietietokanta. Hakupäivä 25.5.2023. <https://verkkopedagogi.net/vanhat/fi/sisalto/materiaalit/access2003/luku0375c6.html?C>

Sopanen, Teemu 2021. Viisi syytä ottaa käyttöön Microsoft Power BI. Hakupäivä 6.3.2023. <https://www.etevat.fi/blogi/viisi-syyta-ottaa-kayttoon-microsoft-power-bi>

SQLBI 2023. Tabular Editor. Hakupäivä 5.4.2023. <https://www.sqlbi.com/tools/tabular-editor/>