

Avaimenhakuprosessin optimointi

LAHDEN AMMATTIKORKEAKOULU

Tietotekniikan koulutusohjelma

Ohjelmistotekniikka

Opinnäytetyö

Kevät 2008

Ari Rämö

Lahden ammattikorkeakoulu

Tietotekniikan koulutusohjelma

RÄMÖ, ARI: Avaimenhakuprosessin optimointi

Ohjelmistotekniikan opinnäytetyö, 35 sivua

Kevät 2008

TIIVISTELMÄ

Opinnäytetyön aiheena oli SSIS 2005 -avaimenhaun yksinkertaistaminen. Avaimenhaku tehdään usealla SSIS 2005 -komponentilla eBIC-konseptituotteen ETL -prosessissa. Koska usean SSIS 2005 -komponentin käyttäminen avaimenhaussa on vaikeaa, piti avaimenhakua yksinkertaistaa. Avaimenhaun yksinkertaistamisen tarkoituksena oli pienentää avaimenhaun toteuttamiseen tarvittavaa työmäärää.

Tutkimusongelmana oli avaimenhaun yksinkertaistaminen SSIS 2005 -ohjelmassa. Tutkimusosassa tutkittiin avaimenhaun toiminnallisuutta ja avaimenhakukomponentin käyttöliittymää.

Työ toteutettiin itse tehdyllä SSIS 2005 -tietovuokomponentilla. Tietovuokomponentissa on graafinen käyttöliittymä, jolla käyttäjä voi määritellä avaimenhaun. Tarvittaessa käyttäjä voi määritellä useita avaimenhakuja samassa komponentissa. Toteutettu avaimenhakukomponentti toteuttaa samat toiminnallisuudet kuin valmiina saatavien SSIS 2005 komponenttien yhteiskäyttö.

Avaimenhaun yksinkertaistaminen onnistui. Käyttäjät ovat antaneet positiivista palautetta ja ottaneet toteutetun avaimenhakukomponentin käyttöön.

Asiasanat: SSIS 2005, avaimenhaku, eBIC

Lahti University of Applied Sciences
Faculty of Technology
RÄMÖ, ARI: Optimizing lookup

Bachelor's Thesis of Software engineering, 35 pages

Spring 2008

ABSTRACT

The objective of this thesis was simplifying SSIS 2005 lookup. Lookup has been done with multiple SSIS 2005 components in LogicaCMG's ETL process of the eBIC concept product. Because lookup with multiple components is difficult, it needs to be simplified. The purpose of simplifying lookup is to reduce the amount required in lookup.

The aim of the theory was to examine how to simplify lookup in SSIS 2005. The functionality of a lookup and the graphical user interface of the lookup component were studied .

The lookup component was implemented with a custom made SSIS 2005 data flow component. The data flow component has a graphical user interface which the user can use to define lookup. If needed, the user can define multiple lookups in the component. The lookup component has the same functionality as lookup that is done with multiple SSIS 2005 components.

Simplifying the lookup was a success. Users have given positive feedback and taken the component in use.

Keywords: SSIS 2005, lookup, eBIC

SISÄLLYSLUETTELO

1 JOHDANTO	1
2 TOIMINTAYMPÄRISTÖ	3
2.1 eBIC	3
2.2 ETL – prosessi	5
3 SSIS 2005	7
3.1 Yleistä	7
3.2 SSIS 2005 arkkitehtuuri yleisellä tasolla	8
3.2.1 Putket ja suodattimet arkkitehtuuri	8
3.3 SSIS 2005 paketti	9
3.3.1 Tehtävävuot	11
3.3.2 Tietovuot	12
4 AVAIMENHAKU	13
4.1 Yleistä	13
4.2 Haettavat arvot	14
4.2.1 Tunnistearvo	14
4.2.2 Tunnistearvo, jolla ei löydetä surrogaattivainta	16
4.2.3 Puuttuva tunnistearvo	17
5 AVAIMENHAKU KOMPONENTIN SUUNNITTELU	19
5.1 Yleistä	19
5.2 Komponentin toiminnalliset vaatimukset	20
5.3 Komponentin rakenne	21
5.4 Komponentin käyttöliittymä	23
6 AVAIMENHAUN TOTEUTUS SSIS 2005 PAKETISSA	28
6.1 Avaimenhaun toteutus SSIS 2005 komponenteilla	28
6.1.1 Ohjausvuon komponentit	28
6.1.2 Tietovuon komponentit	30
6.2 Avaimenhaun toteutus itse tehdyllä tietovuokomponentilla	32
7 YHTEENVETO	34

LYHENNELUETTELO

Avain	Sarake tai sarakekokelma tietokannassa, jolla voidaan yksilöidä tietokannassa oleva rivi.
BI/DW	Business Intelligence / Data Warehouse. Johdon raportointijärjestelmät / tietovarastointi.
Dimensiotaulu	Tietokantataulu, joka sisältää ohjaustietoja.
eBIC	LogicaCMG:n BI/DW konseptituote.
ETL	Extract-Transform-Load. Tiedon hakeminen lähdejärjestelmästä, tiedon muuntaminen ja tiedon lataaminen kohdejärjestelmään
Faktataulu	Tietokantataulu, joka sisältää tapahtumatietoja.
Lastausalue	Katso työaluealue.
Perusavain	Sarake, jolla voidaan yksilöidä tietokantataulun rivi
SSIS 2005	SQL server 2005 Integration services
Surrogaattiavain	Keinotekoisesti muodostettu avain. eBIC-konseptituotteessa surrogaattiavaimet ovat juoksevia numerosarjoja.
Työaluetaulu	Tietokantataulu, jota käytetään tiedon varastoimiseen ennen tiedon käsittelemistä
Vierasavain	Tietokantataulun sarake, joka viittaa tietokantataulun perusavaimen.

1 JOHDANTO

LogicaCMG:n Suomen Lahden toimiyksikkö Compsol31200 on ottanut SSIS 2005 (SQL Server 2005 Integration Services) -tuotteen käyttöön Microsoft SQL server 2005 tietokantapalvelimen kanssa. Toimiyksikön yhtenä tavoitteena on siirtyä käyttämään Microsoft SQL server 2005 -tietokantapalvelinta Microsoft SQL server 2000 tietokantapalvelimen sijasta. Osana siirtymistä on ottaa käyttöön SSIS 2005 paketit ETL-prosessissa (Extract-Transform-Load).

Toimiyksikkö Compsol31200 tarvitsi itse toteutetun tietovuokomponentin SSIS 2005 -ohjelmaan yksinkertaistamaan avaimenhaku-prosessia SSIS 2005 -ohjelmassa. SSIS 2005 -ohjelmaa käytetään toimiyksikön Compsol31200 eBIC-konseptituotteen ETL-prosessissa.

LogicaCMG on merkittävä kansainvälinen IT-palveluyritys. Sen palveluksessa on noin 40 000 henkilöä 41 maassa. Toiminnan tavoitteena on mahdollistaa asiakkaiden kilpailukykyinen toiminta. Toimialalähtöiseen tarjontaan kuuluu muun muassa liiketoimintakonsultointi, järjestelmäintegraatiopalvelut sekä tietotekniikan ja liiketoimintaprosessien ulkoistamispalvelut. LogicaCMG:n palveluksessa on Pohjoismaissa noin 9 000 henkeä, joista Suomessa noin 2 600. Suomessa WM-data, a LogicaCMG company toimii noin 20 paikkakunnalla. (LogicaCMG 2007.)

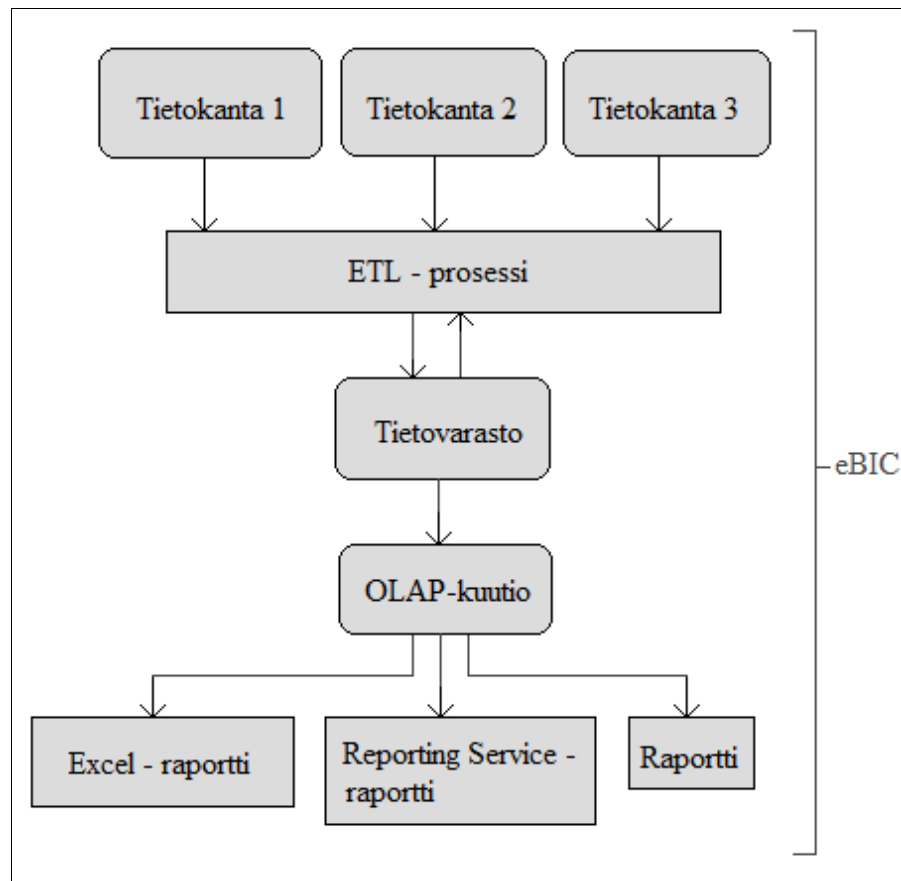
Työn tavoitteena on toteuttaa itse tehty avaimenhakukomponentti SSIS 2005 -ohjelmaan. Työn alussa avaimenhakuprosessi toteutettiin usealla erillisellä SSIS 2005 -komponentilla, ja tätä prosessia haluttiin yksinkertaistaa toteuttamalla avaimenhakuprosessi yhdellä komponentilla. Yksinkertaistamisen tarkoituksena on tehdä avaimenhakuprosessista käytettävyydeltään helpompi operaatio. Koska toteutus avaimenhakulle oli jo olemassa, käytettiin olemassa olevaa toteutusta määrittelemään uuden toteutuksen toiminta.

Tutkimusongelmana on yksinkertaistaa LogicaCMG:n eBIC-järjestelmän ETL-prosessin avaimenhakutoimenpidettä SSIS 2005 -ohjelmassa. Yksinkertaistaminen toteutetaan siirtämällä usean SSIS 2005 -komponentin toiminnallisuus yhden itse toteutetun tietovuokomponentin sisälle. Kehitystyökaluksi valittiin Microsoft Business Intelligence Studio, koska se on SSIS 2005 -järjestelmän käyttöliittymä.

2 TOIMINTAYMPÄRISTÖ

2.1 eBIC

Avaimenhakukomponentin toimintaympäristö sijoittuu eBIC-konseptituotteen ETL-prosessiin. eBIC-konseptituote on tietovarastointi- ja raportointituote, joka on tehty toimiyksikkö Compsol31200-toimintaperiaatteiden mukaisesti. Jokaista tietovarastointi- ja raportointijärjestelmää, jonka toimiyksikkö Compsol31200 on valmistanut, kutsutaan eBIC konseptituotteeksi. Kuviossa 1 on esitetty eBIC konseptituotteen yleinen kokoonpano.



KUVIO 1. eBIC konseptituote

Kuviosta 1 voidaan nähdä, että eBIC konseptituote on laajahko käsite, eikä eBIC konseptituote rajoitu pelkästään toimiyksikön CompSol31200 toteuttamiin

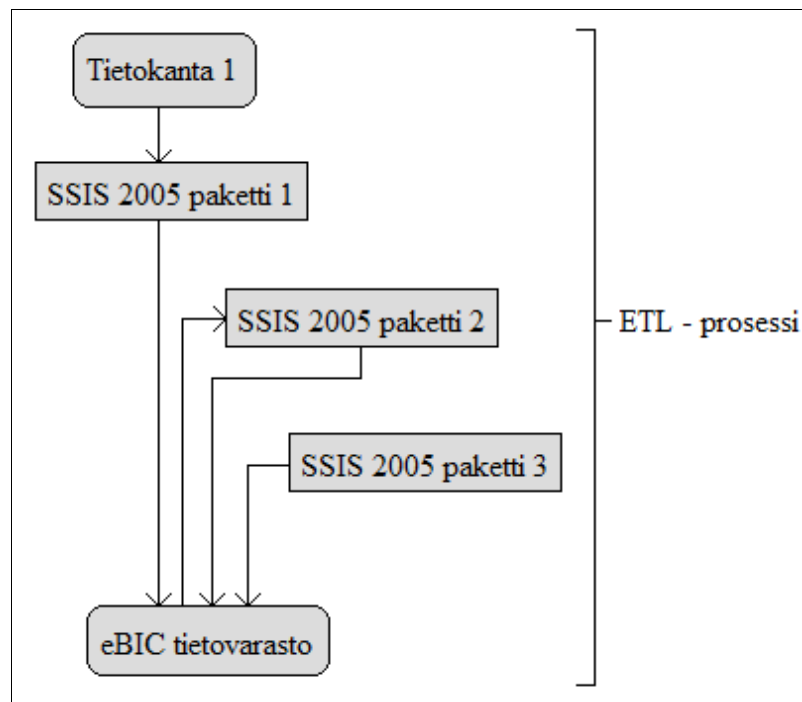
tuotteisiin. Esimerkkinä tästä voidaan antaa Excel-raportti. Excel raportti kuuluu eBIC konseptituotteeseen, jos sen on toteuttanut toimiyksikön CompSol31200 työntekijä toimiyksikön yleisten periaatteiden mukaisesti.

Tässä työssä käsitellään eBIC konseptituotteesta tarkemmin vain ETL – prosessia ja avaimenhakukomponenttia, joka sijaitsee ETL – prosessissa. Tietovarastoa, OLAP-kuutiota ja raportteja ei käsitellä, koska ne eivät ole tärkeitä avaimenhakukomponentin toteutuksen kannalta. eBIC-konseptituotteen tietovarastosta käytetään tässä dokumentissa myös työnimeä eBIC-tietovarasto.

Toimiyksikön toteuttamat eBIC-konseptituotteet jakautuvat kahteen eri luokkaan: Perusversioon, jota myydään asiakkaille valmiina ratkaisuna, sekä asiakaskohtaisesti tehtyyn eBIC sovellukseen, joka toteutetaan asiakkaan toiveiden mukaisesti. Avaimenhakukomponenttia käytetään sekä perusversiossa että asiakaskohtaisissa toteutuksissa.

2.2 ETL – prosessi

eBIC konseptituotteen ETL-prosessi vastaa tiedon siirtämisestä lähdejärjestelmästä kohdejärjestelmään. Tässä työssä kohdejärjestelmänä on eBIC-tietovarasto. ETL-prosessi lyhenteenä tarkoittaa tietojen hakemista, muuntamista ja lataamista. Kuviossa 2 on esitetty ETL – prosessin osa.



KUVIO 2. ETL – prosessin osa

Kuvion 2 kuvaama ETL – prosessin osa voisi kuvata seuraavanlaista ETL – tapahtumaa. ”SSIS 2005 paketti 1” lataa tiedot ”Tietokanta 1” tietokannasta eBIC tietovarastoon ”eBIC tietovarasto”. ”SSIS 2005 paketti 2” muuntaa ”SSIS 2005 paketti 1” lataamat tiedot eBIC tietovaraston käyttämään muotoon. Kun ”SSIS 2005 paketti 2” on muuntanut tiedot haluttuun muotoon, ”SSIS 2005 paketti 3” suorittaa eBIC tietovarastossa olevan tallennetun proseduurin, joka suorittaa jälkilaskennan tietovarastoon ladatuille tiedoille. Jälkilaskenta voisi olla esimerkiksi myyntitilausrivien summien laskeminen kuukausitason summatauluun.

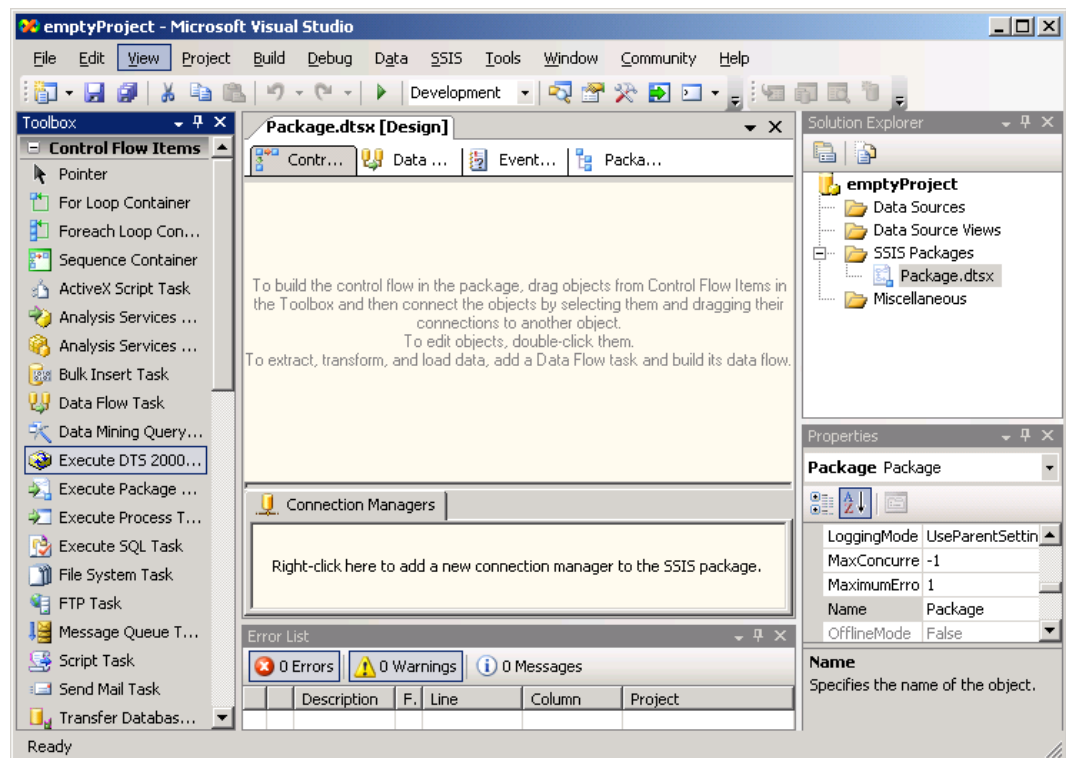
Kuvion 2 ETL – prosessi on yksinkertaistettu tapaus ETL – prosessin tapahtumasta eikä sitä pidä mieltää itsessään ETL – prosessiksi. Tässä työssä ei käsitellä tarkemmalla tasolla ETL – prosessia, koska se ei ole olennainen avaimenhakukomponentin toiminnallisuuden kannalta.

Avaimenhakukomponentti voisi sijoittua kuviossa 2 esimerkiksi ”SSIS 2005 -paketti 2” SSIS 2005 -pakettiin. SSIS 2005 -pakettia käsitellään tarkemmin luvussa 3.3.

3 SSIS 2005

3.1 Yleistä

SSIS 2005 on Microsoft:in kehittämä tuote tiedon siirtämiseen lähdejärjestelmästä kohdejärjestelmään. Lähde- ja kohdejärjestelmänä voi toimia tekstitiedosto, tietokantayhteys tai rajapinta, jolle on toteutettavissa SSIS 2005 järjestelmän liittymä. SSIS 2005 tuote pohjautuu putket ja suodattimet rakenteeseen. Tuotteen käyttöliittymänä toimii Microsoft Business Intelligence Studio -ohjelma. Kuviossa 3 on esitetty SSIS 2005 -ohjelman käyttöliittymä.



KUVIO 3. SSIS 2005 käyttöliittymä

SSIS 2005 -ohjelma tuotiin markkinoille SQL Server 2005 -tietokantapalvelimen yhteydessä. SSIS 2005 -ohjelman edeltäjä oli SSIS 2000.

3.2 SSIS 2005 arkkitehtuuri yleisellä tasolla

SSIS 2005 koostuu tehtävävuoto- ja tietovuoto-moottorista. Tietovuoto on osana tehtävävuotoa. Tehtävävuoto tarkoituksena on ohjata tietovuotoa ja tarjota tietovuotoille ajonaikaiset resurssit. Tehtävävuoto ja tietovuoto yhdistelmä mahdollistaa SSIS 2005 ohjelman tehokkuuden perinteisissä ETL-sovelluksissa. (SSIS 2005 esittely 2007.)

Tehtävävuoto voi sisältää useita tietovuotoa, mutta tietovuoto voi kuulua vain yhteen tehtävävuotoon. Tietovuoto voi kuulua vain yhteen tehtävävuotoon, koska SSIS 2005 ohjelmassa voi olla vain yksi tehtävävuoto.

3.2.1 Putket ja suodattimet arkkitehtuuri

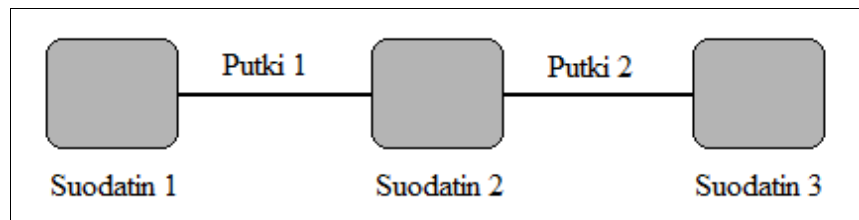
SSIS:n ytimenä on tietomuunnosputki. Putkella on puskuriorientoitunut arkkitehtuuri, joka on nopea manipuloimaan tietoa sen jälkeen, kun tieto on ladattu muistiin (SSIS 2005 esittely 2007.)

Tieto liikkuu putkia pitkin suodattimelta toiselle SSIS 2005 -ohjelmassa. Suodatin voi sijaita joko tehtävävuotoissa tai tietovuotoissa. Tiedon liikkuminen putkia pitkin suodattimelta toiselle tapahtuu muistissa. Poikkeuksena on tapaus, jossa putkessa liikkuva tietomäärä on niin suuri, että putkessa oleva tieto kirjoitetaan kovalevyllä sijaitsevaan välimuistiin tavallisen muistin loppuessa.

Tehtävävuotoissa olevat putket välittävät tiedon suorituksen aloituksesta. Tietovuotoissa olevat putket välittävät tietoa. Putkia käytetäänkin tehtävävuotoissa tehtävien ohjaamiseen ja tietovuotoissa tiedon välittämiseen.

Suodatin on ohjelma, joka kiinnittyy putkeen. Tietovuossa suodatin voi lukea ja kirjoittaa putkeen. Suodatin voi välittää, muuntaa, lisätä ja poistaa putken välittämiä tietoja. Poikkeuksena on tilanne, jossa putken välittämä tieto tai tiedonosa on merkitty vain luettavaksi olevaksi tiedoksi. Tässä tapauksessa suodatin ei voi muuntaa tietoa.

Kuviossa 4 on esitetty putket ja suodattimet -arkkitehtuurin perusrakenne. Sovellus, joka on toteutettu putket ja suodattimet arkkitehtuurilla sisältää suodattimia ja putkia. Suodatin käsittelee tietoa. Putki välittää tietoa suodattimien välillä.

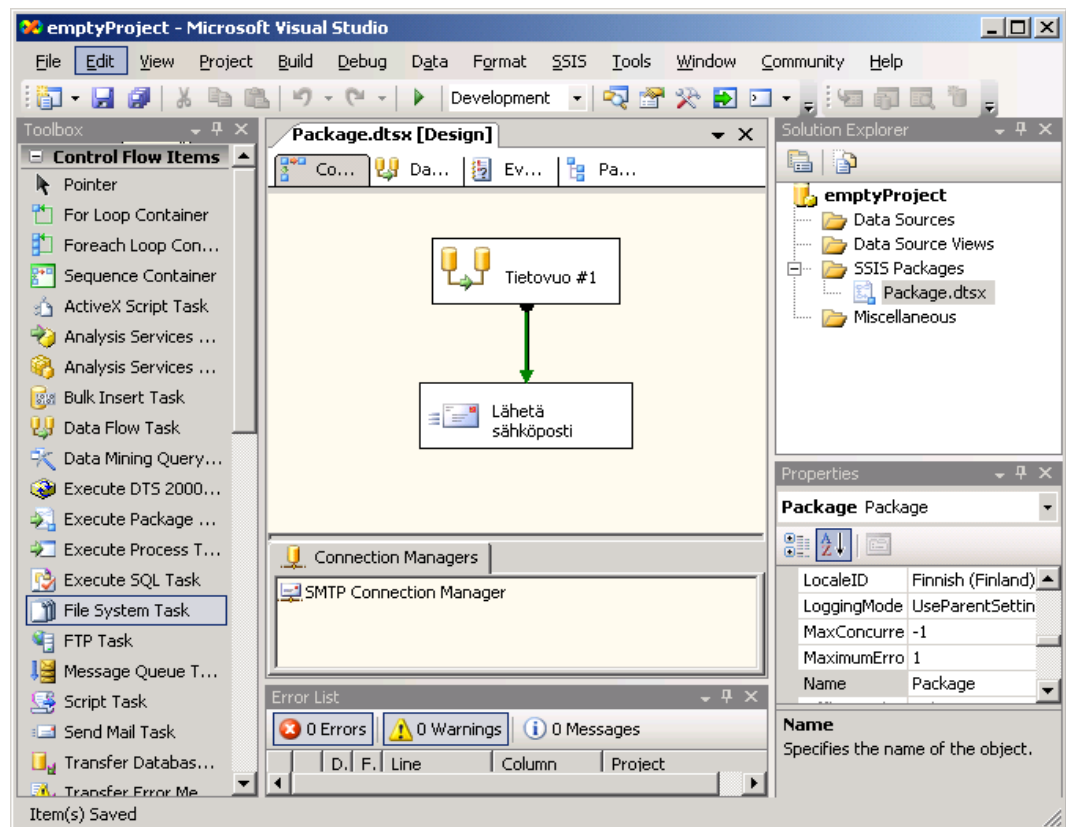


KUVIO 4. Putket ja suodattimet

SSIS 2005 -ohjelmassa suodattimet ovat ohjaus- ja tietovuohon sijoitettuja komponentteja. Ohjaus- ja tietovuoto voidaan rinnastaa putket ja suodattimet arkkitehtuurin putkiin.

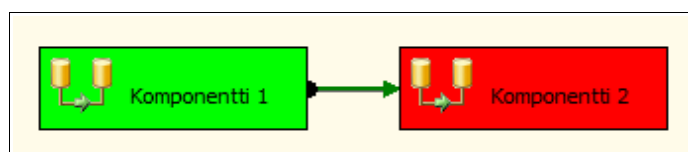
3.3 SSIS 2005 paketti

SSIS 2005 -paketti on XML-muotoinen määritelmätiedosto, joka määrittelee tehtävävuon ja tehtävävuohon mahdollisesti liitetyt tietovuot suodattimineen. SSIS 2005 käyttöliittymässä toteutetut ohjelmat ovat paketteja. Paketteihin liitettävät komponentit ovat suodattimia ja yhteydet, joilla komponentit liitetään toisiin komponentteihin, ovat putkia. Kuviossa 5 on esitetty SSIS 2005 paketin tehtävävuonäkymä, johon on liitetty tietovuokomponentti ja sähköpostinlähetyskomponentti.



KUVIO 5. SSIS 2005 -paketti

SSIS 2005 -paketissa olevat komponentit ilmoittavat paketin suoritusaikana visuaalisesti suoritustilansa. Epäaktiiviset komponentit värikoodataan valkoisiksi ja suoritustilassa olevat paketit värikoodataan keltaisiksi. Suoritettut komponentit värjätään joko vihreiksi tai punaisiksi sen mukaan, minkä tilakoodin komponentti palauttaa. Onnistuneen suorituksen tilakoodin palauttaneet komponentit värjätään vihreiksi ja epäonnistuneet punaisiksi. Kuviossa 6 on esitetty tilanne, jossa ”komponentti 1” on suoritettu onnistuneesti, mutta ”komponentti 2” on palauttanut virhekoodin.



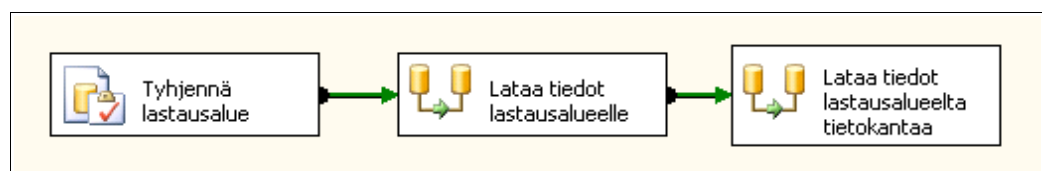
KUVIO 6. SSIS 2005 -komponenttien väritilakoodit

3.3.1 Tehtävävu

Tehtävävu ohjaa paketin suoritusta. Tehtävön kenties merkittävin ero tietovuohon on se, että tehtävävuoossa putket eivät niinkään välitä varsinaista tietoa kuin ohjaustietoa. Komponenttien suoritus aloitetaan ja lopetetaan tehtävävuoossa sen mukaan, miten niille välitetään suorituskäsky putkea pitkin. Komponentti aloittaa suoritustilansa, kun se saa kuuntelemaltaan putkelta aloituskäskyn. Suoritettuaan toimintonsa komponentti palauttaa SSIS 2005 -paketille tilakoodin suorituksestaan ja kirjoittaa aloituskäskyn omaan ulosmenoputkeensa.

Komponentin SSIS 2005 -paketille välittämä tilakoodi voi lopettaa paketin suorittamisen, jos tilakoodi tulkitaan paketin lopettavaksi virhetilaksi. Paketin lopettava virhetila voisi olla esimerkiksi tietokantayhteyden puuttuminen yritettäessä tietokannasta lukemista.

Kuviossa 7 on esitetty tehtävävu, joka sisältää kolme komponenttia. Ensimmäinen komponentti on SQL-tehtävä ja loput kaksi ovat tietovuokomponentteja, jotka toteuttavat tietovuon. Kuvio 7 esittämän tehtävön suoritus aloitetaan SQL-tehtävästä ”Tyhjennä lastausalue”. Kun SQL-tehtävä on suoritettu siirrytään tehtävävuoossa tietovuokomponenttiin, joka on nimetty ”Lataa tiedot lastausalueelle”. Kun tietovuokomponentti ”Lataa tiedot lastausalueelle” on suoritettu, suoritetaan tehtävävu ”Lataa tiedot lastausalueelta tietokantaan”.



KUVIO 7. Esimerkki tehtävystä

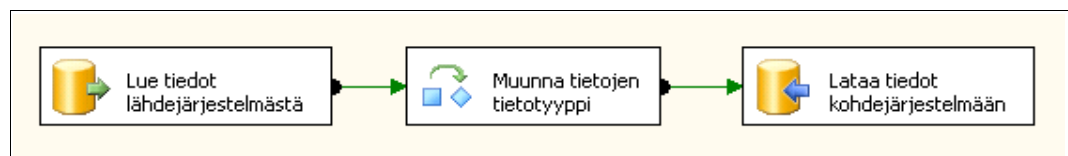
3.3.2 Tietovuo

Tietovuossa välitetään tietoja komponenttien eli suodattamien välillä. Komponenttien suoritus aloitetaan ja lopetetaan tietovuossa sen mukaan, miten niille välitetään tietoa. Komponentti on epäaktiivisessa tilassa niin kauan kuin sille ei välitetä tietoa putkea pitkin. Kun komponentille välitetään tietoa, suorittaa se sille määritellyn tehtävän ja palauttaa SSIS 2005 -paketille tilakoodin tehtävän suorituksen onnistumisesta.

Komponentin paketille välittämä tilakoodi voi lopettaa paketin suorituksen, jos se tulkitaan paketin lopettavaksi virhetilaksi. Tietovuossa tämänkaltainen virhe voisi olla esimerkiksi tilanne, jossa putkeen ei pystytä kirjoittamaan tietoa.

Tietovuon komponentti lopettaa suorituksensa ja palauttaa tilakoodin, kun sen lukema putki asettaa itsensä suljettuun tilaan. Kun komponentin lukema putki on suljetussa tilassa, tietää komponentti, että putki ei välitä enää lisätietoja.

Kuviossa 8 on esitetty tietovuo, joka sisältää kolme komponenttia. Tietovuo voisi olla kuviossa 7 esitetty tietovuokomponentti ”Lataa tiedot lastausalueelle”. Tietovuon suoritus aloitetaan komponentista ”Lue tiedot lähdejärjestelmästä”. Tietovuon suorituksessa siirrytään seuraavaksi komponenttiin ”Muunna tietojen tietotyyppi”. Viimeisenä tietovuossa suoritetaan komponentti ”Lataa tiedot kohdejärjestelmään”.

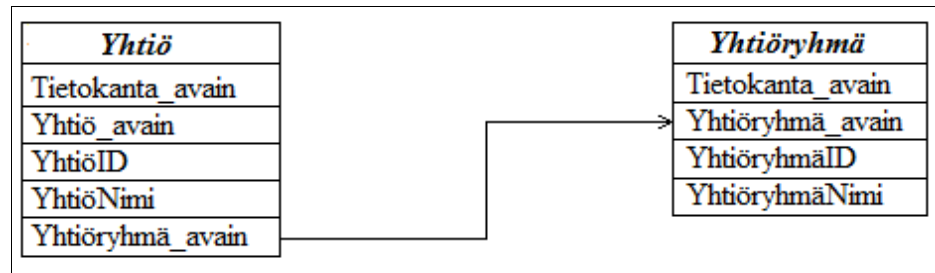


KUVIO 8. Esimerkki tietovuosta

4 AVAIMENHAKU

4.1 Yleistä

Avaimenhakua käytetään eBIC-sovelluksen tietokantataulujen yhdistämisessä. Jokaisessa tietokantataulussa on surrogaatti-perusavain, jota voidaan käyttää liittäessä kahta taulua. Liitos tehdään hakemalla liitettävästä taulusta surrogaatti-perusavain ja lisäämällä se vierasavaimeksi liitostauluun. Kuviossa 9 on esitetty tilanne, jossa liitettävästä taulusta haetaan perusavain liitostauluun. Kuvion 9 taulussa liitettävä taulu on yhtiöryhmä ja liitostaulu on yhtiö.



KUVIO 9. Avainliitos kahden taulun välillä

Taulukossa 1 on esimerkki tilanteesta, jossa yhtiötaulun tiedot on liitetty yhtiöryhmätaulun tietoihin. Yhtiöryhmä-taulu on liitetty yhtiö-tauluun sarakkeella yhtiöryhmä_avain. Liitos on tehty hakemalla yhtiö-taulun riville yhtiöryhmä-taulusta rivi, jolla on sama yhtiöryhmä_avain arvo kuin yhtiö-taulun rivillä.

TAULUKKO 1. Yhtiö- ja yhtiöryhmä-taulun tiedot liitettynä toisiinsa

Yhtiö_avain	YhtiöNimi	Yhtiöryhmä_avain	YhtiöryhmäNimi
1	Yhtiö_1	1	Yhtiöryhmä_1
2	Yhtiö_2	1	Yhtiöryhmä_1
3	Yhtiö_3	2	Yhtiöryhmä_2
4	Yhtiö_4	3	Yhtiöryhmä_3

4.2 Haettavat arvot

eBIC-sovelluksen ETL-prosessissa voidaan hakea kolmea erityyppistä arvoa. Kukin haettava arvotyyppi ilmaisee yksiselitteisesti sen, miten taulujen liitos on onnistunut.

eBIC-sovelluksen ETL-prosessissa löydetyille arvoille palautetaan löydetyn arvon surrogaattiavain. Surrogaattiavain on keinotekoisesti muodostettu avainsarake, jolla voidaan yksiselitteisesti yksilöidä tietokantataulun rivi. eBIC-sovelluksessa surrogaattiavaimet ovat juoksevia numerosarjoja.

4.2.1 Tunnistearvo

Tunnistearvo voi olla yhden tai useamman sarakkeen muodostama arvojoukko, jolla haetaan tietokantataulusta vastaavaa arvojoukkoa. Yhdestä sarakkeesta muodostettu tunnistearvo voi olla esimerkiksi Yhtiö-taulun YhtiöID-sarake. YhtiöID-sarake on sarake, jolla Yhtiö on yksilöity lähdejärjestelmässä.

Kuviossa 10 on esitetty tilanne, jossa Yhtiö-taulusta haetaan tunnistearvoa ”Yht002”. Kun tietokantataulusta on löydetty rivi, jolla on tunnistearvo ”Yht002”, palautetaan kyseisen rivin perusavain.

Tietokanta_avain	Yhtiö_avain	YhtiöID	Yhtiönimi	Yhtiöryhmä_avain
1	1	Yht001	Yhtiö_1	1
1	2	Yht002	Yhtiö_2	1
1	3	Yht003	Yhtiö_3	2

KUVIO 10. Surrogaattiavaimen palautus yhdellä sarakkeella.

Jos kuvion 10 taulussa olisi useampi YhtiöID arvo ”Yht002”, palautettaisiin taulusta ensimmäisen löydetyn rivin perusavain.

Kahdesta sarakkeesta muodostettu arvojoukko voi muodostua esimerkiksi Yhtiötaulun sarakkeista Tietokanta_avain ja YhtiöID. Kahdella tai useammalla sarakkeella muodostettua arvojoukkoa tarvitaan tapauksessa, jossa yhdellä sarakkeella ei voida yksilöidä haettavaa arvoa. Tämänkaltainen tilanne voi muodostua muun muassa seuraavalla tavalla: ETL-prosessi hakee kahdesta lähdejärjestelmästä tiedot Yhtiö-tauluun. Koska lähdejärjestelmät ovat tässä tapauksessa toisistaan erillisiä järjestelmiä, yhtiöillä voi olla sama yhtiöID-arvo. Tässä tapauksessa tunnistearvona on käytettävä YhtiöID- ja Tietokanta_avain-saraketta. YhtiöID-sarake ilmaisee haettavan yhtiön tunnistenumeron ja Tietokanta_avain-sarake sen, minkä lähdejärjestelmän tietoja vastaan YhtiöID-tietoa halutaan verrata.

Kuviossa 11 on esitetty tilanne, jossa Yhtiö-taulusta haetaan perusavainta tunniste-arvojoukolla; YhtiöID on ”Yht002” ja Tietokanta_avain on ”1”. Kun tietokanta-taulusta on löydetty rivi jolla YhtiöID on ”Yht002” ja Tietokanta_avain on ”1”, palautetaan kyseisen rivin perusavain.

Tietokanta_avain	Yhtiö_avain	YhtiöID	Yhtiönimi	Yhtiöryhmä_avain
1	1	Yht001	Yhtiö_1	1
①	②	Yht002	Yhtiö_2	1
1	3	Yht003	Yhtiö_3	2
2	4	Yht001	Yhtiö_1	3
2	5	Yht002	Yhtiö_2	4

KUVIO 11. Surrogaattiavaimen palautus kahdella sarakkeella.

Jos kuvion 11 taulusta olisi haettu YhtiöID arvoa ”Yht002” ilman tietokanta_avain lisärajausta, olisi perusavaimena palautettu joko ”2” tai ”5” riippuen hakumoottorin toiminnasta.

4.2.2 Tunnistearvo, jolla ei löydetä surrogaattiavainta

Jos tunnistarvohaku ei palauta surrogaattiavainta, on määritelty, että eBIC-sovelluksen ETL-prosessi hakee tunnistearvoa, joka on määritelty ”Avainta ei löydy” arvoiseksi. ”Avainta ei löydy”-arvo on määritelty '??'-merkiksi eBIC-sovelluksessa.

”Avainta ei löydy”-tunnistearvoa käytetään, jotta tietokannasta muodostetulla raportilla saataisiin näkymään mahdollisimman paljon tietoa. Esimerkkitilanteena voisivat olla laskutusrivit ja niihin liittyvät tilausnumerot. Esimerkiksi jos lähdejärjestelmästä tulee tilausnumero, jota ei löydy tilausnumero taulusta, ei puuttuvan tilausnumeron laskutusriviä voitaisi näyttää raportilla. Tämä sen takia, että laskutusrivien ja tilausnumeron välille tehdään liitos. Tietokannasta muodostetulla raportilla voidaan näyttää vain sellaiset rivit, jotka täyttävät liitoksen tilausnumero- ja laskutusrivi-taulun välillä.

”Avainta ei löydy” tunnistearvon käyttäminen on raportointitekkinen ratkaisu, jolla varmistetaan raportoinnin kannalta tärkeiden tietojen näyttäminen raportilla. Esimerkkitapauksessa tilausnumeron puuttuminen raportilta on mielletty eBIC-sovelluksessa pienemmäksi haitaksi kuin laskutusrivin laskutustietojen puuttuminen.

Jotta '?' merkkiarvoa voitaisiin etsiä tarvittaessa, on tarvittaviin tietokantatauluihin määriteltävä käsin rivi joka sisältää '?' merkkiarvon. Kuviossa 12 on näytetty '?' merkkiarvon lisääminen tietokantatauluun.

Tietokanta_avain	Yhtiö_avain	YhtiöID	Yhtiönimi	Yhtiöryhmä_avain
1	1	??	Unknown ID	1
2	2	Yht001	Yhtiö_001	2
3	3	Yht002	Yhtiö_002	2

KUVIO 12. ”Avainta ei löydy” - tunnistearvo tietokantataulussa

Tietokantataulun rivillä, joka sisältää '?'-merkkiavaimen, on seuraava erikoisominaisuus. Jokainen tietokantarivin vierasavain viittaa osoittamansa vierastaulun riviin, joka on '?'-merkkiavaimen sisältävä rivi. Kuvion 12 tapauksessa '?' merkki-rivillä on yhtiöryhmä_avain "1". Yhtiöryhmä-taulussa on siten oltava avaimen "1" rivillä yhtiöryhmä, joka on määritelty '?' merkkiavaimella arvoon "Avainta ei löydy"

On määritelty, että '?' merkkisarvoa haetaan tietokantataulusta aina yhdellä sarakkeella. Vaikka alkuperäinen tunnistarvojoukko olisikin koostunut useammasta sarakkeesta haetaan '?' merkkisarvoa vain yhdellä sarakkeella. Sarake, jota käytetään haussa on aina tapauskohtainen, ja se pitää erikseen päättää eBIC-sovelluksen ETL-prosessia määriteltäessä.

4.2.3 Puuttuva tunniste-arvo

Jos tunniste-avaimelle ei välitetä tunniste-arvoa, on määritelty, että eBIC-sovelluksen ETL-prosessi hakee tunniste-arvoa joka on määritelty "Puuttuva tunniste-arvo" arvoiseksi. "Puuttuva tunniste-arvo" arvo on määritelty '?'-merkiksi eBIC-sovelluksessa.

"Puuttuva tunniste-arvo" tunniste-arvoa käytetään samasta syystä, kuin "Avainta ei löydy" eli '?' -merkkiä. Kyseessä on siis raportointitekniinen seikka, jolla varmistetaan tietojen näkyminen loppuraportilla.

Jotta '?' merkkisarvoa voitaisiin etsiä tarvittaessa, on tarvittaviin tietokantatauluihin määriteltävä käsin rivi joka sisältää '?' merkkisarvon. Kuviossa 13 on näytetty '?' merkkisarvon lisääminen tietokantatauluun.

Tietokanta_avain	Yhtiö_avain	YhtiöID	Yhtiönimi	Yhtiöryhmä_avain
1	1	?	Missing ID	1
2	2	Yht001	Yhtiö_001	2
3	3	Yht002	Yhtiö_002	2

KUVIO 13. ”Puuttuva tunnistearvo” - tunnistearvo tietokantataulussa

Tietokantataulun rivillä, joka sisältää '?' merkkiavaimen on seuraava erikoisominaisuus. Jokainen tietokantarivin vierasavain viittaa osoittamansa vierastaulun riviin, joka on '?' merkkiavaimen sisältävä rivi. Kuvion 13 tapauksessa '?' merkkirivillä on yhtiöryhmä_avain ”1”. Yhtiöryhmä-taulussa on siten oltava avaimen ”1” rivillä yhtiöryhmä, joka on määritelty '?' merkkiavaimella arvoon ”Puuttuva tunnistearvo”.

On määritelty, että '?' merkkiarvoa haetaan tietokantataulusta aina yhdellä sarakkeella. Vaikka alkuperäinen tunnistarvojoukko olisikin koostunut useammasta sarakkeesta haetaan '?' merkkiarvoa vain yhdellä sarakkeella. Sarake, jota käytetään haussa on aina tapauskohtainen, ja se pitää erikseen päättää eBIC-sovelluksen ETL-prosessia määriteltessä.

5 AVAIMENHAKU KOMPONENTIN SUUNNITTELU

5.1 Yleistä

Avaimenhakukomponentin suunnittelu on tehty yhteistyössä toimiyksikön Compsol31200 kanssa. Komponentin suunnittelu on painottunut käytettävyysslähtöisesti ja suunnittelu aloitettiin määrittelemällä ne toiminnallisuudet, jotka komponentin on toteutettava. Sen jälkeen mietittiin, miten määritellyt toiminnallisuudet saadaan helposti käytettäväksi komponentissa.

Suunnittelu aloitettiin tässä tapauksessa käyttöliittymäkerroksesta, eikä bisneslogiikka kerroksesta, koska yhdelläkään toimiyksikön Compsol31200 työntekijällä ei ollut kokemusta SSIS 2005 tietovuokomponentin toteutuksesta. Käyttöliittymäkerroksen suunnittelussa oli useita iteraatiokertoja. Käytännössä tämä tarkoittaa sitä, että komponentin käyttöliittymää muutettiin käytännön kokemusten myötä. Käyttäjille esiteltiin toteutettu komponentti, ja käyttäjät saivat esittää parannusehdotuksensa komponentin käyttöliittymään. Tätä toistettiin niin kauan, että käyttäjät olivan pääosin tyytyväisiä komponentin käyttöliittymään.

Bisneslogiikan suunnitteleminen SSIS 2005 tietovuokomponenttiin oli prosessi, joka tehtiin yrityksen ja erehdyksen kautta. Suunnitteluun tarvittiin useita iteraatioita, koska oppiminen tapahtui osittain toteutettaessa tietovuokomponenttia. Toteutuksen aikana tehtyä suunnitelmaa muutettiin useamman kerran, koska havaittiin, että komponenttia ei ollut mahdollista toteuttaa suunnitelmien mukaisesti.

Komponentin suunnittelu kokonaisuudessaan oli useita iteraatioita vaativa operaatio, joka vietiin läpi komponentin toteutuksen aikana saatujen kokemusten tuomien lisätietojen avulla. Suunnittelun aikana toimiyksikkö Compsol31200 piti useita kontrollipalaverejä avainhenkilöiden kanssa. Kontrollipalavereilla todettiin, että komponentin suunnittelu ja toteutus ovat toimiyksikön kannalta oikeilla linjoilla.

Suunnittelun päätavoitteena oli luoda toimiva ja helppokäyttöinen avaimenhaku-komponentti. Avaimenhaku oli jo toteuttavissa SSIS 2005 -komponenteilla, ja sitä haluttiin yksinkertaistaa itse toteutetulla SSIS 2005 -tietovuokomponentilla.

5.2 Komponentin toiminnalliset vaatimukset

Komponentin toiminnalliset vaatimukset on otettu toimiyksikön Compsol31200 ”parhaista käytännöistä” ja suunnittelupalaverien tuomista ideoista. ”Parhaat käytännöt” tarkoittavat tässä tapauksessa toimiyksikön käytännöksi muodostuneita toimintatapoja. Käytännöksi muodostuneita toimintatapoja ovat muun muassa

- '?' ja '??' avainarvojen hakeminen ja asettaminen
- virheenkäsittely virhetilanteissa.

Komponentin suunnittelun aikana ja komponenttiin liittyneissä suunnittelu- ja kontrollipalavereissa komponentille määriteltiin seuraavat toiminnalliset vaatimukset:

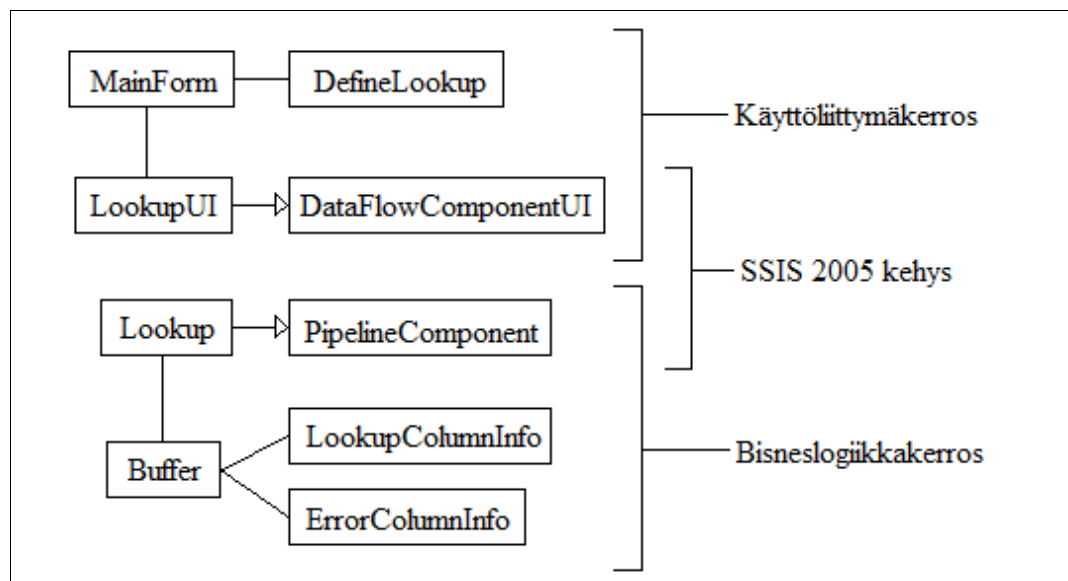
- Komponentin on suoriuduttava '?' ja '??' avainarvojen hakemisesta.
- Komponentin on suoriuduttava avainarvojen hakemisesta yhdellä tunniste- arvolla tai useamman tunniste- arvon joukolla.
- Komponentin on palautettava virhetilanne, jos komponentti ei löydä '?' tai '??' avainarvoja niitä haettaessa.
- Komponentin on oltava suorituskyvältään samaa tasoa kuin SSIS 2005 komponenteilla toteutettu avaimenhaku.
- Komponentissa on oltava debug-toiminto, jolla voidaan jäljittää mahdollisia virhetilanteita.
- Komponentti ei saa hukata tietoja ilmoittamatta siitä virhetilanteella.

Määritetyillä toiminnallisilla vaatimuksilla päästiin tilanteeseen, jossa komponentilla on vastaavat toiminnallisuudet SSIS 2005 -komponenteilla toteutetun avaimenhaun kanssa.

5.3 Komponentin rakenne

Komponentin rakenne on jaettu käyttöliittymä- ja bisneslogiikkakerrokseen. Käyttöliittymäkerros tarjoaa käyttäjälle käyttöliittymän avaimenhakukomponenttiin. Bisneslogiikkakerros vastaa komponentin toiminnallisuudesta. Käyttöliittymäkerros käyttää bisneslogiikkakerroksen toiminnallisuuksia osana käyttöliittymäkerroksen toiminnallisuuksia.

Kuviossa 14 on esitetty avaimenhakukomponentin rakenne luokkatasolla. Luokkarakenne koostuu kahdesta pääluokasta ja niihin liittyvistä apuluokista. Mainitut kaksi pääluokkaa ovat Lookup ja LookupUI, jotka toteuttavat komponentin rungon. Lookup on bisneslogiikkaluokka, joka toteuttaa tietovuokomponentille määritellyt rajapinnat. LookupUI on käyttöliittymäkerroksen luokka, joka käynnistää varsinaisen käyttöliittymän ja keskustelee bisneslogiikkaluokan kanssa.



KUVIO 14. eBIC avaimenhakukomponentin rakenne

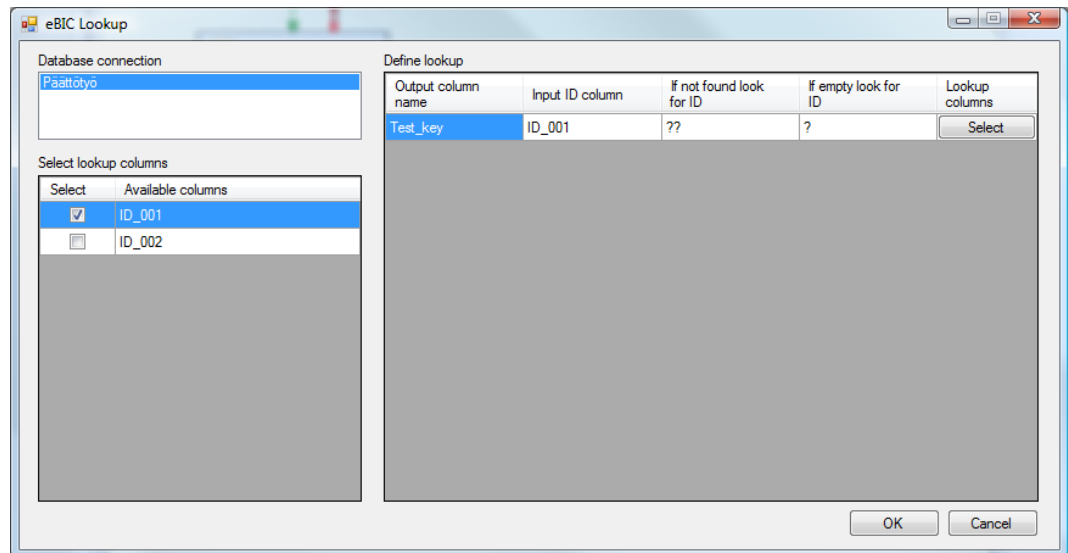
Kuviossa 14 oleva rajattu alue ”SSIS 2005 kehys” rajaa luokat, jotka toteuttavat SSIS 2005 kehysten toiminnallisuuksia. Huomionarvoista on myös se, että Lookup- ja LookupUI-luokat eivät näe toisiaan. Keskustelu luokkien välillä tapahtuu täysin SSIS 2005-kehysten tarjoamien toiminnallisuuksien kautta. Taulukossa 2 on listattu avaimenhakukomponentin luokat ja niiden kuvaukset.

TAULUKKO 2. Avaimenhakukomponentin luokat ja niiden kuvaukset

Luokan nimi	Luokan kuvaus
MainForm	Käyttöliittymäluokka, joka toteuttaa komponentin perusnäkyvän
DefineLookup	Käyttöliittymäluokka, joka toteuttaa avaimenhaun määrittämisen
LookupUI	Käyttöliittymäluokka, joka on periytetty DataFlow-ComponentUI luokasta.
DataFlowComponentUI	Käyttöliittymäluokka, joka toteuttaa rajapinnan IdtsComponentUI.
Lookup	Bisneslogiikkaluokka, joka on periytetty Pipeline-Component luokasta.
PipelineComponent	Bisneslogiikkaluokka, joka toteuttaa tietovuokomponentin rajapinnat
Buffer	Puskuriluokka, jota käytetään tiedon puskuroimiseen bisneslogiikkaluokassa
LookupColumnInfo	Tietorakenneluokka, jota käytetään puskuriluokassa
ErrorColumnInfo	Tietorakenneluokka, jota käytetään puskuriluokassa

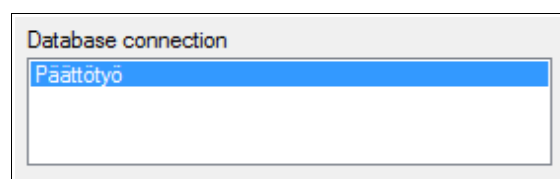
5.4 Komponentin käyttöliittymä

Avaimenhakukomponentin käyttöliittymästä on tehty yksinkertainen ja havainnollinen käytettävyydeltään. Lukuisten iteraatikertojen jälkeen perusnäkyksi avaimenhakukomponenttiin on muodostunut kuviossa 15 esitetty näkymä.



KUVIO 15. Avaimenhakukomponentin perusnäky

Avaimenhakukomponentin perusnäkyssä määritellään käytettävä tietokantayhteys. Lisäksi näkyssä määritellään tietovuosarakkeet, joilla suoritetaan avaimenhaku ja käytettävät avaimenhakuarvot. Kuviossa 16 on esitetty avaimenhakukomponentin perusnäkyssä esitety tietokantayhteyden valinta. Tietokantayhteys valitaan valitsemalla haluttu tietokantayhteys valintalistasta. Valintalistan tietokantayhteydet ovat samoja, jotka on määritelty paketin ja SSIS- projektin tietokantayhteyksiksi. Valitun tietokantayhteyden on oltava OleDb-muotoinen. Jos käyttäjä valitsee muun kuin OleDb-tietokantayhteyden, käyttäjälle annetaan virheilmoitus.



KUVIO 16. Tietokantayhteyden valinta

Kuviossa 17 on esitetty kuviossa 15 näkyvä avaimenhakusarakkeiden valintalista. Avaimenhakusarake valitaan rastittamalla Select-sarakkeessa oleva rasti. Rastitettu valinta siirtyy kuviossa 18 kuvattuun listaan.

Select	Available columns
<input checked="" type="checkbox"/>	ID_001
<input type="checkbox"/>	ID_002

KUVIO 17. Avaimenhakusarakkeiden valinta

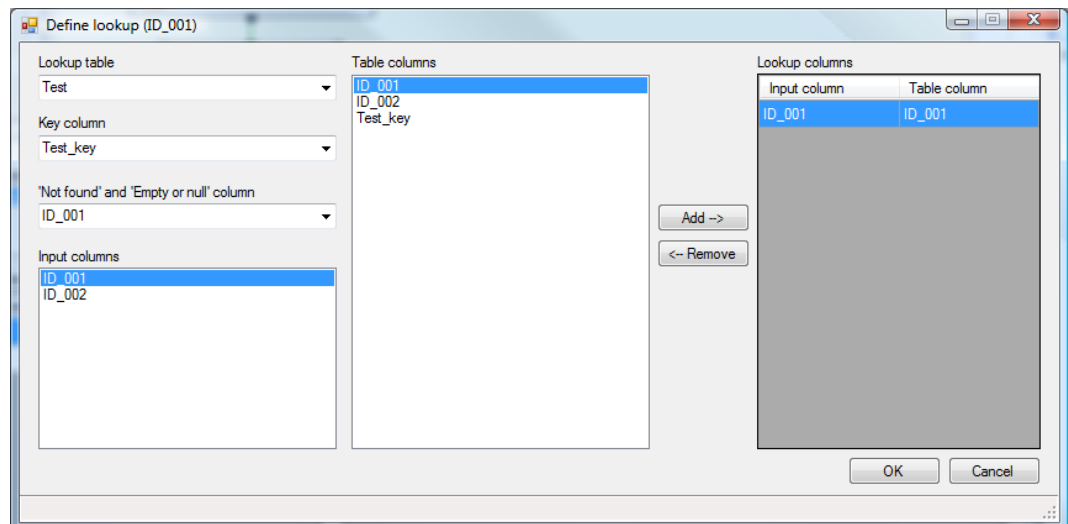
Kuviossa 18 on esitetty valitut avaimenhakusarakkeet. Kuvio 18 on osa kuviossa 15 esitettyä perusnäkyä. Kuvion 18 näkymässä voidaan määrittellä avaimenhaun perustiedot. Itse avaimenhaku määritellään lomakkeessa, joka avautuu painamalla Select-nappia. Avaimenhaun perustiedoissa määritellään avaimenhakusarakkeen nimi, arvo jota haetaan, jos surrogaattivainta ei löydetä ja arvo, jota haetaan, jos haettava arvo puuttuu.

Output column name	Input ID column	If not found look for ID	If empty look for ID	Lookup columns
Test_key	ID_001	??	?	Select

KUVIO 18. Avaimenhakusarakkeen määrittely perusnäkyssä

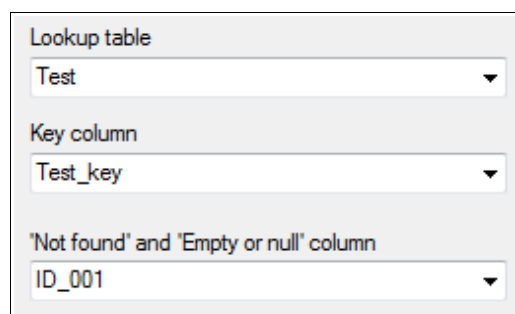
Kuviossa 18 ”Input ID Column” on ainoa sarake, jonka tietoa ei voida muuttaa. Sarake ”Input ID Column” kertoo avaimenhaussa ensisijaisesti käytettävän tietovuosarakkeen nimen.

Kuviossa 19 on esitetty avaimenhaun määrittelemisen. Lomake avautuu, kun kuviossa 18 näkyvää Select-nappia painetaan. Lomakkeella voidaan määrittellä taulu, johon avaimenhaku tehdään, palautettava avainsarake ja käytettävä sarake, jos esitetään '?' tai '??' arvoa. Lomakkeessa määritellään myös vertailut tietovuosarakkeiden ja avaimenhakutaulun sarakkeiden välillä.



KUVIO 19. Näkymä avaimenhaun määrittämiseen

Avaimenhakutaulu ja sen sarakkeet määritellään kuviossa 20 esitetyillä valinnoilla. Avaimenhakutaulu valitaan kohdasta "Lookup table". Kun taulu on valittu, lomake päivittää automaattisesti lomakkeesta kohdat, joissa näytetään avaimenhaku-taulun sarakkeet. Jos käyttäjä on tehnyt avaimenhakumäärittelyn tauluun ja valitsee toisen taulun, poistetaan aikaisemmat avaimenhakumäärittelyt jotta avaimenhaku voidaan tehdä vain yhteen tauluun.



KUVIO 20. Avaimenhakutaulun määrittelyt

Kuvion 21 ”Key column” alasetalista määrittelee sarakkeen, joka palautetaan avaimenhakuarvona. ”Not found' and 'Empty or null' column” määrittelee taulun sarakkeen, jota käytetään, jos surrogaattivainta ei löydetä, tai jos avaimenhakuarvo on tyhjä tai null.

Kuvion 21 punaisella viivalla ympyröidyllä näkymällä määritellään avaimenhakuoperaatioissa käytettävät sarakkeet. Kuvion 21 vasemmassa alalaidassa kohdassa ”Input columns” näytetään sarakkeet, jotka ovat käytettävissä tietovuon puolelta. ”Table columns” valintalistassa näytetään valitun avaimenhakutaulun sarakkeet. Kohdassa ”Lookup columns” esitetään avaimenhakuoperaatioissa käytettävät sarakkeet.

Avaimenhaku määritellään valitsemalla tietovuon sarake ”Input columns” -listasta. Valitsemisen jälkeen valitaan listasta ”Table columns” -sarake, joka halutaan verrata valittuun tietovuosarakkeeseen. Kun molemmat sarakkeet on valittu, painetaan lomakkeella näkyvää ”Add -->” nappia, joka lisää valitut sarakkeen avaimenhaku-operaatioksi.

Lookup table	Table columns	Lookup columns
Test	ID_001	Input column Table column
Key column Test_key	ID_002	ID_001 ID_001
'Not found' and 'Empty or null' column ID_001	Test_key	
Input columns ID_001 ID_002		

KUVIO 21. Avaimenhakusarakkeiden valinta

Avaimenhauksessa käytettävän sarakkeen vertailun voi poistaa valitsemalla ”Lookup columns” näkymästä poistettavan sarakkeen vertailun ja painamalla ”<-- Remove” nappia.

Kuviossa 20 esitetty sarake ”'Not found' and 'Empty or null' column” on oltava osa avaimenhakuoperaatiota. Jos mainittu sarake ei esiinny näkymässä ”Lookup columns”, lomaketta ei voida tallentaa.

6 AVAIMENHAUN TOTEUTUS SSIS 2005 PAKETISSA

6.1 Avaimenhaun toteutus SSIS 2005 komponenteilla

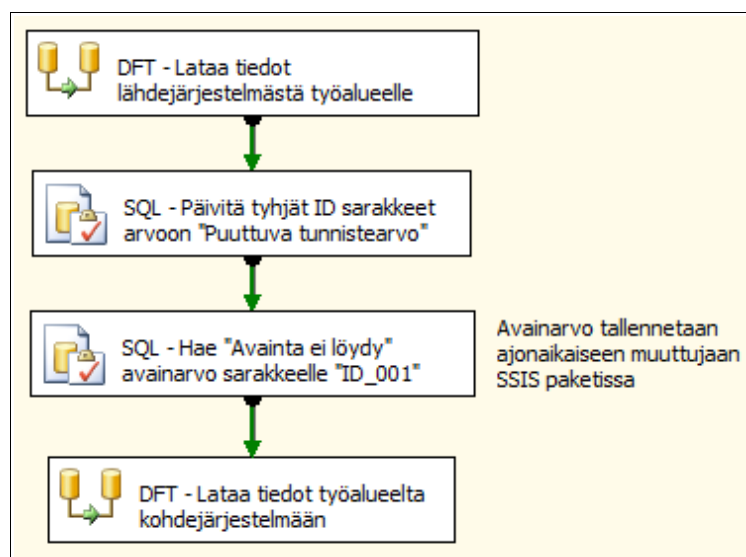
Avaimenhaun toteutus SSIS 2005 -komponenteilla vaatii usean komponentin määrittämistä niin ohjaus- kuin tietovuo rajapinnoille. Ohjausvuorajapinnalla haetaan avaimenhaussa tarvittavat ohjaustiedot ja tietovuorajapinnalla suoritetaan itse avaimenhaku. (Compsol31200 2007.)

Avaimenhaussa tarvitaan ohjausvuon komponentteja, koska eBIC-järjestelmän ETL-prosessille on määritelty, että se suorittaa '?'-ja '??'-haut tarvittaessa. Jos '?'-ja '??'-hakuja ei suoritettaisi, voitaisiin lookup tehdä pelkästään tietovuorajapinnan komponenteilla. (Compsol31200 2007.)

6.1.1 Ohjausvuon komponentit

Kuviossa 22 on esitetty yksinkertaistettu tapaus ohjausvuosta ja siinä olevista komponenteista, joita tarvitaan yhden avaimenhaun toteuttamiseen. Kuvion 22 ohjausvuossa on kaksi tietovuokomponenttia ”DFT – Lataa tiedot lähdejärjestelmästä työalueelle”, joka lataa tiedot eBIC-järjestelmän työalueelle lähdejärjestelmästä. Tämä komponentti vain hakee tiedot, joille tehdään avaimenhaku, joten tässä ei keskitytä sen toteutukseen tarkemmalla tasolla. Toinen tietovuokomponentti ”DFT – Lataa tiedot työalueelta kohdejärjestelmään” suorittaa avaimenhaun ja siirtää tiedot kohdejärjestelmään, tässä tapauksessa eBIC-järjestelmään. Tämän tietovuokomponentin toteutus käydään tarkemmin läpi kappaleessa 6.1.2.

Kuvion 22 ohjausvuon SQL – komponentit alustavat tietoja avaimenhakua varten ja hakevat tarvittavia ohjaustietoja. Ensimmäisen suoritettava SQL – komponentti ”SQL- Päivitä tyhjät ID sarakkeet arvoon 'Puuttuva tunniste arvo’” päivittää kaikki tyhjät ID sarakkeet arvoon '?'. Päivitys tehdään, koska ETL-prosessissa on määritetty, että puuttuvaa tunniste arvoa haetaan arvolla '?'. Ohjausvuon toinen SQL – komponentti hakee avainarvon tuntemattomalle eli '??-arvolle. Arvo haetaan, koska ETL-prosessissa on määritetty, että tuntemattomat arvot päivitetään '??'-arvoon. (Compsol31200 2007.)

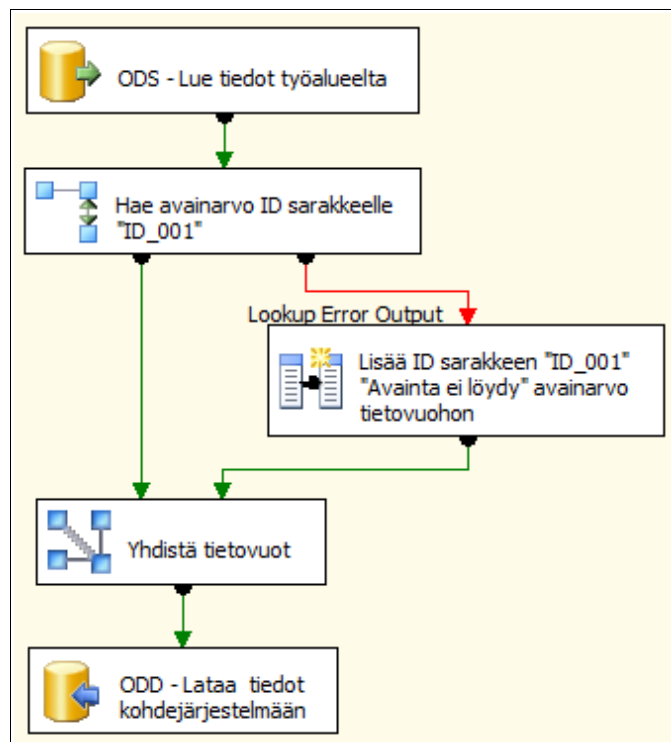


KUVIO 22. Ohjausvuon komponentit SSIS 2005 avaimenhakussa

Kuvion 22 SQL-komponentti ”SQL – Hae 'Avainta ei löydy' avainarvo sarakkeelle 'ID_001’” tallentaa hakemansa arvon SSIS 2005 paketin ajonaikaiseen muuttujaan. Esimerkissämme olemme nimenneet muuttujan ”ID_001_Avainta_ei_Löydy_avain” nimiseksi. Tieto tallennetaan ajonaikaiseen muuttujaan, jotta voisimme tietovuossa lukea kyseisen arvon tarvittaessa avaimenhakuoperaatiossa. Käsittelemme kyseisen muuttujan käyttöä tarkemmin luvussa 6.1.2.

6.1.2 Tietovuon komponentit

Tietovuossa olevat komponentit suorittavat avaimenhakuoperaation. Avaimenhakuoperaatio on kuvattu kuviossa 23. Kuvan tietovuo on yksinkertaistettu, jotta avaimenhakuprosessi voitaisiin kuvata selkeämmin. Kuviossa 23 on suoritettu avaimenhakuoperaatio ID sarakkeelle ”ID_001”.

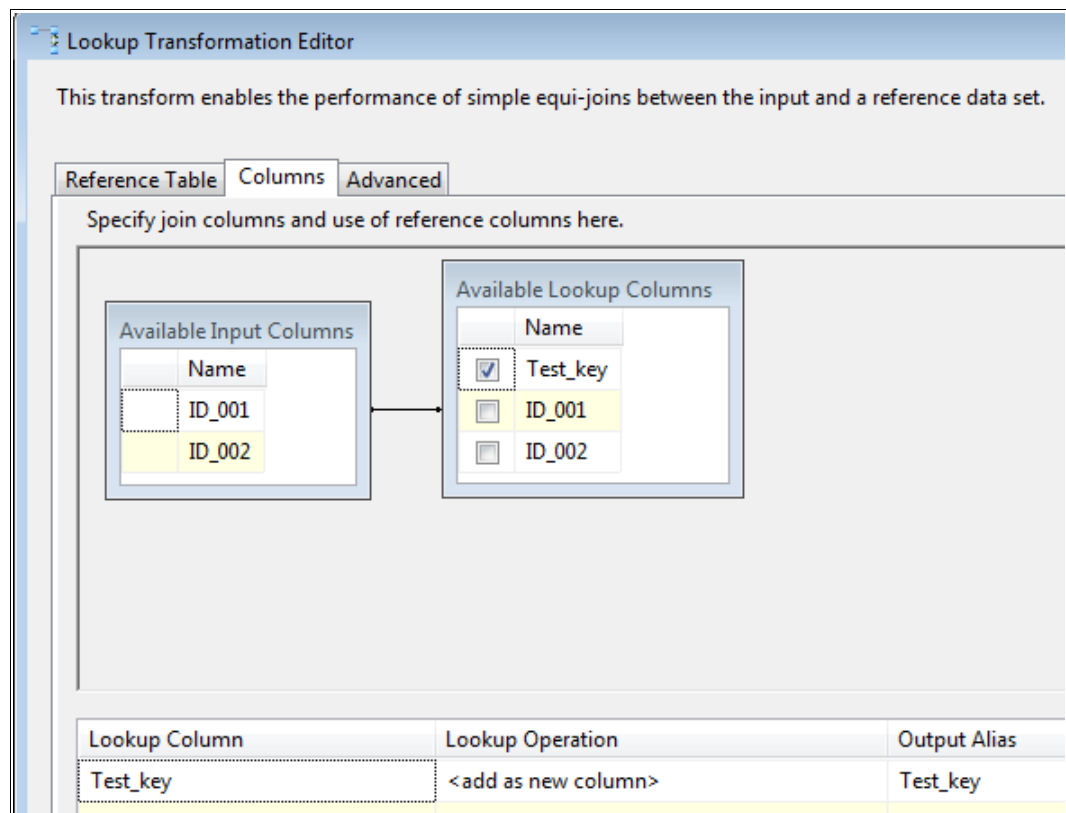


KUVIO 23. Tietovuon komponentit SSIS 2005 avaimenhaussa

Kuvion 23 komponentti ”ODS- Lue tiedot työalueelta” hakee tiedot eBIC-järjestelmän työalueelta tietovuoputkeen. Huomaa, että ohjausvuo on tässä vaiheessa päivittänyt tietokannan työalueella olevat tyhjät sarakkeet arvoon ’?’. Komponentti ei ole avaimenhauksen toteutuksen kannalta tärkeä, eikä se keskity tässä sen toteutukseen tarkemmin.

Komponentti ”Hae avainarvo ID sarakkeelle 'ID_001'” suorittaa avaimenhauksen ID sarakkeelle 'ID_001'. Komponentti on SSIS 2005 ohjelman avaimenhaku-tietovuokomponentti, joka tulee SSIS 2005 -ohjelman mukana. Avaimenhakukomponentti

toimii kappaleessa 3 mainituilla periaatteilla. Kuviossa 24 on havainnollistettu avaimenhakukomponentin avaimenhakuoperaatio. Avaimenhakuoperaatio hakee tietokannasta ”Test_key”-avaimen vertaamalla tietovuon ID saraketta ”ID_001” tietokannan ID sarakkeeseen ”ID_001”. Haettu avain lisätään tietovuohon uutena sarakkeena.



KUVIO 24. SSIS 2005 avaimenhakukomponentin asetuksia.

Jos kuviossa 24 esitetty avaimenhakuoperaatio epäonnistuu, ohjaa avaimenhakukomponentti etsityn arvon komponentin virhevuohon. Virhevuon on esitetty kuviossa 23 nimellä ”Lookup error output”.

Avaimenhakukomponentti ohjaa tiedon joko tietovuohon tai virhevuohon. Komponentin ohjatessa tiedon tietovuohon lisätoimet eivät ole tarpeen. Sen sijaan jos komponentti ohjaa tiedon virhevuohon tarvitaan lisäoperaatio. Kyseinen lisäoperaatio on lisätä puuttuva tunniste arvo tietovuohon. Kyseisellä operaatiolla saadaan

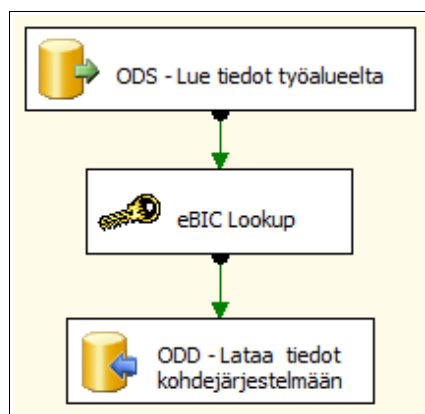
aikaiseksi tilanne, jossa avaimenhakukomponentilta saadaan palautettua aina arvo.

Kuviossa 23 esitetty komponentti ”yhdistä tietovuot” yhdistää avaimenhakukomponentilta tulevan tietovuon ja tietovuon, joka sisältää puuttuvan tunniste- arvon avaimen. Komponentti toimii siten, että se ohjaa oletuksena avaimenhaku komponentilta tulevat tiedot eteenpäin. Jos tietovuosta, joka sisältää puuttuvan tunniste- arvon avaimen, tulee tieto, asetetaan se avaimenhakukomponentilta tulevaan avain-sarakkeeseen.

6.2 Avaimenhaun toteutus itse tehdyllä tietovuokomponentilla

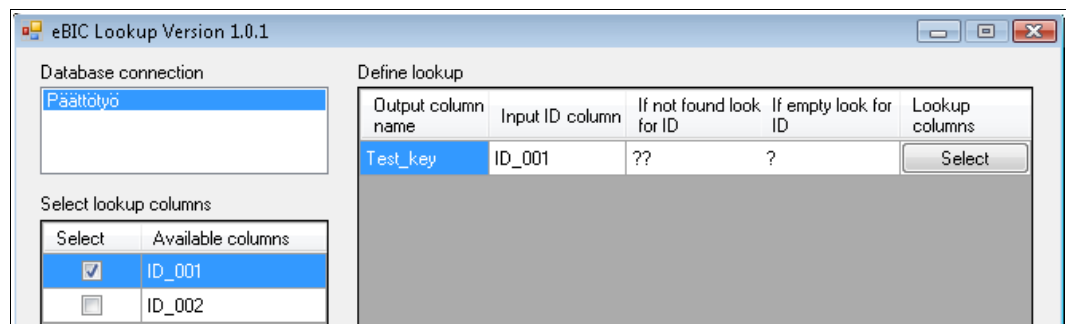
Avaimenhaku itse tehdyllä tietovuokomponentilla voidaan toteuttaa yhdellä tietovuon komponentilla. Avaimenhakuoperaatio määritellään luvussa 5.4 kuvatuilla käyttöliittymän kentillä.

Kuviossa 25 on kuvattu avaimenhaun toteutus tietovuossa. Avaimenhakukomponentti suorittaa '?' ja '??' käsittelyn avaimenhaun lisäksi, joten muita komponentteja ei tarvita.



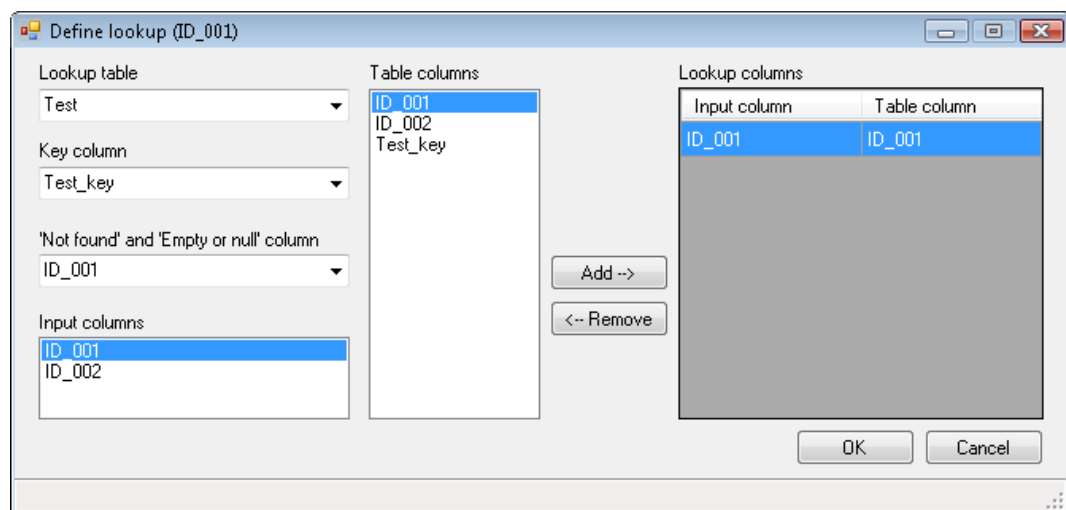
KUVIO 25. Tietovuon komponentit itse toteutetulla tietovuokomponentilla

Kuviossa 26 on esitetty osa kuvion 25 ”eBIC lookup” tietovuokomponentin perusnäkymästä. Kuviossa 26 käyttäjä on valinnut sarakkeen ”ID_001” avaimenhakusarakkeeksi. Sarakkeeseen ”If not found look for ID” käyttäjä on määritellyt, että ”Tunnistearvo, jolla ei löydetä surrogaattivainta” on ”??”. Vastaavasti sarakkeeseen ”If empty look for ID” käyttäjä on määritellyt, että ”Puuttuva tunnistearvo” on ”?”. Käyttäjä voi määrittellä sarakkaisiin myös muut arvot tarvittaessa.



KUVIO 26. Osa eBIC lookup avaimenhakukomponentin perusnäkymästä

Kuviossa 27 nähdään Kuviossa 26 kuvatun avaimenhaun määrittelmä. Avaimenhaku on vastaava luvussa 7.1 toteutetun avaimenhaun kanssa.



KUVIO 27. Avaimenhaun määrittely

7 YHTEENVETO

Tämän työn tavoitteena oli yksinkertaistaa avaimenhakua SSIS 2005 -ohjelmassa itse tehdyllä tietovuokomponentilla. Tarve avaimenhaun yksinkertaistamiseen lähti toimiyksikön CompSol31200 siirtymisestä SSIS 2005 -järjestelmään eBIC-konseptituotteen ETL – prosessissa.

Toteutettu tietovuokomponentti eBIC lookup on otettu käyttöön useassa asiakas-ympäristössä, ja komponentin toiminnallisuus vastaa sille asetettuja tavoitteita. Komponentin käytöstä kerättyjen kokemusten perusteella avaimenhaun yksinkertaistaminen on onnistunut. Käyttäjät ovat kiittäneet komponentin helppokäyttöisyyttä verrattuna SSIS 2005 -komponenteilla toteutettuun vaihtoehtoon.

eBIC lookup -komponentti on vähentänyt avaimenhaun toteuttamiseen tarvittavaa työmäärä verrattaessa SSIS 2005 -komponenteilla toteutettuun avaimenhakuun. Jos SSIS 2005 -paketissa tarvitaan esimerkiksi kymmenen eri avaimenhakua, on pie-neneminen työmäärässä huomattava. Tämä sen takia, että SSIS 2005 -komponenteilla toteutettu avaimenhaku vaatii neljä komponenttia yhtä avaimenhakua varten. eBIC lookup -komponentti sen sijaan voi toteuttaa kaikki avaimenhaut samassa komponentissa. eBIC lookup -komponentti vähentää tehdyn työmäärän neljäsosaan verratessa sitä SSIS 2005 -komponenteilla toteutettuun avaimenhakuun.

Avaimenhaun yksinkertaistaminen on toteutunut tarvittavien komponenttien vähentyneellä määrällä ja sillä, että avaimenhaku voidaan toteuttaa kokonaan yhdellä komponentilla. Toteutetun komponentin käyttöliittymän toteutus on myös osaltaan yksinkertaistanut avaimenhakua. Komponentin käyttöliittymä on toteutettu toimiyksikkö Compsol31200 toiveiden mukaan, ja se näkyy käyttäjiltä tullessa palautteessa. Käyttäjät pitävät käyttöliittymää toimivana ja helppokäyttöisenä.

eBIC lookup -komponentin kehitystyön aloittaminen oli teknisesti haastava operaatio, koska dokumentaatiota kyseisestä aiheesta oli saatavilla niukasti. Kehitystyö kuitenkin saatiin vietyä läpi useista takaiskuista ja pitkittyneestä aikataulusta huolimatta. Kehitystyö olisi ollut helpompi, jos aiheesta olisi ollut saatavilla laaja ja järjestelmällinen dokumentaatio.

Toteutettua komponenttia voisi vielä jatkokehittää esimerkiksi muuttamalla käyttöliittymän toiminnallisuutta. Nykyisessä toteutuksessa käyttöliittymän kokoa ei pysty vaihtamaan, mikä joissain tapauksissa hankaloittaa komponentin käyttämistä.

LÄHTEET

LogicaCMG 2007, LogicaCMG yritysesittely [Verkkodokumentti], [viitattu 21.10.2007]. Saatavissa: <http://www.wmdata.fi/>

SSIS 2005 esittely 2007, SSIS 2005 ohjelman esittely [Verkkodokumentti], [viitattu 4.12.2007]. Saatavissa:

<http://www.microsoft.com/technet/prodtechnol/sql/2005/intro2is.msp>

Compsol31200 2007, LogicaCMG kustannuspaikan Compsol31200 parhaat käytännöt.