



SAVONIA

OPINNÄYTETYÖ - AMMATTIKORKEAKOULUTUTKINTO
TEKNIIKAN JA LIIKENTEEN ALA

YLÄ-SAVON AMMATTIOPISTON TIETOVARASTON JATKOKEHITYS

TEKIJÄ: Eetu Kallio

Koulutusala Tekniikan ja liikenteen ala	
Koulutusohjelma/Tutkinto-ohjelma Tietotekniikan koulutusohjelma	
Työn tekijä Eetu Kallio	
Työn nimi Ylä-Savon ammattiopiston tietovaraston jatkokehitys	
Päiväys	23.4.2020
Sivumäärä/Liitteet	
Ohjaajat Keijo Kuosmanen, Pasi Liimatainen	
Toimeksiantaja Ylä-Savon ammattiopisto	
Tiivistelmä	
<p>Tämän opinnäytetyön aiheena oli Ylä-Savon ammattiopiston olemassa olevan tietovaraston jatkokehittäminen. Tavoitteena oli kasvattaa tilaajan tietovarastosta saamaa hyötyä. Painopisteenä oli raportoinnin helpottaminen ja tiedon visualisointi opiskelijatietojen osalta. Lisäksi työssä perehdyttiin tietovarastoinnin teoriaan ja käytettävissä oleviin tekniikoihin.</p> <p>Opinnäytetyö jakautui kolmeen osaan: asiakkaan tarpeiden kartoitukseen, kehitystöihin ja tulosten arviointiin. Asiakkaalla oli jo ennestään käytössään tietovarasto ja joitain työkaluja tiedollajohtamiseen, joten aluksi perehdyttiin niihin ja kartoitettiin, mitä sellaisia tarpeita asiakkaalla on, joihin silloiset työkalut eivät anna ratkaisua. Kolme suurinta teemaa olivat tiedon visualisointi, tiedon helppo saatavuus ja mahdollisuudet porautua yksityiskohtaisemmalle tasolle. Kehitystyöt veivät suurimman työmäärän ja niihin sisältyi mm. uusien ETL-prosessien luomista ja aiempien muokkaamista uuden tiedon tuomiseksi tietovarastoon ja uusien OLAP-kuutioiden luominen. Lisäksi opinnäytetyön aikana organisaatiossa otettiin käyttöön Power BI -ohjelmisto tiedon raportoinnin ja visualisoinnin avuksi. Lopuksi käytiin saadut palautteet läpi, arvioitiin tuloksia ja mietittiin mahdollisia jatkokehityskohteita.</p> <p>Opinnäytetyö antoi tilaajalle lisää työkaluja tiedolla johtamiseen ja säästi raportointiin kuluva työaika. Tietovarasto paransi ja helpotti tiedon saatavuutta ja käytettävyyttä. Palaute on ollut positiivista. Uudet kuutiot on otettu hyvin käyttöön ja ne on koettu helpompina käyttää kuin aiemmat.</p>	
Avainsanat Tietovarasto, OLAP, Power BI	

Field of Study Technology, Communication and Transport			
Degree Programme Degree Programme in Information Technology			
Author Eetu Kallio			
Title of Thesis Further Development of the Ylä-Savo Vocational College's Data Warehouse			
Date	April 23, 2020	Pages/Appendices	
Client Organisation /Partners Ylä-Savo Vocational College			
<p>Abstract</p> <p>The subject of this thesis was the development of the Ylä-Savo Vocational College's data warehouse. The goal was to increase the client's benefits gained from their data warehouse. The focal points were to make reporting easier and to improve the visualization regarding the information about students. Additionally, there was some research on the theoretical background and available techniques of data warehousing.</p> <p>The thesis work was divided into three parts: mapping the client's needs, the development work and analyzing the results. The client had a pre-existing data warehouse and some tools for knowledge-based management and thus the thesis was started with familiarizing with them and resolving what kind of needs the client has that are not met with the existing tools. Three most important areas were the visualization of information, the availability of information and possibilities to drill into more specific data. Development itself took the most part of working hours and it consisted of the creation and modification of ETL, extraction transformation and loading, processes and creating new OLAP, online analytical processing, cubes. During the process, the Power BI software suite was deployed into the organization to assist in visualization, availability, and reporting. Finally, the feedback was went through and the results were evaluated. Further development possibilities were also considered.</p> <p>As a result of this thesis, the client was provided with more knowledge-based management tools which will save time in reporting. The improved data warehouse improved the availability and usability fo the information. Feedback was mostly positive. The new cubes have been deployed into the organization well and they are seen easier to use than the previous ones.</p>			
<p>Keywords Data warehouse, OLAP, Power BI</p>			

SISÄLTÖ

1	JOHDANTO	5
1.1	Lyhenteet ja määritelmät	5
2	TIETOVARASTOT JA OLAP	6
2.1	Tietokuutiot ja OLAP.....	6
2.2	Tietomallit.....	8
2.3	Tietovarastoinnin hyödyt.....	9
2.4	Power BI.....	10
3	TILANNE JA TILAAJAN TARPEET	11
3.1	Ympäristö, ohjelmistot ja tekniikat.....	11
3.2	Ylä-Savon ammattiopiston tietovaraston tilanne.....	11
3.3	Primus	11
3.4	Raportointitarpeet	11
4	KEHITYSTYÖ.....	13
4.1	Integraatiot	13
4.1.1	PrimusQuery	13
4.1.2	ETL –prosessi.....	15
4.2	Kuutiot ja SSAS.....	17
4.2.1	Opintojen etenemisen seuranta	18
4.2.2	Opintojen keskeytykset.....	19
4.3	Kuutioiden käyttöönotto.....	19
4.4	Käyttöoikeuksien hallinta.....	20
5	TULOSTEN ARVIOINTI JA SAADUT PALAUTTEET	22
6	YHTEENVETO.....	23
7	LÄHDELUETTELO.....	24
8	KUVALUETTELO	25

1 JOHDANTO

Työn tarkoituksena on kehittää lisää työkaluja tiedollajohtamiseen olemassa olevan tietovaraston päälle. Tilaajana toimii Ylä-Savon ammattiopisto. Ylä-Savon koulutuskuntayhtymä on perustettu vuonna 1962 ja sen omistaa kahdeksan kuntaa: Iisalmi, Keitele, Kiuruvesi, Lapinlahti, Pielavesi, Rautavaara, Sonkajärvi ja Vieremä. YSAO tarjoaa toisen asteen ammatillista koulutusta sekä aikuiskoulutusta vuosittain noin 2300 opiskelijalle. Henkilökuntaan kuuluu noin 200 henkilöä. Tilaajan edustajana toimii tietohallintopäällikkö Ari Partanen.

Kehitys ja suunnittelu työt tehdään arkisin työajalla ja opinnäytetyöraportin kirjoitus iltaisin ja viikonloppuisin. Kehitystyötä tehdään tiiviissä yhteistyössä YSAO:n pedagogisen pääkäyttäjän Antti Korolaisen kanssa. Alkuperäisen tietovaraston toimittanut TCD Consulting and Research Oy tarjoaa teknistä tukea tarvittaessa, yhteyshenkilönä heidän päässään toimii Arto Pärnänen.

Opinnäytetyöprojekti jakautuu kolmeen päävaiheeseen. Ensimmäisenä tehtävänä on asiakkaan tilanteen ja tarpeiden kartoitus. Seuraavana on varsinaiset kehitystyöt, jota tulee olemaan määrällisesti paljon ja se vie suurimman osan ajasta. Tässä opinnäytetyössä keskitytään näiden osalta asiakkaalle merkittävimpiin ja muut käydään vain ohimennen. Lopuksi arvioidaan projektin onnistumista saatavien tulosten ja palautteiden perusteella ja pohditaan mahdollisia jatkokehityskohteita. Asiat käydään raportissa läpi tässä samaisessa järjestyksessä, mutta todellisuudessa projektin osia tehdään sekalaisessa järjestyksessä, ja osa valmistuu toisia aiemmin. Välissä tehdään myös muita, opinnäytetyöhön liittymättömiä, työtehtäviä.

1.1 Lyhenteet ja määritelmät

YSAO = Ylä-Savon ammattiopisto

OLAP = Online analytical processing

OLTP = Online transaction processing

ETL = Extraction, transformation & loading

SSIS = SQL Server integration services

SSAS = SQL Server analysis services

DW = Data warehouse, tietovarasto

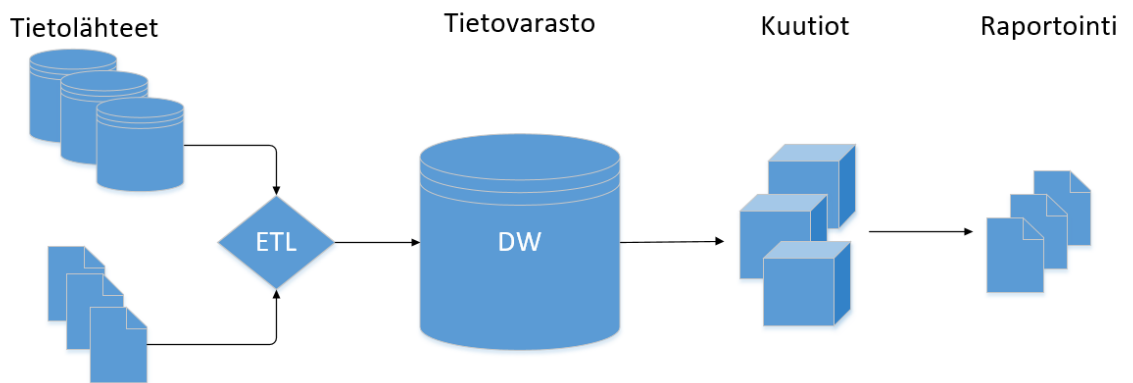
KPI = Key Performance Indicator, mittari.

BSC = Balanced Scorecard, tuloskortti, suorituskykymittaristo. Tavoitejohtamisen työkalu.

SCCM = Microsoft System Center Configuration Manager

2 TIETOVARASTOT JA OLAP

Tietovarasto on iso yhtenäinen tietokanta, johon kootaan tietoa useista erillisistä lähdejärjestelmistä. Yleensä tietovarasto on toteutettu relaatiomallin mukaisen tietokannan päälle, johon voidaan suorittaa kyselyitä SQL –kielellä. Varastossa olevat tiedot pidetään ajan tasalla siirtämällä uutta ja muuttunutta tietoa säännöllisesti lähdejärjestelmistä ETL –prosessin avulla. Tietovaraston rakenne on suunniteltu raportointia varten ja se helpottaa kyselyitä ja tiedon yhdistelyä. Seuraavassa kuvassa on kuvattuna tietovarastojärjestelmän yleinen toimintaperiaate.



Kuva 1 Tietovaraston peruseräite. (Kallio 2019)

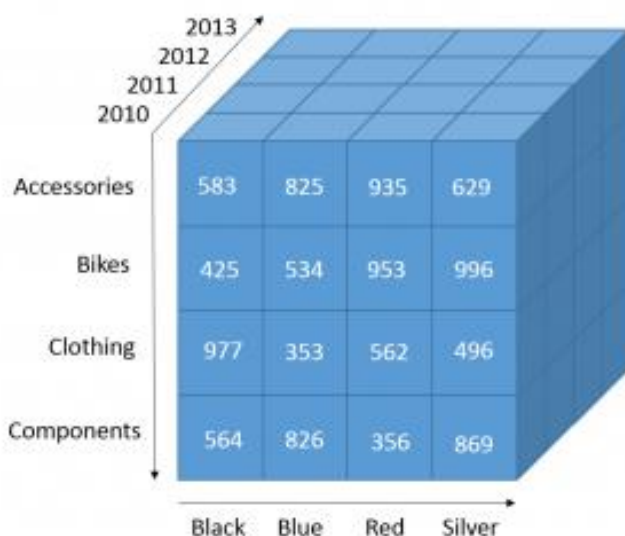
Kuvassa vasemmalla on tietolähteet, joista tietoa haetaan tietovarastoon. Tietolähteet ovat yleensä organisaation operatiivisia järjestelmiä, joissa syntyy tietoa liiketoiminnasta. Monia muunkin tyyppisiä tietolähteitä voidaan hyödyntää, kuten avointa dataa tai yksittäisiä tiedostoja. Esimerkiksi staattinen organisaatorakenne voitaisiin ladata erillisestä tiedostosta tai avointa säädettä voitaisiin viedä jostain ulkopuolisesta järjestelmästä tietovarastoon.

Tietolähteiden tieto siirretään ETL –prosessin avulla tietovarastoon. ETL –prosessi huolehtii tiedon hakemisesta ja lataamisesta tietovarastoon. ETL –prosessin aikana tietoa jalostetaan, tietotyypit muutetaan vastaamaan tiedon sisältöä ja tiedon rakenne yhtenäistetään vastaamaan tietovaraston rakennetta riippumatta alkuperäisestä järjestelmästä ja sen tietorakenteista. Kun tieto on siirretty tietovarastoon, tietovaraston päälle muodostetaan yleensä tietokuutioita tai erillisiä data martteja helpottamaan raportointia ja tiedon saatavuutta.

2.1 Tietokuutiot ja OLAP

OLAP on lyhenne sanoista Online Analytical Processing. Tietovarastojen päälle rakennetaan usein OLAP –tietokantoja helpottamaan raportointia. Jos OLAP –kanta verrataan operatiivisissa järjestelmissä yleisesti käytettyihin OLTP –kantoihin, OLAP on suunniteltu suoritumaan nopeasti raportoinnissa tarvittavista monimutkaisista analyttisistä kyselyistä, kun taas OLTP –kannat pyrkivät suorittamaan mahdollisimman paljon erityyppisiä yksinkertaisia kyselyitä. OLAP –kantojen päivittäminen on

hidasta, mutta se voidaan automatisoida tapahtumaan silloin kun palvelimen käyttö on mahdollisimman vähäistä, esimerkiksi öisin. OLAP –kannoista puhutaan myös monesti OLAP –kuutioina, koska niitä voidaan havainnollistaa kuution avulla. Seuraavassa kuvassa on havainnollistettu yksinkertaista OLAP-kantaa.



Kuva 2 Esimerkki OLAP -kuutiosta. (Lutz 2018)

Kuvan kuutio on havainnollistus myyntidataa sisältävästä OLAP –tietokannasta. Kuutio koostuu pienemmistä kuutioista, jotka sisältävät faktatietoa, jota halutaan tarkastella. Myyntiä mitataan yleensä kappale- tai euro-määräisesti, joten tämä on valittu kuution faktatiedoksi. Faktatiedon lisäksi kuutiossa on yleensä useita dimensioita, joiden mukaan faktatietoa voidaan jaotella. Tämän kuution dimensioina ovat aika (vuosi), tuotekategoria ja tuotteen väri. Pienemmät faktatietoa sisältävät kuutiot on järjestetty dimensioiden mukaan ja niihin on laskettu valmiiksi erilaisia tunnuslukuja dimensioiden mukaan, esimerkiksi summia ja keskiarvoja. Seuraavassa taulukossa on sama OLAP-kanta esitettyä tauluna.

Taulukko 1: Esimerkki OLAP -tietokannan taulusta. (Kallio 2020)

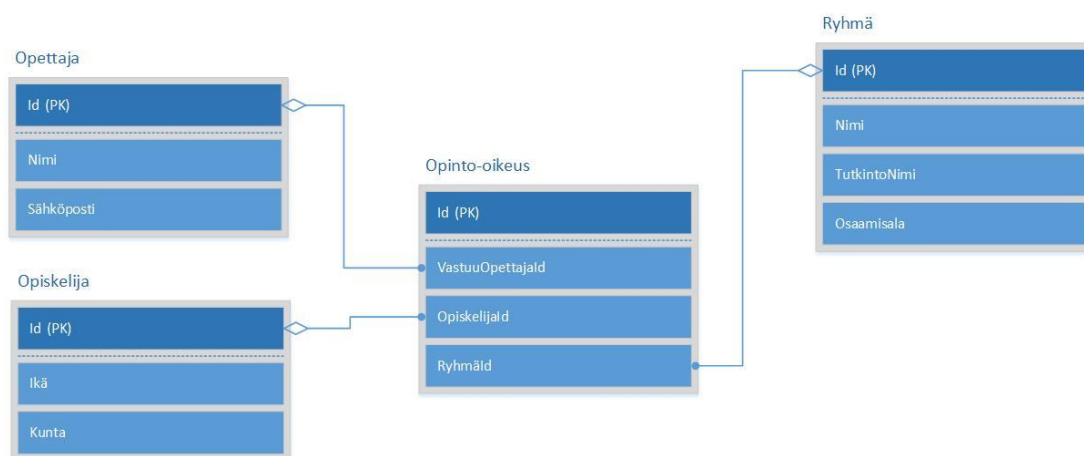
<u>Aika</u>	<u>Tuotekategoria</u>	<u>Väri</u>	<u>Summa</u>
2018	Accessories	Blue	825.00 €
2018	Accessories	Red	935.00 €
2018	Accessories	NULL	1,760.00 €
2018	Clothing	Blue	353.00 €
2018	Clothing	Red	562.00 €
2018	Clothing	NULL	2,388.00 €
2018	NULL	Blue	1,178.00 €
2018	NULL	Red	1,497.00 €
2018	NULL	NULL	10,883.00 €

Taulussa on jokaista dimensiota ja faktaa vastaavat sarakkeet ja osalle riveistä on laskettu valmiiksi summia. Jos taulusta haluttaisiin lukea esimerkiksi vuoden 2018 koko myynti, etsitään vain sellainen rivi, jolla tuotekategoria ja väri sarakkeissa on arvona NULL, siis esimerkki taulukossa viimeisenä

oleva rivi. Tämän rivin faktasarakkeesta voidaan lukea arvo suoraan, jolloin säästytään laskennalta, kun summa löytyy valmiina.

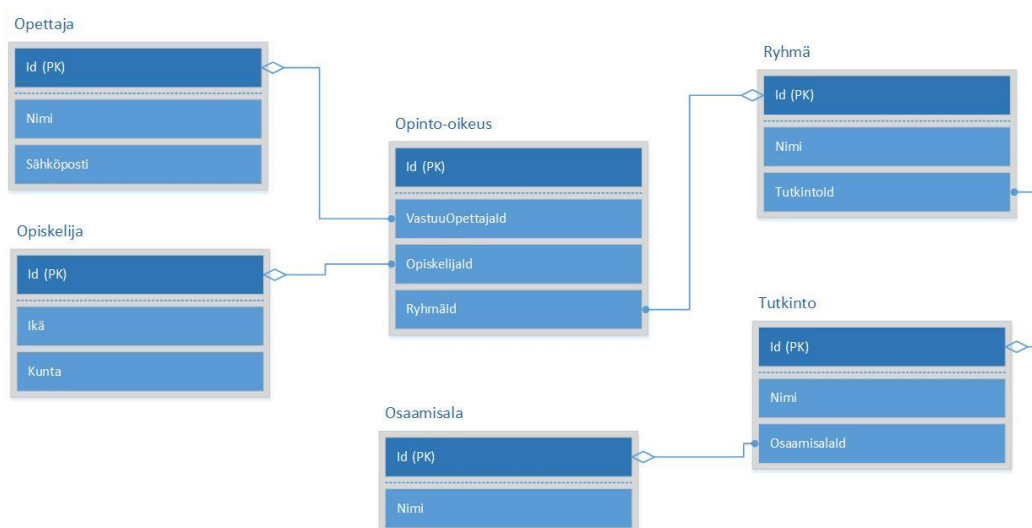
2.2 Tietomallit

Yleensä tiedon rakenne tietovarastossa on erilainen kuin operatiivisissa lähdejärjestelmissä. Tietovaraston tietomalli suunnitellaan vastaamaan raportoinnin tarpeisiin. Yksinkertaisimmillaan voidaan käyttää ns. tähtimallia. Seuraavassa kuvassa esimerkki tämän mallisesta tietokannasta.



Kuva 3 Luonnos tähtimallin tietokannasta. (Kallio 2020)

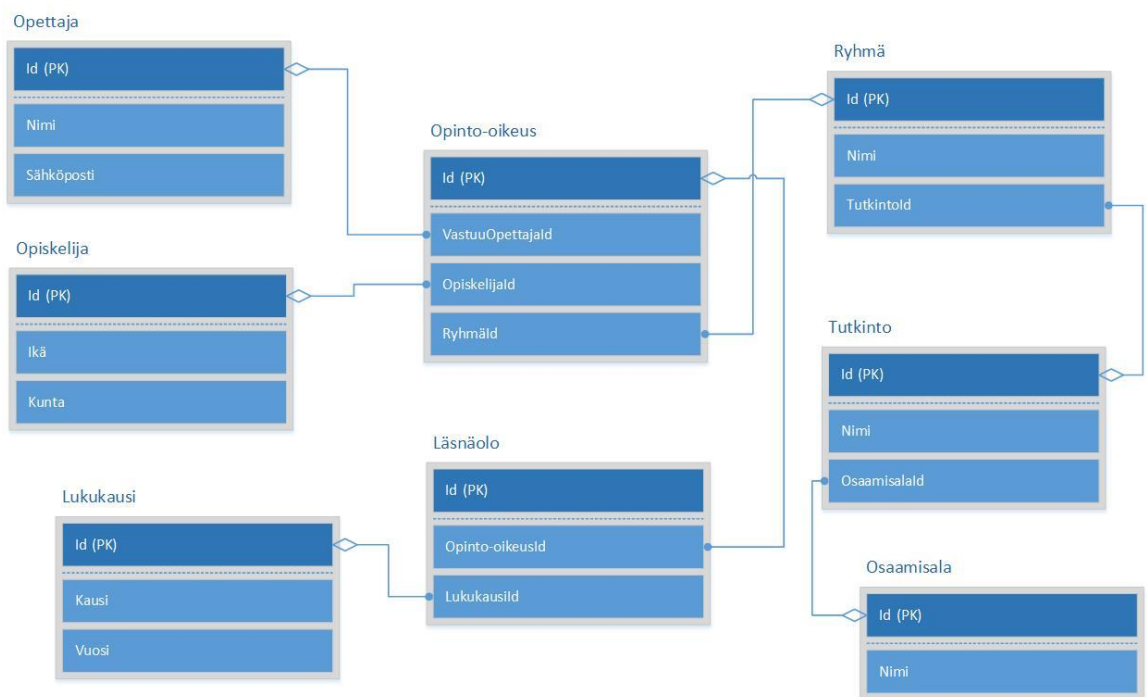
Tähtimallisessa kannassa faktatiedon sisältävä taulu on keskellä ja siihen on liitetty jokaista dimensiota kohden yksi taulu. Tällaisen tietokannan pohjalta on hyvin helppo muodostaa OLAP-kuutioita, jolloin se sopii hyvin tietovarastossa käytettäväksi. Tämän mallisessa tietokannassa dimensiot kuitenkin sisältävät päällekkäistä tietoa, joka kasvattaa tarpeettomasti tallennustilan tarvetta. Esimerkin Ryhmä -dimensiossa todennäköisesti toistuu samat tutkinnot useasti, sillä samaa tutkintoa opiskelevia ryhmiä voi olla useita. Päällekkäisen tiedon määrää voidaan vähentää normalisoimalla dimensio-taulut, tällöin tietokanta muuttuu tähtimallisesta lumihuutalemallin mukaiseksi.



Kuva 4 Lumihuutalemallin mukainen tietokanta. (Kallio 2020)

Tässä Ryhmä dimensioon kuuluneet tutkinto ja osaamisala tiedot on normalisoitu jakamalla tieto useampiin tauluihin. Lumihiutalemallin etuja tähtimalliin nähden on tiedon helpompi ylläpito ja sääs-
töt tallennustilassa. Lumihiutalemallisen tietokannan pohjalta on edelleen helppo muodostaa tieto-
kuutioita, joten tähtimallin hyviä puolia ei menetetä.

Tähti- tai lumihiutalemallin mukaista tietokantaa täytyy monesti laajentaa lisäämällä faktatauluja sillä
yhteen 'tähteen' tai 'lumihiutaleeseen' on vaikea sovittaa kaikkia haluttuja käsitteitä. Aiempaan kan-
taan voitaisiin liittää esimerkiksi läsnäolo uutena faktana.



Kuva 5 Uudella faktatiedolla laajennettu tietokanta. (Kallio 2020)

Tällöin opinto-oikeus -taulu toimii samanaikaisesti omana faktatauluna ja läsnäolo -taulun dimen-
siona. Samaan tapaan tietokantaa voidaan laajentaa rajattomasti, kunnes kaikki tarvittavat käsitteet
on saatu sisällytettyä kantaan.

(Tutorials Point 2020)

2.3 Tietovarastoinnin hyödyt

Yrityksillä ja organisaatioilla on yleensä käytössään monia erillisiä tietojärjestelmiä ja tällöin myös
organisaation toiminnan kannalta olennainen tieto on hajautettuna eri tietokannoissa. Tietovarasto
pyrkii keräämään nämä tiedot yhteen paikkaan, josta ne ovat helpommin saatavilla. Tämä parantaa
tiedon saatavuutta ja samalla mahdollistaa eri järjestelmistä peräisin olevan tiedon yhdistelyn. Li-
säksi tietovarastoon on mahdollista säilöä myös historiatieto, mitä operatiiviset järjestelmät eivät
yleensä tee.

Tiedonsiirto tietovarastoon on mahdollista automatisoida, jolloin käytettävissä on aina mahdollisimman ajantasaista tietoa ilman että sitä täytyy manuaalisesti hakea lähdejärjestelmistä. Samoin tietovaraston päälle rakennetut raportit voidaan päivittää automaattisesti.

2.4 Power BI

Rapostointia varten Microsoftilla on tarjolla Power BI –ohjelmisto. Tämä sisältää työpöytäsovelluksen ja selaimella käytettävän Power BI online –alustan, jonka kautta raportteja on helppo jakaa. Raportit rakennetaan työpöytäsovelluksessa, josta ne saadaan jaettua suoraan Power BI online:n puolelle, josta loppukäyttäjät pääsevät tarkastelemaan niitä selaimella ja luomaan itselleen koontinäyttöjä omien tarpeidensa mukaan. Tarvittaessa Power BI mahdollistaa tiedon yhdistelyn useista eri lähteistä olivatpa ne sitten omissa konesaleissa, pilvessä tai avointa dataa. Power BI on myös melko helppokäyttöinen, jolloin käyttäjät voivat luoda myös itse tarvitsemiaan raportteja ns. Self-Service BI -periaatteen mukaisesti. (Sulava Oy, 2019)

Power BI desktopin käyttöönotto vaatii vain kyseisen työpöytäsovelluksen asentamisen. Sovellus voidaan Microsoft -ympäristössä jakaa keskitetysti SCCM:n kautta, jolloin kaikki sitä tarvitsevat voivat asentaa sen itsenäisesti. Jotta Power BI online:n kanssa voidaan käyttää on-premises tietolähteitä, Microsoftilta on tarjolla tietokantapalvelimelle asennettava Power BI on-premises gateway –ohjelma, jonka kautta Power BI online pääsee käsiksi tietokantapalvelimella oleviin SSAS tietokantoihin. Gatewayn asennuksen jälkeen kuutioita voidaan lisätä tietolähteiksi Power BI online:n puolella.

(Microsoft Oy, 2019)

3 TILANNE JA TILAAJAN TARPEET

3.1 Ympäristö, ohjelmistot ja tekniikat

YSAO:lla on käytössään VMwaren hyperkonvergoitu palvelinympäristö. Ympäristössä ajettavat käyttöjärjestelmät ja ohjelmistot ovat pääasiassa Microsoftin tuotteita. Tietovarastopalvelimen käyttöjärjestelmänä on Windows Server 2012 R2, johon on asennettu Microsoft SQL Server 2012. Kehitystyötä varten palvelimella on asennettuna Visual Studio 2013, jota voi käyttää etätyöpöydän kautta. ETL-prosessit ja tietokuutiot on toteutettu SSIS- ja SSAS –työkaluilla, ajantasaisen tiedon siirto lähdejärjestelmistä on automatisoitu SQL Server Agent:n avulla.

Tietovarasto olisi voitu toteuttaa toisenlaiseenkin ympäristöön ja useilla erilaisilla tekniikoilla. Timo Jalkanen (2016) vertailee opinnäytetyössään erilaisia ympäristöjä, mutta koska YSAO:lla oli jo ennestään laajalti Microsoftin tuotteisiin perustuva ympäristö, ja niiden päälle kehitetty tietovarasto, ei ollut muita järkeviä vaihtoehtoja, kuin jatkaa kehitystyötä näiden päälle. Lisäksi tämä säästi merkittävästi työaikaa, sillä suuri osa tiedosta oli jo ennestään saatavilla tietovarastossa. Koska Microsoft SQL Server on laajalti käytössä, myös tukimateriaalia on saatavilla runsaasti ongelmatilanteiden varalta.

3.2 Ylä-Savon ammattiopiston tietovaraston tilanne

YSAO:lla oli ennestään käytössä TCD Consulting and Research Oy:n, lyhennettynä TCD, kehittämä tietovarasto ja sen päällä toimivia selainpohjaisia suunnittelu ja BSC-työkaluja. Lisäksi TCD on toteuttanut tietovarastoon muutamia kuutioita ja kouluttanut niiden käyttöä henkilökunnalle. Kuutiot eivät kuitenkaan ole päätyneet käyttöön ja ne ovat rakenteeltaan melko vaikeakäyttöisiä. Selainpohjainen ohjelmistokokonaisuus sen sijaan on organisaatiossa laajalti käytössä ja sen jatkokehitystä tehdään jatkuvasti yhteistyössä TCD:n kanssa.

3.3 Primus

Opinnäytetyön kannalta olennaisin lähdejärjestelmä on YSAO:n opiskelijahallintojärjestelmä Primus, joka pyörii saman ympäristön toisella palvelimella. Primuksessa tieto on tallennettuna rekistereihin, jotka vastaavat periaatteeltaan SQL –tietokannan tauluja. Rekistereissä tieto on tallennettu kortteihin, jotka vastaavat SQL –kannan rivejä. Jokaisella kortilla on rekisterin sisällä uniikki korttinumero, jolla kyseiseen korttiin voi tarvittaessa viitata. Primuksesta on mahdollista hakea tietoa PrimusQuery –työkalun avulla. PrimusQuery:ä voidaan käyttää myös komentoriviltä, joka mahdollistaa siirtojen automatisoinnin helposti.

3.4 Raportointitarpeet

Olemassa olevien työkalujen lisäksi halutaan kehittää erityisesti tietovarastosta saatavaa raportointi-hyötyä. Nykyinen tuloskortti antaa hyvän yleisnäkymän toiminnasta ja mahdollistaa porautumisen eri

organisaatiotasoinne, mutta asiakas haluaa lisäksi tarkempaa tietoa toiminnan eri osa-alueilta visuaalisten raporttien muodossa.

Yksi keskeinen ongelma on ollut opintonsa keskeyttävät opiskelijat. Ongelman ratkaisemisen tueksi halutaan raportteja, joista nähtäisiin tarkemmin missä mahdolliset ongelmakohdat ovat, jotta tiedettäisiin, minne resursseja kannattaa kohdistaa.

Toinen merkittävä kokonaisuus, jota halutaan saada raportoitua, on opiskelijavuosien ja suoritteiden toteumat organisaatorakenteen mukaisesti. Ammatillisen koulutuksen rahoitus on uudistumassa siten, että jatkossa toiminnan tehokkuus ja vaikuttavuus korostuu aiemman vahvasti opiskelijavolyymiin pohjautuneen mallin sijasta. Vastaavaa rahoitusmallia halutaan jatkossa käyttää myös sisäisen rahoituksen jaossa ja tämän perusteeksi tarvitaan kyseisiä tietoja. Lisäksi raporttiin halutaan liittää suorituskykykymittareita.

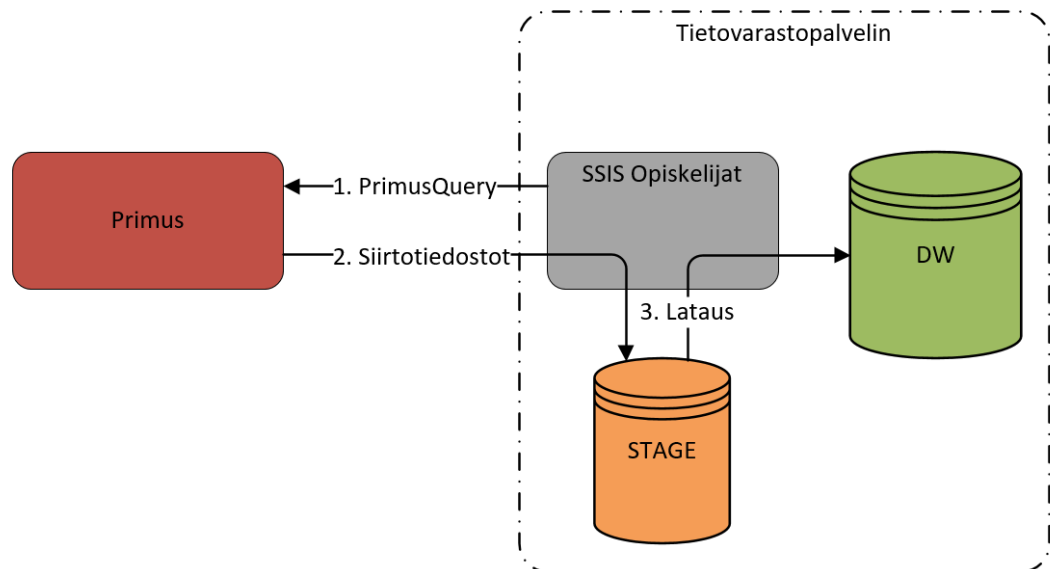
Projektin aikana nousi esiin monia muitakin pienempia raportointikohteita, joita sivutaan tässä opinäytetyössä. Pääpaino pidetään kuitenkin näissä kahdessa merkittävimmissä kokonaisuudessa.

4 KEHITYSTYÖ

Asiakkaan raportointitarpeita ratkaisemiseksi tietovarastoa täytyi laajentaa puuttuvien tietojen osalta. Lisäksi kuutioita täytyisi muokata tai luoda uusia, jotta halutut raportit voitaisiin niiden pohjalta muodostaa. Lopuksi täytyi ratkaista, miten raportit saadaan jaettua tarvittaville henkilöille ja toisaalta, kuinka käyttöoikeuksia hallitaan niin, ettei asiattomat henkilöt pääse käsiksi sensitiiviseen tietoon.

4.1 Integraatit

Integraatit, eli tiedonsiirrot lähdejärjestelmä Primuksesta tietovarastoon, on toteutettu SSIS –paketteina. SSIS –paketit huolehtivat koko ETL –prosessin ajamisesta ja SQL Server Agentin avulla voidaan ajastaa kyseiset paketit ajettavaksi halutulla syklillä. Tätä prosessia muokkaamalla Primuksesta saadaan tuotua tarvittavat lisäkentät. Kuvassa 1 on havainnollistettuna prosessin päävaiheet.



Kuva 6 Tiedon kulku eri järjestelmien välillä ETL-prosessin aikana. (Kallio 2019)

Ensimmäisessä vaiheessa tieto haetaan Primuksesta PrimusQuery –kyselyiden avulla. Seuraavaksi generoidut siirtotiedostot luetaan sellaisenaan tietovarastopalvelimen STAGE –tietokantaan. Lopuksi tieto ladataan tietovaraston varsinaiseen DW –kantaan, samalla tietotyypit muutetaan merkkijonoista vastaamaan tiedon sisältöä ja tehdään muita tarvittavia muunnoksia.

4.1.1 PrimusQuery

Primuksessa oleviin tietoihin pääsee käsiksi PrimusQuery –työkalun avulla. PrimusQuery mahdollistaa kyselyt suoraan Primuksen tietokantoihin erillisen kyselykielen avulla. Kysely tuottaa tekstitiedoston, jonka muoto voidaan määrittellä vapaasti kyselyssä, tiedosto voi olla esimerkiksi CSV tai HTML muotoinen. PrimusQuerylla on versiosta 2.2 alkaen ollut mahdollista myös tallentaa tietoa Primuksen tietokantoihin.

Windows ympäristössä PrimusQuery vaatii primusquery.exe –tiedoston kyselyä tekeväälle laitteelle. Lisäksi haluttu kysely pitää tallentaa erilliseen tekstitiedostoon. Ohessa yksi esimerkkikysely, jossa palvelimen ja käyttäjän tiedot ovat keksittyjä.

```
#HOST 10.20.30.40
#PORT 1234
#USER Username
#PASS Salasana12345
#OUTPUT tulokset\osaamisalat.txt
#CHARSET UTF-8
#SORT V1
#DATABASE osaamalat
#HEADER_START
ID";"Arkistoitu";"Nimi";"Koodi";"Kustannusryhma
#HEADER_STOP
#DATA{V1}";"#DATA{K1}";"#DATA{K2}";"#DATA{K3}";"#DATA{K23}
```

Kuva 7 Esimerkki PrimusQuery –kyselystä. (Kallio 2019)

Kyselyn alussa määritellään Primus –palvelimen tiedot, käyttäjätunnus, salasana, mihin tietokantaan kysely tehdään ja minne tulokset tallennetaan. Lopussa määritellään minkä muotoinen tulostiedostosta tehdään ja mitä kenttiä siihen haetaan. Primuksen tietokannan kenttiin viitataan kenttänumeroilla K. Avaimena käytetään vakioista löytyvää korttinumeroa V1. Lisäksi voidaan käyttää myös Primuksen funktioita, joihin viitataan kirjaimella F.

Tässä tapauksessa luodaan CSV:n tapainen tiedosto, jossa kentät on eroteltu käyttäen lainausmerkeissä olevaa puolipistettä. Samaa formaattia on käytetty opinnäytetyön kaikkien kyselyjen kohdalla. Tämän jälkeen kysely voidaan ajaa antamalla komentorivillä seuraava komento.

primusquery kyselytiedosto.txt

Komento ajaa kyselyn Primuksen tietokantaan ja tallentaa tulokset halutussa formaatissa tulostiedostoon. Edellinen kysely tuottaa seuraavanlaisen tulostiedoston.

```
ID";"Arkistoitu";"Nimi";"Koodi";"Kustannusryhma
1";"Kyllä";"Sähkö- ja energiatekniikan osaamisala";"1534";"0
2";"Kyllä";"Elektroniikan ja tietoliikennetekniikan osaamisala";"1533";"0
3";"Kyllä";"Automaatiotekniikan ja kunnossapidon osaamisala";"1522";"0
4";"Kyllä";"Ruoka- ja asiakaspalvelun osaamisala";"1570";"0
5";"Kyllä";"Teollisuuspuusepän osaamisala";"1536";"0
6";"Kyllä";"Maatilatalouden osaamisala";"1580";"0
7";"Kyllä";"Hevostalouden osaamisala";"1581";"0
8";"Kyllä";"Ravintolapalvelun osaamisala";"1574";"0
```

Kuva 8 Muutama ensimmäinen rivi PrimusQueryn tulostiedostosta. (Kallio 2019)

Tämän muotoisia tekstitiedostoja on myöhemmin helppo siirtää SSIS –paketin avulla SQL tietokantaan.

Komentorivin lisäksi PrimusQuerya voidaan käyttää myös skriptien kautta. Skripteillä on helppo automatisoida kyselyjen ajot osaksi SSIS –pakettia. Tätä varten on tehty seuraava .bat -skripti.

```
@ECHO off
title PrimusQuery
set "dir1=D:\PROJECT\PRIMUS"

cls
echo PRIMUSQUERY HAKU
echo -----
del tulokset\*.txt

FOR %%X in ("%dir1%\*_lahde.txt") DO (
    echo Ajetaan tiedostoa %%~nX.txt ..
    primusquery %%~nX.txt
    if %ERRORLEVEL% GEQ 1 EXIT /B 1
)
echo -----
echo VALMIS
```

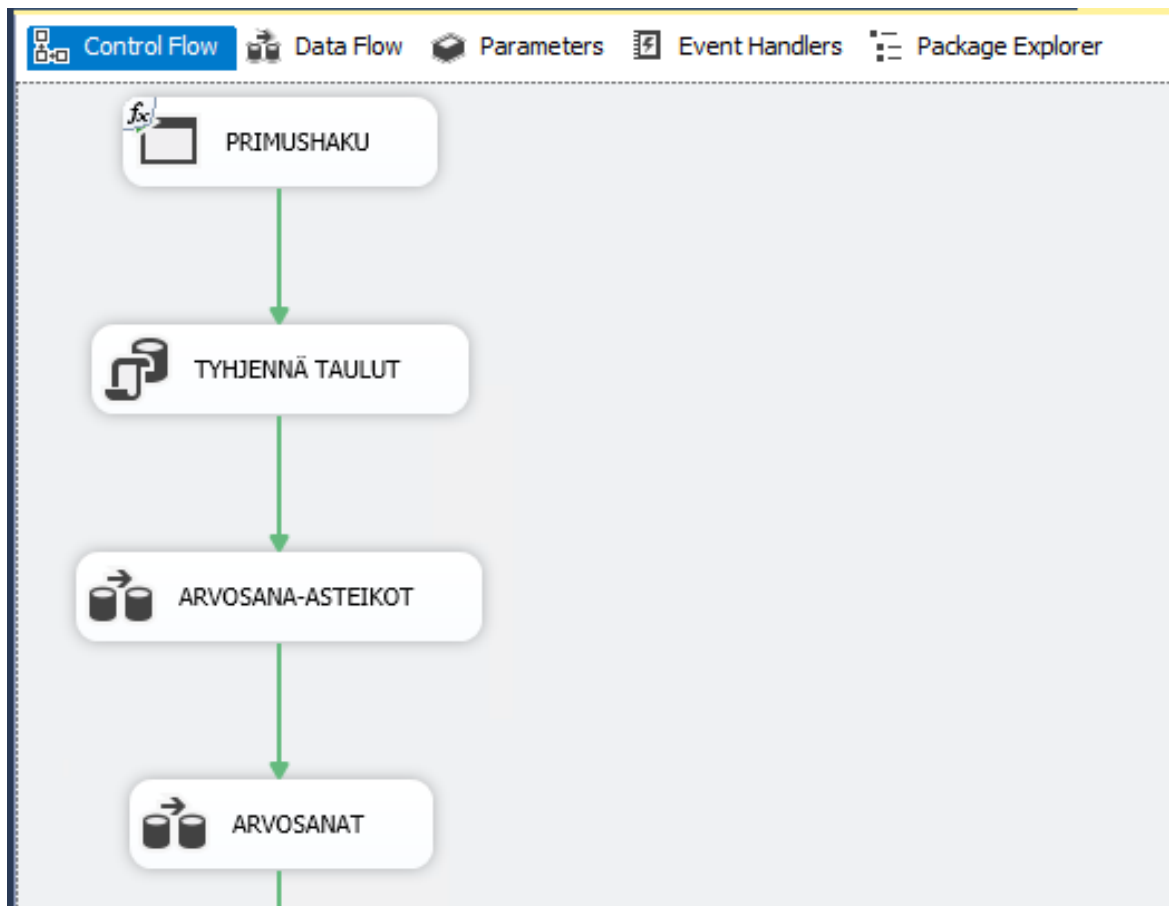
Kuva 9 Komentojono Primus -kyselyjen suorittamiseksi. (Kallio 2019)

Skriptillä saadaan ajettua hakemistosta kaikki sellaiset PrimusQuery –kyselyt, joiden tiedostonimi loppuu *_lahde.txt*. Uusia kyselyjä voidaan lisätä kirjoittamalla lisää kyselytiedostoja kyseiseen hakemistoon ja nimeämällä ne saman formaatin mukaan.

(StarSoft Oy, 2014)

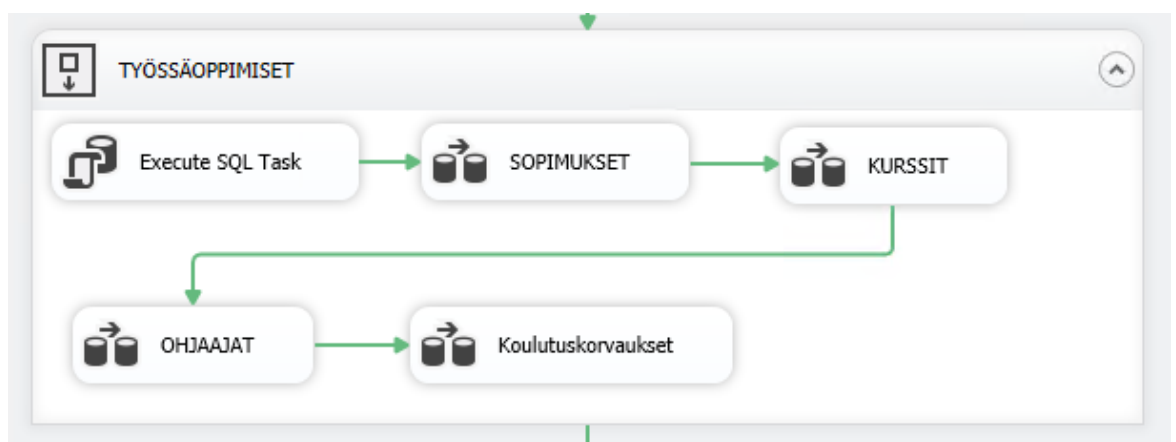
4.1.2 ETL –prosessi

ETL –prosessi vastaa myös tiedon siirtämisestä Primuksen rekistereistä tietovarastokantaan. Prosessin aikana myös tiedon rakenne muutetaan vastaamaan tietovaraston tietomallia. ETL –prosessi on jaettu useisiin erillisiin SSIS –paketteihin, joista jokainen vastaa tietyistä ETL –prosessin osista tai tietokokonaisuudesta. SSIS –paketit tehdään Visual Studiossa graafisesti, samalla syntyy kaavio, jota voidaan käyttää paketin dokumentaationa. SSIS –paketti jakautuu Control Flow ja Data Flow –osiin. Control Flow:ssa määritellään paketin suorituksen vaiheet ja niiden järjestys. Control Flow:ssa määritetään yleensä yksi tai useita Data Flow –tehtäviä, jotka vastaavat tiedon siirtämisestä ja käsitteilystä.



Kuva 10 Ensimmäisen SSIS -paketin Control Flow -osion alku. (Kallio 2020)

ETL –prosessin ensimmäisessä vaiheessa suoritetaan PrimusQuery –kyselyt ja tyhjennetään STAGE –tietokannan taulut. Tämän jälkeen on lukuisia Data Flow -tehtäviä, jotka siirtävät Primuksen rekistereistä haetut kortit STAGE –tietokannan tauluihin. Kun kaikki rekisterit on siirretty STAGE –kantaan, siirrytään ETL –prosessin seuraavaan vaiheeseen. Seuraavassa kuvassa esimerkki ETL –prosessin seuraavasta vaiheesta.



Kuva 11 Control Flow työssäoppimisten siirtämiseksi STAGE –kannasta tietovarastokantaan. (Kallio 2020)

Opiskelijatietojen osalta TCD oli toteuttanut tietovarastoon yhden kuution, joka pyrki sisältämään kaikki tarvittavat faktat ja dimensiot. Tätä olisi voitu lähteä laajentamaan yhden kuution periaatteen mukaisesti, mutta tässä projektissa päädyttiin aloittamaan kehittäminen puhtaalta pöydältä ja luoda useita yksittäisiä kuutioita vastaamaan eri raportointitarpeita. Käyttäjät olivat kokeneet TCD:n luoman kuution liian monimutkaisena eikä sitä ollut otettu käyttöön. Usean kuution ratkaisulla haluttiin tehdä tietovarastosta helpommin lähestyttävä, jotta käyttäjät ottaisivat sen laajemmin käyttöön kuin mitä tähän mennessä. Jokaisesta kuutiosta päätettiin kuitenkin tehdä tarvittavan laaja, jotta vältytään ristiinkyselyn tarpeelta vaikkakin kääntöpuolena jouduttiin tekemään päällekkäistä työtä aina kun samoja dimensioita tai faktoja tarvittiin useissa erillisissä kuutioissa.

Halutut tietokuutiot voidaan toteuttaa Microsoft ympäristössä Visual Studiolla SSAS -työkalujen avulla. SSAS-projekti voi sisältää mm. yhteysasetukset kuution pohjana käytettäviin tietolähteisiin, laskentakaavat kuution muodostamiseen ja roolit, joiden avulla käyttöoikeudet määritellään.

4.2.1 Opintojen etenemisen seuranta

Opiskelijoiden opintojen etenemisen seuranta haluttiin saada raportoitua organisaatorakenteen mukaisesti ja tätä varten luotiin tietokanta SSAS Suoritukset, joka sisältää yhden tietokuution. Kuution pääasiallisena faktatietona on kertyneet opiskelijavuodet, tutkinnot ja osaamispisteet. Tärkeimpinä dimensioina on mm. organisaatorakenne, aika, tutkinto ja tutkintotyyppi. Lisäksi kuutioon lasketaan KPI-mittarit opiskelijavuosista ja tutkinnoista ja näitä varten tarvittiin myös tavoitearvot kummallekin. Excelillä voidaan kuution tiedoista muodostaa seuraavanlainen raportti.

Pvm.Hierarchy		Kalenteri 2019					
OPV 2019							
Riviotsikot	OPV Toteuma	OPV Ennuste	OPV Tavoite	OPV Status	OPV Trendi	OPV 2018	
Ylä-Savon koulutuskuntayhtymä	1548	1548	1632	🔴	📉	1552	
Kehittämispalvelut	165	165	160	🟢	📈	149	
Kehittämispalvelut	165	165	160	🟢	📈	149	
Osaamispalvelut	1382	1382	1472	🔴	📉	1403	
Logistiikkaosaaminen	285	285	286	🔴	📉	288	
Luonnonvaraosaaminen	367	367	394	🔴	📉	360	
Palveluosaaminen	158	158	191	🔴	📉	161	
Rakentamisaosaaminen	139	139	136	🟢	📈	143	
Teknologiaosaaminen	324	324	303	🟢	📈	303	
Työelämäosaaminen	110	110	162	🔴	📉	148	
Kaikki yhteensä	1548	1548	1632	🔴	📉	1552	

Kuva 13 Excel -raportti opiskelijavuosikertymistä. (Kallio 2020)

Organisaatorakenne -dimension avulla saadaan opiskelijavuosikertymät raportoitua organisaatorakenteen mukaisesti jaoteltuna. Kuutioon liitettiin myös yksinkertainen opiskelijamääriin pohjautuva ennuste tulevista opiskelijavuosista. KPI-mittareilla raportista saatiin helposti hieman visuaalisempi, jolloin on helpompi nähdä nopeasti yleistilanne. Samaan kuutioon koottiin lisäksi opiskelijoiden tutkinnon suoritusajkoja ja näitä voidaan tarkastella samojen dimensioiden avulla kuin muitakin kuution faktatietoja.

4.2.2 Opintojen keskeytykset

Opintonsa keskeyttävien opiskelijoiden tarkastelua ja raportointia varten tehtiin SSAS-projekti SSAS Eronneet. Tämän kuution faktatiedoksi otettiin eronneiden opiskelijoiden lukumäärä. Dimensioiksi valittiin opiskelijaa luokittelevan tiedon lisäksi sellaisia asioita, jotka voisivat selittää erosyitä, esimerkiksi poissaolokertymät.

Lisäksi haluttiin saada raportoitua, miten erityisen tuen piirissä olevien opiskelijoiden tuen saanti toteutuu ja mahdollisesti ehkäistä keskeyttämisiä huolehtimalla erityisen tuen suunnitelmien ajantasaisuutta ja tukitoimien toteutumisesta. Tätä varten luotiin projekti SSAS Erityinen tuki. Tässä faktatiedoksi valittiin lukumäärät erityisen tuen suunnitelmista ja toteutetuista tukitoimista. Dimensioiden avulla saatiin jaoteltua esimerkiksi ajan tasalla olevien suunnitelmien osuus kaikista suunnitelmista.

4.3 Kuutioiden käyttöönotto

Valmiit tietokuutiot eivät vielä itsessään tuo juuri hyötyä organisaatiolle, mutta niitä voidaan hyödyntää erilaisilla raportointityökaluilla. Raportointityökalut mahdollistavat raporttien muodostamisen helposti kuutioiden pohjalta. Pääasialliseksi raportointityökaluksi valitui Microsoftin Power BI, jolla on helppo muodostaa visuaalisia ja interaktiivisia raportteja. Power BI online mahdollistaa raporttien helpon jakamisen niin organisaation sisällä kuin myös ulkopuolisille tahoille. Lisäksi Power BI online toimii selaimessa, jolloin käyttäjien ei tarvitse asentaa ohjelmistoa laitteelleen, vaan he voivat katella raportteja suoraan selaimella.

Kuutioiden käyttöönotto päätettiin toteuttaa ensisijaisesti Self-Service BI -periaatteen mukaisesti, eli käyttäjät muodostivat itse tarvitsemansa raportit kuutioiden pohjalta. Tarvittaessa käyttäjille annettiin tukea Power BI:n käytössä, mutta käyttöönotto sujui käyttäjiltä pääasiassa itsenäisesti ja tarvittavia raportteja saatiin muodostettua. Seuraavassa esimerkki opiskelijavuosien seurantaan keskittyvästä raportista.



Kuva 14 Power BI raportti opiskelijavuosiennustuksista. (Korolainen 2019)

Erilaiset visualisoinnit ja värit auttavat hahmottamaan nopeasti yleistä tilannetta ja liikennevalo väreillä on havainnollistettu tilannetta suhteessa tavoitteisiin ja samoin on esitettyä myös trendi eli tilanne suhteessa edellisen vuoden samaan aikaan verrattuna. Power BI mahdollistaa raporteissa porautumisen myös tarkemmalle tasolle. Raporttiin on liitetty myös hakukenttä ja tarkastelujaksoa voidaan muuttaa sliderin avulla tai syöttämällä halutut päivämäärät niille varattuihin kenttiin. Näin raportin kautta voidaan hakea helposti haluttuja tietoja.

4.4 Käyttöoikeuksien hallinta

Käyttäjien pääsyä kuution tietoihin voidaan hallita SSAS tietokannan roolien kautta. Windows ympäristössä tässä voidaan käyttää AD autentikointia ja määrittellä oikeudet olemassa olevien käyttäjätunusten ja ryhmien kautta. Lisäksi käyttäjille pitää myöntää oikeudet erikseen myös Power BI online:n tietolähteisiin, jossa voidaan käyttää käyttäjien Office365 -tunnuksia.

SSAS tietokannan käyttöoikeudet voidaan määrittää joko SSAS-projektissa tai projektin pohjalta luotun tietokantaan. Tässä oikeudet määriteltiin Visual Studion puolella SSAS-projektiin, jotta oikeudet säilyvät myös silloin, jos tietokantaan tehdään muutoksia. Käyttöoikeudet määritellään tietokannan roolien (Role) kautta. Tietokantaan voidaan luoda tarvittava määrä rooleja, ja jokaiselle roolille myönnetään halutut oikeudet. Tämän jälkeen rooliin voidaan liittää AD-käyttäjiä ja ryhmiä, joiden kautta roolin oikeudet kohdistuvat halutuille käyttäjille.

Lähtökohtaisesti SSAS tietokantoihin luotiin vain yksi rooli, jolle annettiin täydet oikeudet kuution tietoihin. Tietokannan roolien kautta käyttäjän pääsyä tietoihin voitiin myös rajata. Tarvittaessa osalle käyttäjiä paljastettiin kuutiosta vain tietyt dimensiot tai dimensioista vain tietyt attribuutit tai

niiden jäsenet. Tällöin yhteen kuutioon voitiin liittää kaikki oleelliset tiedot, mutta pitää arkaluontoiset kentät piilossa sellaisilta käyttäjiltä, joilla ei niihin kuulunut olla pääsyä.

(Microsoft Oy, 2018)

5 TULOSTEN ARVIOINTI JA SAADUT PALAUTTEET

Kehitystyön jälkeen kerättiin palautetta henkilöstöltä. Palaute oli positiivista, raportit oli otettu käyttöön ja ne koettiin hyödyllisinä. Raportit mahdollistivat haluttujen asioiden tarkastelun ja seurannan, joka ei aiemmin olisi ollut mahdollista. Raportteihin saatiin toteutettua hyvin kaikki halutut toiminnallisuudet ja visualisoinnit.

Power BI:n käyttöönotto on helpottanut huomattavasti tietovaraston hyödyntämistä, jolloin suurempi osa henkilöstöstä on huomannut tietovaraston hyödyllisyyden. Lisäksi raporttien katselu ja jakaminen on helpottunut huomattavasti. Joitain raportteja olisi haluttu jakaa myös ulkopuolisille tahoille, mutta nykyisin käytössä oleva lisenssi ei mahdollista tätä. Lisenssin muuttaminen on kuitenkin mahdollista, mikäli tälle toiminnallisuudelle tulee enemmänkin kysyntää.

6 YHTEENVETO

Projektin toteutus onnistui asiakkaan toiveiden mukaisesti. Esille nousseet kehitystyöt saatiin toteutettua ja kehitystyön tulokset otettiin organisaatiossa hyvin käyttöön. Toteutetut kehitystyöt myös koettiin hyödyllisinä organisaatiossa. Tästä huolimatta tietovaraston ei voida sanoa olevan valmis vaan uusia ideoita ja tarpeita nousee esiin tasaiseen tahtiin. Myös esimerkiksi muuttuva lainsäädäntö tuo uusia velvoitteita ulkoiseen raportointiin oppilaitosympäristössä ja näihin pyritään vastaamaan mahdollisuuksien mukaan jatkossakin tietovaraston avulla.

Tietovarastossa on paljon hyödyntämätöntä potentiaalia organisaation toiminnan eri osa-alueilta. Opinnäytetyöprosessin aikana keskityttiin opiskelijatietoihin, mutta esimerkiksi talous- tai henkilöstötiedoista voitaisiin toteuttaa hyvin vastaavia kehitystyitä. Kevään 2020 aikana tullaankin kehittämään mm. talouspuolen raportointia yhdessä kyseisen alan henkilöstön kanssa. Tietovarastosta saatavaa hyötyä voisi kasvattaa myös ottamalla käyttöön tiedon automaattista analytiikkaa esimerkiksi tiedonlouhinnan keinoin, kun toistaiseksi on keskitytty vain perinteiseen raportointiin.

7 LÄHDELUETTELO

Tutorials Point. Data Warehousing Tutorial / DWH – Schemas. Haettu 23. Huhtikuu 2020 osoitteesta <https://www.tutorialspoint.com>: https://www.tutorialspoint.com/dwh/dwh_schemas.htm

Sulava Oy. (15. Tammikuu 2019) *Power BI – Kaikki mitä sinun tulee tietää aloittaaksesi*. Haettu 30. Marraskuu 2019 osoitteesta <https://www.sulava.com>: <https://www.sulava.com/power-bi-kaikki-mita-sinun-tulee-tietaa-aloittaaksesi>

Microsoft Oy. (15. Heinäkuu 2019) *What is an on-premises data gateway?* Haettu 1. Joulukuu 2019 osoitteesta <https://docs.microsoft.com>: <https://docs.microsoft.com/en-us/power-bi/service-gateway-onprem>

Jalkanen T. (Helmikuu 2016) *Tietovarastointi- ja Business Intelligence -ympäristöjen vertailu*. Karelia-ammattikorkeakoulu. Tietojenkäsittelyn koulutusohjelma. Opinnäytetyö. Haettu 1. Joulukuuta 2019 osoitteesta: <https://www.theseus.fi/>: https://www.theseus.fi/bitstream/handle/10024/106335/Jalkanen_Timo.pdf?sequence=1&isAllowed=y

StarSoft Oy (3. Kesäkuu 2014) *PrimusQuery V2.2.3 -ohje*

Serra J. (15. Tammikuu 2013) *SSAS: One giant cube or many small ones?* Haettu 1. Joulukuu 2019 osoitteesta: <http://www.jamesserra.com/>: <http://www.jamesserra.com/archive/2013/01/ssas-one-giant-cube-or-many-small-ones>

Microsoft Oy. (5. Helmikuu 2018) *Grant cube or model permissions (Analysis Services)*. Haettu 1. Joulukuu 2019 osoitteesta <https://docs.microsoft.com>: <https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/grant-cube-or-model-permissions-analysis-services>

Kuva 1. Kaavio tietovaraston toimintaperiaatteesta. Kallio E. (9. Tammikuu 2019)

Kuva 2. Havainnollistava kuva OLAP-kuutiosta. Lutz G. (17. Joulukuu 2018) Haettu 9. Tammikuu 2019 osoitteesta: <https://www.grapacity.com/blogs/working-with-olap-cubes>

Kuva 3: Tietokantakaavio. Kallio E. (20. Huhtikuu 2020)

Kuva 4: Tietokantakaavio. Kallio E. (20. Huhtikuu 2020)

Kuva 5: Tietokantakaavio. Kallio E. (20. Huhtikuu 2020)

Kuva 6. Kaavio YSAO:n tietovaraston ETL-prosessista. Kallio E. (2. Joulukuu 2019)

Kuva 7. Kuvakaappaus PrimusQuery -kyselystä. Kallio E. (2. Joulukuu 2019)

Kuva 8. Kuvakaappaus PrimusQueryn tulostiedostosta. Kallio E. (2. Joulukuu 2019)

Kuva 9. Kuvakaappaus Windows komentojonosta. Kallio E. (2. Joulukuu 2019)

Kuva 10. Kuvakaappaus SSIS -projektista. Kallio E. (24. Tammikuu 2020)

Kuva 11. Kuvakaappaus SSIS -projektista. Kallio E. (24. Tammikuu 2020)

Kuva 12. Kuvakaappaus SSIS -projektista. Kallio E. (24. Tammikuu 2020)

Kuva 13. Kuvakaappaus Excel -raportista. Kallio E. (10. Huhtikuu 2020)

Kuva 14. Kuvakaappaus Power BI -raportista. Korolainen A. (11. Joulukuu 2019)

Taulukko 1. Esimerkki rivejä OLAP-kannan taulusta. Kallio E. (20. Huhtikuu 2020)