KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
INFORMACIJOS SISTEMŲ KATEDRA

Kristina Paulavičiūtė

Transformacijų šablonais grindžiamas duomenų saugyklos projektavimo procesas

Magistro darbas

Vadovas
doc. dr.L. Nemuraitė

KAUNAS, 2006

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
INFORMACIJOS SISTEMŲ KATEDRA

Transformacijų šablonais grindžiamas duomenų saugyklos projektavimo procesas

Magistro darbas

Vadovas
doc. dr. L. Nemuraitė

Recenzentas
doc. dr. S. Maciulevičius

Atliko
IFM 0/4 gr. stud.
K. Paulavičiūtė

KAUNAS, 2006

Summary

Data warehouse building process based on data transformation templates.

Growing amount of data and needs of data analysis starts needs of data warehouses. A lot of organizations operational data cumulates in OLTP DBMS databases. Organization historical data are cumulating in data warehouses. These data are adjusted for data analysis. DBMS ETL tools don't have good data warehouse building opportunities. Created ETL tool for MS SQL Server makes data warehouse building process easier and speedier.

Turinys

Lentelių sąrašas

Paveikslėlių sąrašas

1. Įvadas

Žmonija jau senai suvokė informacijos kaupimo ir saugojimo svarbą. Buvo kuriamos specialios informacijos saugyklos: bibliotekos, muziejai, fonotekos. Informacinių technologijų atsiradimas ir plėtra suteikė galimybę kaupti duomenis kompiuteriuose. Tai leido supaprastinti duomenų kaupimą, apdorojimą, atnaujinimą.

Tačiau pradėjus vystytis informacinėms technologijoms atsirado didelių informacijos kiekių apdorojimo problema. Sukaupiamų duomenų kiekiai didėja daug greičiau nei vystosi duomenų saugojimo technologijos. Fiziškai neįmanoma sukaupti ir saugoti visus bet kokios organizacijos duomenis. Praktiškai neįmanoma išanalizuoti tokių didelių duomenų kiekių įvairiais pjūviais, nes šios operacijos būtų per brangios, užimtų per daug laiko ir kai kuriais atvejais galėtų turėti neigiamų padarinių visai veikiančiai sistemai. Dar viena problema yra ta, kad analizei reikalingi duomenys dažniausia yra kaupiami skirtinguose formatuose (operacinių duomenų transakcijų apdorojimo sistemose OLTP (*angl*. *On-line Transaction Processing*), liktinėse sistemose, tekstiniuose failuose, MS Excel failuose ir t.t.) ir skirtingose fizinėse vietose.

Šias, duomenų kaupimo, analizės, surinkimo iš skirtingų šaltinių, problemas sprendžia duomenų saugyklos (angl. data warehouse). Jose duomenys analizuojami naudojant specialią programinę įrangą – OLAP (*angl*. *On-Line Analytical Processing*) sistemas.

Duomenų saugyklos autoriumi laikomas B. Inmon, kuris apibūdino duomenų saugyklas taip: „Duomenų saugykla yra į analizės sritis (temas) orientuotas (*angl*. *subject oriented*), integruotas, laiko chronologijos tvarka išdėstytas nemodifikuojamų duomenų rinkinys, skirtas valdymo sprendimams paremti“. R. Kimball duomenų saugyklą apibūdina taip: „Duomenų saugykla yra specialiai analizei struktūrizuotų operacinių duomenų kopija“.

Duomenų saugyklos kūrimo procesas susideda iš: duomenų šaltinio identifikavimo, saugyklos projektavimo ir duomenų transformavimo ir perkėlimo. Saugyklos kūrimo procesas yra gana ilgas, o norint turėti kuo efektyvesnę duomenų analizę, reikia trumpinti šitą procesą.

Duomenų transformavimas ir perkėlimas į duomenų saugyklą, jų atnaujinimas vykdomas naudojant ETL (a*ngl. extract/tranform/load - Išgauk/Transformuok/Įdėk*) programines priemones.

Duomenų saugyklos projektavimas – tai sudėtingiausias duomenų saugyklos kūrimo etapas. Jis apima duomenų saugyklos schemos ir duomenų transformacijų projektavimą. Duomenų saugyklos schemos tipas priklauso nuo saugykloje vykdomų užklausų. Tuo tarpu duomenų transformacijos priklauso nuo daugelio dalykų: kokius duomenis vartotojas nori matyti saugykloje, kokio detalumo tie

duomenys turi būti ir t.t. Todėl duomenų transformacijų kūrimas yra pakankamai sudėtingas ir ilgai trunkantis.

Išanalizavus ETL priemonių rinką, tapo aišku, kad ETL priemonės, kurios duomenų saugyklos kūrimo procesą galėtų pagreitinti ir palengvinti yra brangios, o populiariausių duomenų bazių valdymo sistemų (DBVS) siūlomos priemonės neturi tokių galimybių. Todėl darbe siekiama patobulinti DBVS ETL priemonę, panaudojant atskirai platinamų ETL priemonių patirtį.

Iškeltas toks darbo tikslas: Sukurti ETL priemonę, pritaikytą automatizuotam duomenų saugyklos kūrimui MS SQL priemonėmis.

Norint pasiekti iškeltą tikslą, buvo atlikta probleminės srities analizė, kuriamos ETL priemonės projektavimas ir realizacija. Darbo eiga ir rezultatai pateikiami dokumente.

Darbo struktūra: analitinė dalis, eksperimentinis duomenų saugyklos kūrimas esamomis priemonėmis, kuriamos ETL sistemos projektavimas, realizacija, sukurtos ETL priemonės įvertinimas.

- Analitinėje darbo dalyje apžvelgiama duomenų saugyklos struktūra, jos kūrimo procesas. Detaliai išanalizuojami duomenų transformavimo operacijų tipai, nustatomos pagrindinės, duomenų saugyklos kūrimui reikalingos, transformacijos. Apžvelgus ETL priemonių rinką, išanalizuojami ETL priemonių skirtumai. Atlikus analizę, nustatomas darbo tikslas - Sukurti ETL priemonę, pritaikytą automatizuotam duomenų saugyklos kūrimui MS SQL priemonėmis. Numatytas darbo tikslas bus pasiektas, jeigu pavyks sukurti ETL priemonę, automatizuojančią duomenų saugyklos kūrimą MS SQL priemonėmis. Šiuo metu norint kurti duomenų saugyklą MS SQL reikia rašyti transformavimo programinį kodą.
- Eksperimentinėje dalyje, norint įsigilinti į išanalizuotus duomenų transformavimo operacijų tipus, realizuotas duomenų saugyklos kūrimas. Duomenų saugyklos kūrimas atliktas panaudojant esamas MS SQL priemones. Duomenų saugyklos kūrimui pasirinkta reliacinė duomenų bazė, kurioje saugomi duomenys apie visų Lietuvos aukštųjų mokyklų dėstytojus, studentus, dėstomus modulius bei studentų įvertinimus. Sukurta duomenų saugykla, skirta atlikti studentų pažangumo analizę. Atliktas duomenų saugyklos kūrimas padėjo patikslinti analizės metu aptartus transformavimo operacijų tipus ir sukurti dažniausiai naudojamų duomenų transformacijų šablonus.
- Remiantis analizės dalies duomenimis ir atliktu eksperimentiniu duomenų saugyklos kūrimu, suprojektuota sistema, skirta automatizuotam duomenų saugyklos kūrimui. Išskirtos sistemos funkcijos, suprojektuota sistemos duomenų bazė, kurioje bus saugomi duomenų šaltinių ir duomenų saugyklų metaduomenys ir transformacijos.

- Realizacijos dalyje pateikiamas detalus sukurtos ETL priemonės aprašymas ir šia priemone atliktas eksperimentinis duomenų saugyklos kūrimas.
- Darbas užbaigiamas sukurtos ETL priemonės palyginimu su MS SQL: išskiriami sukurtos ETL priemonės duomenų saugyklos kūrimo privalumai lyginant su MS SQL. Sukurta ETL priemonė realizuoja tik pagrindines duomenų saugyklos kūrimui reikalingas transformacijas ir skirta sukurti tik žvaigždės schemos saugyklą, todėl darbas baigiamas sukurtos ETL priemonės plėtimo galimybių ir perspektyvų aptarimu.

Darbo analitinė dalis ir numatomi rezultatai buvo pristatyti 10 – ojoje tarpuniversitetinėje magistrantų ir doktorantų konferencijoje „Informacinės technologijos“ ir išspausdintas straipsnis (K. Paulavičiūtė, 2005).

2. Transformacijų šablonais grindžiamo duomenų saugyklos projektavimo proceso analizė

2.1. *Analizės tikslas*

Analitinėje dalyje apžvelgiama duomenų saugyklos struktūra, jos kūrimo procesas. Detaliai analizuojami duomenų transformavimo operacijų tipai ir jų reikalingumas kuriant duomenų saugyklą. Išanalizavus ETL priemonių rinką, išsiaiškinti atskirai platinamų ETL priemonių privalumai lyginant su kartu su DBVS platinamomis priemonėmis. Išanalizuota MS SQL ETL priemonė, ir galimybė ją patobulinti pirmaujančių ETL priemonių pavyzdžiu.

2.2. *Tyrimo sritis, objektas ir problema*

Tyrimo sritis: Verslo informacijos saugyklos (*angl. data warehouse*) kūrimo procesas.

Tyrimo objektas: ETL duomenų transformacijos MS SQL sistemoje ir jų tobulinimo galimybės.

Sprendžiama problema: Duomenų saugyklų kūrimo procesas trunka per ilgai ir negali patenkinti nuolat kintančių duomenų analizės poreikių. Pavienės ETL priemonės šiuo požiūriu lenkia universalių, plačiai naudojamų DBVS ETL priemones. Duomenų saugyklos projektavimas – tai sudėtingiausias duomenų saugyklos kūrimo etapas. Norint suprastinti ir sutrumpinti šitą etapą, pirmiausia reikia išanalizuoti saugyklų schemų tipus, duomenų transformavimo operacijas. Norint išbandyti duomenų transformavimo operacijas reikia atlikti duomenų saugyklos kūrimą, panaudojant esamas priemones duomenų transformacijų kūrimui. Realizavus duomenų saugyklos kūrimą, bus galima apibrėžti duomenų transformacijų operacijų vykdymo šablonus ir juos panaudoti tobulinant transformacijų kūrimą. Greičiausiai nebus įmanoma rasti visų reikalingų duomenų transformacijų operacijų, bet nustatytos operacijos – bus būdingiausios duomenų saugyklų kūrimo procesuose. Sukurti duomenų transformacijų šablonai ir galbūt jų panaudojimo grafinė vartotojo sąsaja, žymiai pagreitins duomenų saugyklų kūrimo procesą.

2.3. *Duomenų saugyklos sąvokos, elementų ir kūrimo proceso analizė*

2.3.1. Duomenų saugyklos sąvoka

Terminą Duomenų saugykla (*angl*. *Data Warehouse*) 1990 metais įvedė B. Inmon. Jis duomenų saugyklą apibrėžė: „Duomenų saugykla yra į veiklos sritis orientuotų, integruotų ir nekintančių, turinčių laiko matą duomenų rinkinys, naudojamas sprendimų priėmimo procese“. B. Inmon įvardino šiuos duomenų saugykloje saugomų duomenų požymius:

- **į veiklos objektus orientuoti:** duomenys suteikiantys informacijos apie tam tikrą sistemos (įmonės, organizacijos) objektą, o ne apie sistemos vykdomas operacijas.
- **integruoti:** duomenys surenkami iš įvairių šaltinių į vieną saugyklą ir joje sudaro prasmingą visumą.
- **turintys laiko matą:** saugykloje saugomos kintančių laike duomenų apibendrintos reikšmės, todėl duomenys turi laiko identifikatorių (pavyzdžiui periodą).
- **nekintantys:** duomenys saugykloje nekinta. Gali būti įdedama naujų duomenų, tačiau saugomų duomenų faktai nemodifikuojami.

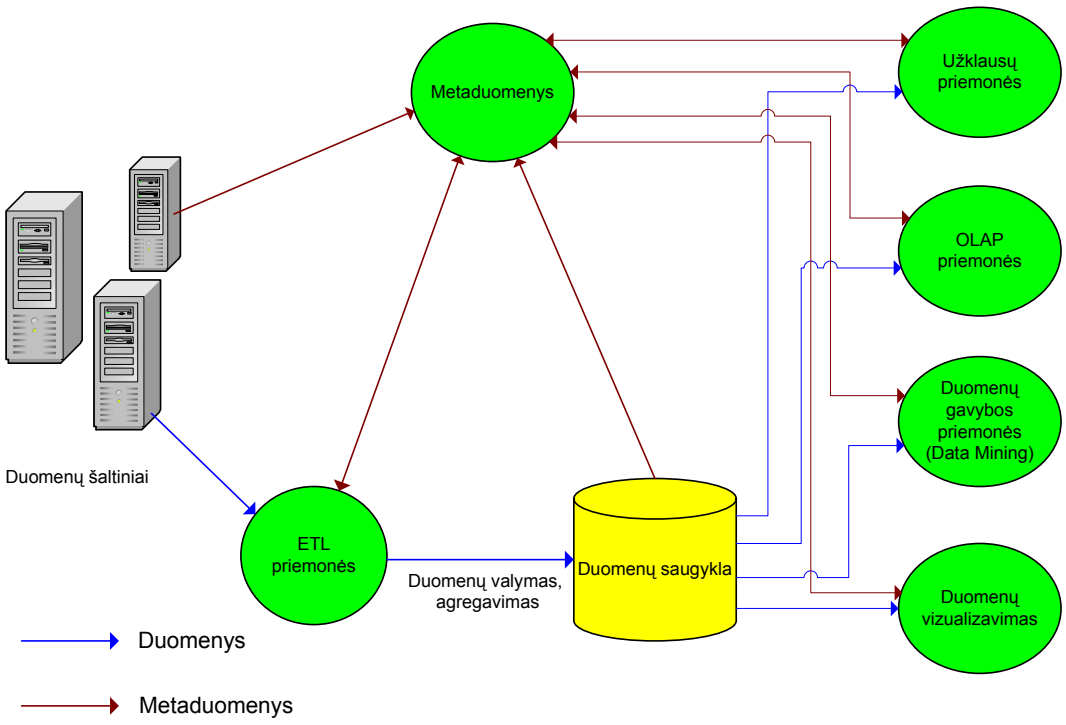
Nors šis apibrėžimas paskelbtas beveik prieš 10 metų, jis ir dabar gana tiksliai charakterizuoja duomenų saugyklą.

Kiek paprastesnį apibrėžimą yra pateikęs R. Kimball: „Duomenų saugykla yra specialiai analizei struktūrizuotų operacinių duomenų kopija“.

Abudu apibrėžimai nėra griežti: pavyzdžiui, duomenys iš duomenų saugyklos gali būti ir ištrinami, dėl per didelio jų kiekio ir brangios saugojimo terpės.

2.3.2. Duomenų saugyklos komponentai ir kūrimo procesas

Duomenų saugyklos komponentai pateikti 1 paveiksle.



1 pav. Duomenų saugyklos sudedamosios dalys

Duomenų saugyklos sukūrimo procesas (*angl. Data Warehousing*) susideda iš:

- **duomenų saugyklos kūrimo poreikių nustatymas.** Analizės poreikiai ateina iš organizacijos narių, kuriems reikalinga vienų ar kitų organizacijos duomenų analizė skirtingais pjūviais.
- **duomenų šaltinio identifikavimas.** Duomenų saugyklos kūrimui reikia turėti reikiamus duomenis. Dažniausiai imami „istoriniai“ ankstesnių periodų duomenys, taip pat kiti reikalingi duomenys, kurie gali būti senose „liktinėse“ sistemose, tekstiniuose ar kituose failuose. Išoriniuose šaltiniuose esančių duomenų išgavimas dažniausiai būna procesas.
- **saugyklos ir duomenų transformavimo operacijų projektavimas.** Tai procesas, kurio metu kuriama saugyklos schema ir duomenų transformavimo. Priklausomai nuo to kokios užklausos saugykloje bus vykdomos, parenkamas daugiadimencinės struktūros tipas. Priklausomai nuo to, kokių duomenų reikia vartotojui, numatomos duomenų transformacijos. Dažniausiai šis žingsnis atliekamas iteracijomis. Vieną kartą sukūrus duomenų modelį ir jį užpildžius dideliais duomenų kiekiais vėliau būna labai sunku, o kartais ir neįmanoma tą modelį keisti. Dažniausiai šis žingsnis yra brangiausias ir ilgiausiai trunkantis.
- **duomenų transformavimas ir perkėlimas.** Tai saugyklos kūrimo, duomenų transformavimo ir perkėlimo iš šaltinio į saugyklą procesas. Naudojamos ETL (*Extract/Transform/Load*) (*Išgauk/Transformuok/Įdėk*) programinės priemonės.
- **pakitimų sekimas.** Periodinis saugyklos atnaujinimas duomenimis iš operacinės aplinkos. Problemos kyla sekant kuriuos duomenis reikia atnaujinti. Ne visose komercinėse sistemose ši problema sėkmingai išspręsta.

2.4.*Duomenų saugyklos kūrimo poreikių nustatymas*

OLTP duomenų bazės, naudojamos einamuosiuose organizacijos procesuose ir sistemose, negali patenkinti duomenų analizės poreikių. Analizei naudojant duomenis, esančius OLTP duomenų bazėse, analizė vyksta labai lėtai, taip pat apkraunamos einamosios programos, kas gali sulėtinti ar netgi neigiamai paveikti einamuosius sistemos procesus. Norint patenkinti organizacijos analizės poreikius, kuriamos duomenų saugyklos. Į duomenų saugyklą perkeliami analizei reikalingi organizacijos duomenys iš OLTP sistemų. Priklausomai, analizuojamų duomenų kiekio, analizės pjūvių, duomenų saugyklos kūrimui parenkamos skirtingos schemos.

2.5.*Duomenų saugyklos schemos pasirinkimas*

Duomenų saugyklos remiasi daugiamačiu duomenų modeliu. Daugiamačiu duomenų modeliu vadinamas toks optimizuotas modelis, kuris naudojamas kuriant duomenų saugyklas ir OLAP kubus. Pagrindinės daugiamačio duomenų modelio dalys yra fakto ir dimensijų lentelės. Fakto lentelėje paprastai saugomi duomenys apie objektus ir įvykius.

Dimensijų lentelės turi nekeičiamus arba retai keičiamus duomenis ir nusako fakto lentelėje saugomų duomenų analizės pjūvius. Kiekviena dimensijų lentelė turi būti susieta su faktų lentele „ vienas su daug“ ryšiu.

Galimos tokios duomenų saugyklos schemos:

- Plokštės schema – pati paprasčiausia, galima be informacijos nuostolių. Suformuojama sujungiant visas duomenų modelio esybes taip, kad gautųsi minimalus esybių skaičius.
- Terasos schema – kiekvienam operaciniam objektui kuriama po vieną lentelę.
- Žvaigždės schema – kiekviena dimensija saugoma vienoje lentelėje.
- Snaigės schema – žvaigždės schemos modifikacija, kai dimensija saugoma keliose lentelėse.
- Žvaigždžių spiečiaus schema.

Sudėtingiausia yra snaigės schema, o pati paprasčiausia – plokštės. Pasirinkdamas vieną ar kitą schemos tipą, naudotojas turi rinktis tarp lentelių skaičiaus (struktūros sudėtingumo) ir faktų skaičiaus (elementų sudėtingumo). Schemos atnaujinimo ir lankstumo požiūriu efektyvesnė yra normalizuota (struktūriškai sudėtingiausia) snaigės schema, naudotojo užklausų paprastumo ir greičio požiūriu – paprasčiausia terasos schema ar tarpiniai variantai.

2.6.*Duomenų transformavimo operacijų analizė*

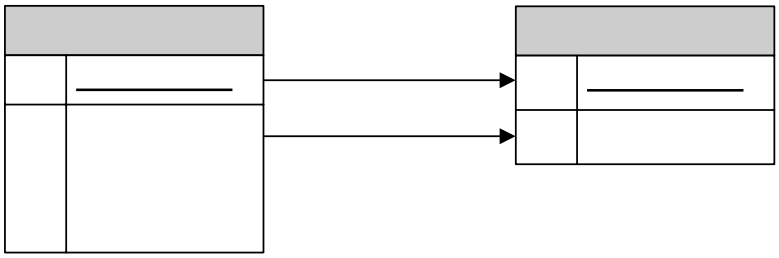
Duomenų perkėlimui iš OLTP sistemų į duomenų saugyklas yra kuriamos duomenų transformacijos, kurias sudaro duomenų transformavimo operacijos.

Duomenų transformcijų kūrimas – tai sudėtingiausias duomenų saugyklos kūrimo etapas. Organizacijos duomenų modelį transformuojant į duomenų saugyklos duomenų modelį, galima atlikti tokais transformavimo operacijas (L. Silverston, 2001):

- operacinių duomenų pašalinimas,
- duomenų saugyklos rakto struktūros papildymas laiko elementu,
- struktūros papildymas išvestiniais (*angl. derive*) duomenimis,
- duomenų ryšių transformavimas į ryšių artefaktus (*angl. artifacts*),
- skirtingų duomenų detalumo lygių pritaikymas duomenų saugykloje,
- skirtingų lentelių suliejimas,
- duomenų masyvų sukūrimas,
- duomenų atskyrimas pagal jų kintamumą.

2.6.1. Operacinių duomenų pašalinimas

Kuriant duomenų transformacijas pirmiausia reikia peržiūrėti organizacijos duomenų modelį ir pašalinti operacinius (*angl. operational*) duomenis. Reikalingi duomenys iš organizacijos duomenų modelio perkeliami į duomenų saugyklos modelį. Pagal 2 paveiksle pateiktą pavyzdį, į duomenų saugyklos modelį perkeliama „Sąskaitos_Nr“ ir „Sąskaitos_data“. Kiti organizacijos duomenų modelio elementai: „Apibūdinimas“, „Pastaba“, „Statusas“, nėra reikalingi duomenų saugyklos modelyje.

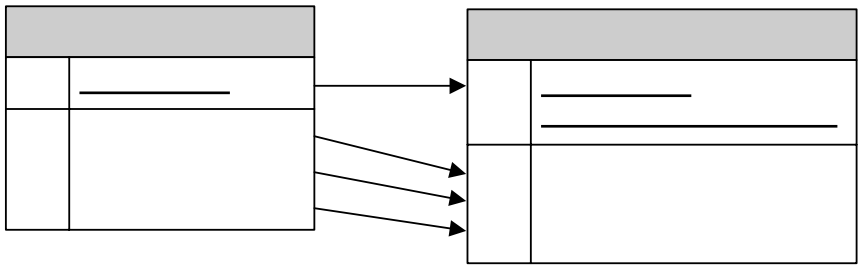


2 pav. Operacinių duomenų pašalinimas

Kiekvienąkart, sprendžiant duomenų pašalinimo klausimą, reikia atsižvelgti į duomenų naudojamumą duomenų saugykloje: perkeliami tik tie duomenys, kurie bus naudojami.

2.6.2. Laiko elemento pridėjimas

Jeigu organizacijos duomenų modelyje nėra laiko elemento, duomenų saugyklos modelyje jį būtina sukurti. Laiko elementas duomenų saugykloje paprastai įeina į raktą. Organizacijos duomenų modelyje saugant informaciją apie asmenį (3 pav.), raktas yra tik asmens kodas („Asmens_AK“). Tuo tarpu duomenų saugykloje raktas susidaro iš dviejų atributų: asmens kodo („Asmuo_AK“) ir duomenų įrašymo į saugyklą datos („Duomenų_įrašymo_data“). Datos dėka yra galimybė sužinoti, kuriuo metu kokie duomenys apie asmenį buvo įvesti.

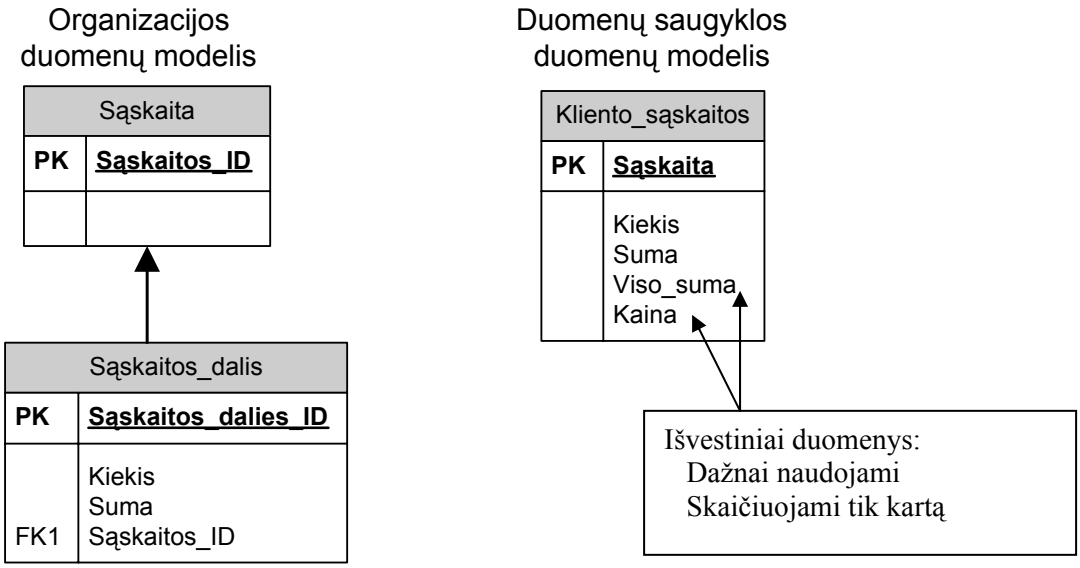


3 pav. Laiko elemento pridėjimas duomenų saugyklos modelyje

Toks laiko elemento įtraukimas į raktą – vienas dažniausių, laiko elemento įtraukimo į duomenų saugyklą, būdų. Kitas būdas – į rakto struktūrą įtraukti laiko periodą žyminčius elementus (pvz., data nuo, data iki). Šitas būdas ypač tinkamas, kai reikia saugoti tęstinius duomenis.

2.6.3. Išvestinių duomenų pridėjimas

Jeigu reikia, duomenų saugyklos modelį galima papildyti išvestiniais duomenimis. Pateiktame pavyzdyje (4 pav.) transformuojant duomenis iš organizacijos duomenų modelio į duomenų saugyklos modelį, papildomai pridedami du skaičiuojami laukai: „Viso_suma“ ir „Kaina“.



4 pav. Išvestinių duomenų pridėjimas

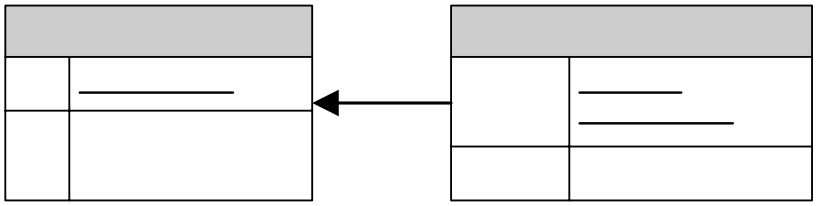
Taisyklingame organizacijos duomenų modelyje išvestiniai duomenys nenaudojami. Į duomenų saugyklos modelį būdinga įtraukti išvestinius duomenis. Išvestinių duomenų naudojimas pagreitina duomenų iš duomenų saugyklos ištraukimą.

Duomenų saugyklą papildant išvestiniais duomenimis, reikia tiksliai žinoti tų duomenų paskirtį. Priešingu atveju, duomenų saugyklos dydis gali labai greitai išaugti, o tai sulėtintų duomenų saugyklos naudojimą.

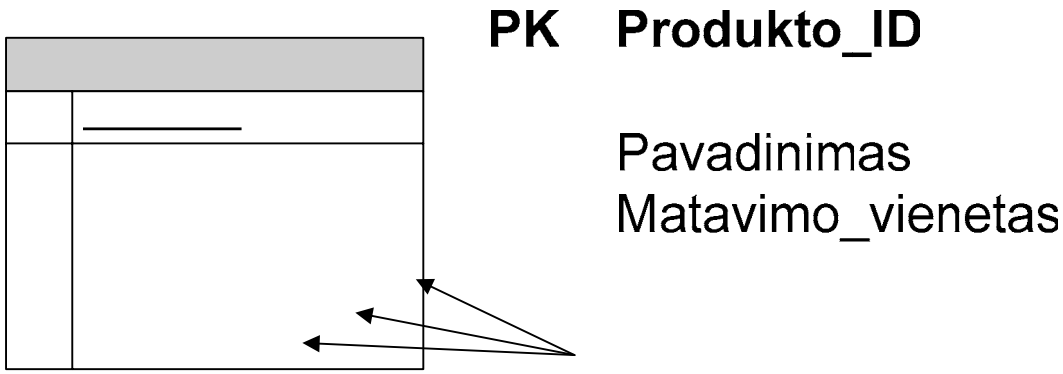
2.6.4. Ryšių artefaktų sukūrimas

Duomenų ryšiai klasikiniame duomenų modeliavimo procese nusako, kad yra tik viena pagrindinė ryšio reikšmė, t.y. yra tik vienas pagrindinis produkto tiekėjas (5 pav.).

Ryšys duomenų saugykloje dažniausiai turi daug reikšmių, nes duomenys saugykloje yra ilgo periodo (yra daug produkto tiekėjų per ilgą laiką). Todėl organizacijos duomenų modelio klasikinis ryšys tarp lentelių nėra adekvatus duomenų saugyklos ryšiui. Ryšiai tarp lentelių duomenų saugykloje yra sukuriami naudojant artefaktus (6 pav.).



5 pav. Operacinis ryšys tarp produkto ir tiekėjo organizacijos duomenų modelyje

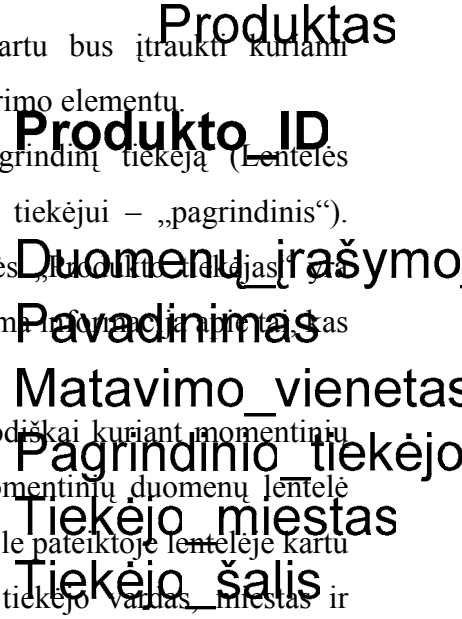


6 pav. Operacinių duomenų ryšio artefaktai duomenų saugyklos duomenų modelyje

Vienas iš ryšio artefakto apibrėžimų – ryšys egzistuoja tik vienu laiko momentu. Pvz., jei kažkas susituokė, o po to išsiskyrė, alimentų mokėjimas yra ryšio artefakto požymis. Tai reiškia, kad ryšys egzistavo tik tą akimirką, kai buvo sumokėti alimentai.

Duomenų saugykloje kuriant momentinius tarpinius duomenis kartu bus įtraukti kuriami duomenys, susiję su ryšiu. Tai vienas sudėtingiausių duomenų saugyklos kūrimo elementų.

5 paveiksle pateiktame pavyzdyje kiekvienas produktas turi pagrindinį tiekėją (Lentelės „Produkto tiekėjas“ atributo „Tiekėjo_privalumas“ reikšmė pagrindiniam tiekėjui – „pagrindinis“). Ryšio integralumas nusako, kad jei tiekėjas yra ištrinamas, tai iš lentelės „Produkto tiekėjas“ yra ištrinami tie įrašai, kurių pagrindinis tiekėjas buvo ištrintasis. Taip prarandama informacija apie tai, kas buvo pagrindinis tiekėjas.

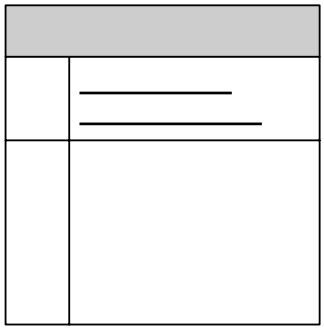


Tokią istorinę informaciją galima saugoti duomenų saugykloje, periodiškai kuriant momentinių duomenų lentelę ir į ją įtraukiant ryšio artefaktus (6 pav.). Produkto momentinių duomenų lentelė kuriama savaitės, mėnesio ar kito reikalingo laikotarpio pabaigoje. 6 paveiksle pateiktoje lentelėje kartu su produkto duomenimis, saugomi su produktu susiję ryšio artefaktai: tiekėjo vardas, miestas ir vietovė.

Aukščiau aptarti momentiniai duomenys turi vieną esminį trūkumą - jie yra nepilni. Jie parodo tik tokį ryšį, koks jis egzistuoja kažkuriuo konkrečiu laiko momentu. Pagrindiniai įvykiai gali įvykti taip, kad jie niekada nebus įtraukti į momentinius duomenis. Pvz., produkto duomenys yra nuskaitomi

kiekvienos savaitės pabaigoje. Per tą savaitę produktas turėjo tris pagrindinius tiekėjus, momentiniai duomenys šito fakto neatspindės.

Norint įtraukti pilną įrašą, siūloma naudoti istorinius įrašus (*angl. historical record*) (7 pav.).

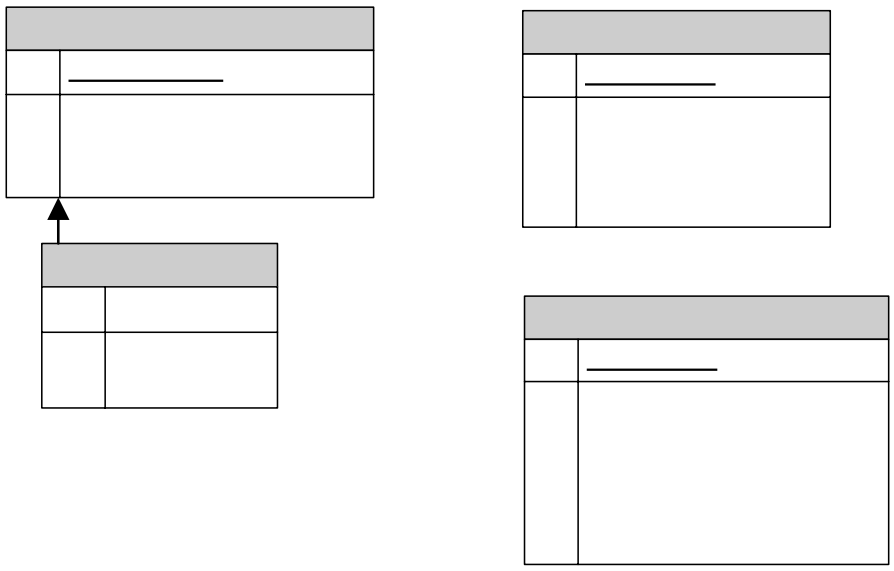


7 pav. Duomenų saugyklos duomenų modelis istoriniams duomenims

„Produkto istorijos“ lentelėje, visa informacija apie produktą yra įtraukiama, kai yra pristatomas produktas ir kartu su juo važtaraštis. Tai kita ryšio artefakto forma, kuria informacija gali būti įtraukta į duomenų saugyklą.

2.6.5. Skirtingų duomenų detalumo lygių pritaikymas duomenų saugykloje

Viena iš duomenų saugyklos ypatybių – skirtingas duomenų detalumo lygis. Kai kuriais atvejais, duomenų detalumo lygis organizacijos duomenis perkeliant į duomenų saugyklą nesikeičia. Kitais atvejais, duomenų detalumo lygis keičiasi, ir tai turi atspindėti duomenų saugyklos duomenų modelyje (8 pav.).

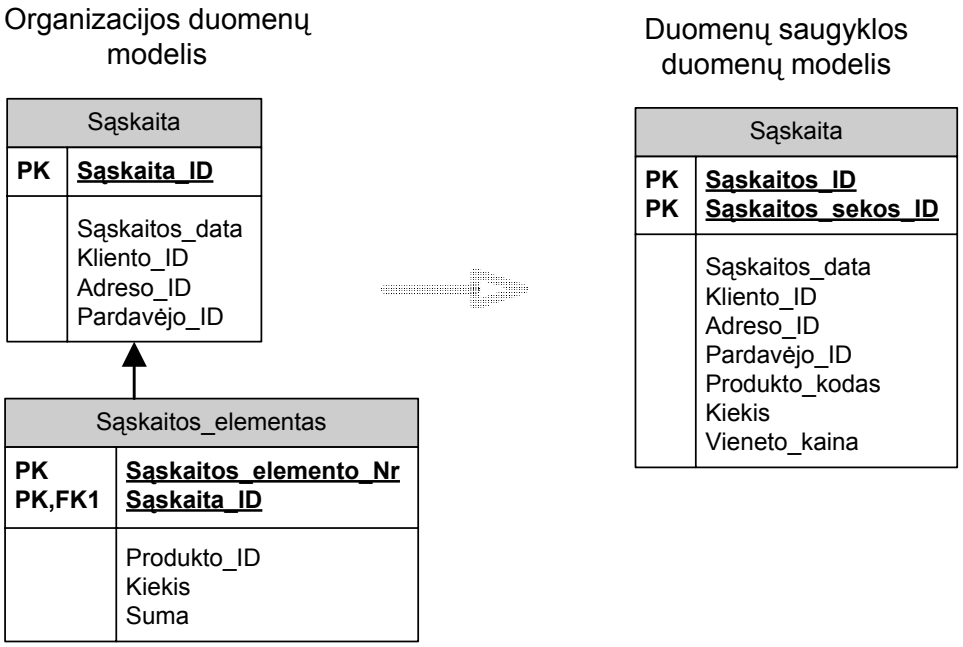


8 pav. Skirtingas duomenų detalumo lygis, perkeliant duomenis iš organizacijos duomenų modelio į duomenų saugyklą

8 paveiksle pateiktas organizacijos duomenų modelis, kuriame saugomi pardavimo duomenys, užfiksuojami kiekvienąkart, sudarant pardavimą. Atsižvelgiant į vartotojo reikalavimus, duomenų detalumas keičiasi duomenis perkeliant į duomenų saugyklą. Atliekami du sumavimai: mėnesinis pardavimų sumavimas ir pardavimų skirtingose vietovėse sumavimas.

2.6.6. Skirtingų lentelių duomenų suliejimas

Kita svarbi duomenų transformavimo operacija – susijusių lentelių suliejimas į vieną duomenų saugyklos lentelę (9 pav.).



9 pav. Susijusių organizacijos duomenų modelio lentelių sujungimas į viena duomenų saugyklos lentelę

Pavyzdyje pateiktos dvi normalizuotos organizacijos duomenų modelio lentelės: „Sąskaita“ ir „Sąskaitos_elementas“. Duomenų saugyklos modelyje šios dvi lentelės yra sujungiamos į vieną. Lentelių suliejimas gali pagreitinti užklausų vykdymą ir supaprastinti duomenų struktūrą.

Lentelių suliejimas yra naudingas, kai:

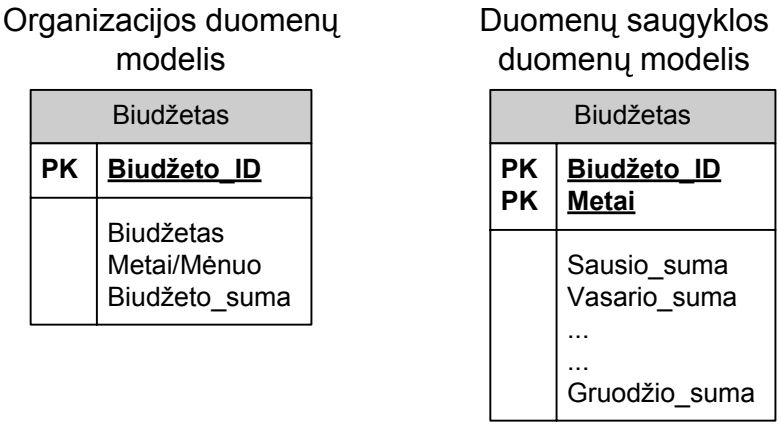
- lentelės dalinasi bendru raktu arba jo dalimi,
- skirtingų lentelių duomenys dažnai naudojami kartu,

Pateiktame pavyzdyje lentelės dalinasi bendru raktu „Sąskaita_ID“, todėl duomenų saugykloje jos yra suliejamos į vieną lentelę.

2.6.7. Duomenų masyvų sukūrimas

Kita duomenų transformavimo operacija – duomenų masyvų sukūrimas duomenų saugyklos modelyje. Duomenys organizacijos duomenų modelyje dažniausiai yra normalizuoti. Tai reiškia, kad

pasikartojančios grupės nėra parodomos duomenų modelyje. Tačiau duomenų saugykloje gali ir net turėtų būti patalpinamos pasikartojančios grupės, t.y. sukuriami duomenų masyvai. 10 paveiksle pateiktas duomenų masyvų kūrimo duomenų saugykloje pavyzdys.



10 pav. Duomenų masyvų sukūrimas duomenų saugyklos modelyje

Pateiktame pavyzdyje, organizacijos duomenų modelyje kiekvieno mėnesio biudžetui yra kuriamas naujas įrašas, t.y. vienų metų laikotarpiui yra sukuriama 12 įrašų. Perkeliant į duomenų saugyklą, duomenys yra perorganizuojami į duomenų masyvą taip, kad kiekvienas metų mėnuo yra to paties masyvo elementas, t.y. vieniems metams yra kuriamas vienas įrašas.

Yra keletas tokios duomenų struktūros privalumų. Vienas iš jų – nekuriant įrašo kiekvienam mėnesiui, o tik metams, yra sutaupoma daug vietos.

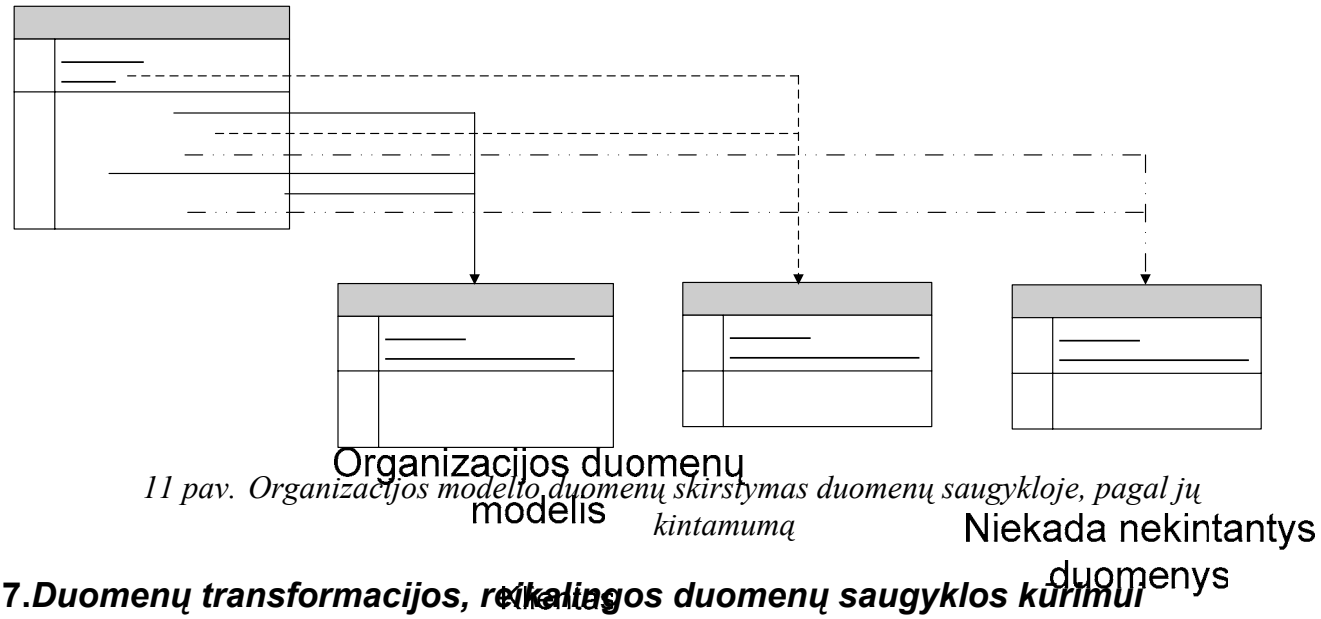
Kurti duomenų masyvus duomenų saugyklos modelyje naudinga, kai:

- iš anksto žinoma, kiek to paties tipo duomenų bus įrašyta,
- duomenų įrašymo kiekis yra mažas (kalbant apie fizinę vietą),
- duomenys dažniausiai naudojami kartu,

2.6.8. Duomenų atskyrimas pagal kintamumą

Paskutinė duomenų transformavimo operacija – suskirstyti duomenis duomenų saugykloje atsižvelgiant į duomenų kintamumą. Organizacijos duomenų modelyje nėra kreipiamas dėmesys į duomenų kintamumo lygį. Bet duomenų saugykla yra labai jautri duomenų kintamumui. Geriausias duomenų išdėstymas duomenų saugykloje yra toks, kai vienoje duomenų lentelėje esantys duomenys keičiasi retai, kitoje – dažnai. Šitokio duomenų skirstymo priežastis – jei vienos esybės duomenys yra saugomi vienoje lentelėje kartu su momentiniais duomenimis, tada bet kurio atributo reikšmės pasikeitimas reikalauja visos lentelės duomenų atnaujinimo. Dažnai besikeičiantys duomenys turi būti patalpinti atskiroje lentelėje, norint išvengti nesikeičiančių duomenų pastovaus kartojimo.

Pavyzdyje, (11 pav.) pateikta, kad organizacijos duomenų modelis kaupia duomenis apie klientus: grupė, kuriai priklauso klientas, kliento vardas, gimimo data ir t.t.. Duomenų saugykloje duomenys apie klientą suskirstomi į tris kategorijas: nekintantys, retai kintantys ir dažnai kintantys. Tai užtikrina, kad kiekvienąsyk atnaujinant duomenis duomenų saugykloje bus perrašyti tik reikiami duomenys: nekintanti informacija (pvz., gimimo_data) nebus perrašoma, o kintanti informacija (pvz., statusas) bus perrašoma.



11 pav. Organizacijos modelio duomenų skirstymas duomenų saugykloje, pagal jų kintamumą

2.7. *Duomenų transformacijos, reikalingos duomenų saugyklos kūrimui*

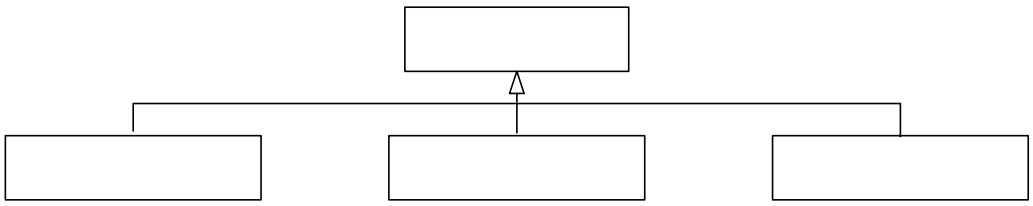
Duomenų saugyklos pagrindinis tikslas yra surinkti į vieną vietą visus vartotojui reikalingus duomenis.

Todėl ne visos išanalizuotos transformavimo operacijos yra būtinos. Kai kurias transformavimo operacijas galima perkelti į analizės (OLAP) lygmenį, kai kurių automatizuoti gali būti neįmanoma.

Pavyzdžiui, operacinių duomenų pašalinimas yra loginis operacinių duomenų atmetimas, kurį atlieka vartotojas, apibrėždamas duomenų saugyklos reikalavimus.

Skirtingų detalumo lygių duomenų įvedimą, masyvų sukūrimą ir pan. tikslinga perkelti į duomenų analizės lygmenį ir realizuoti OLAP operacijomis.

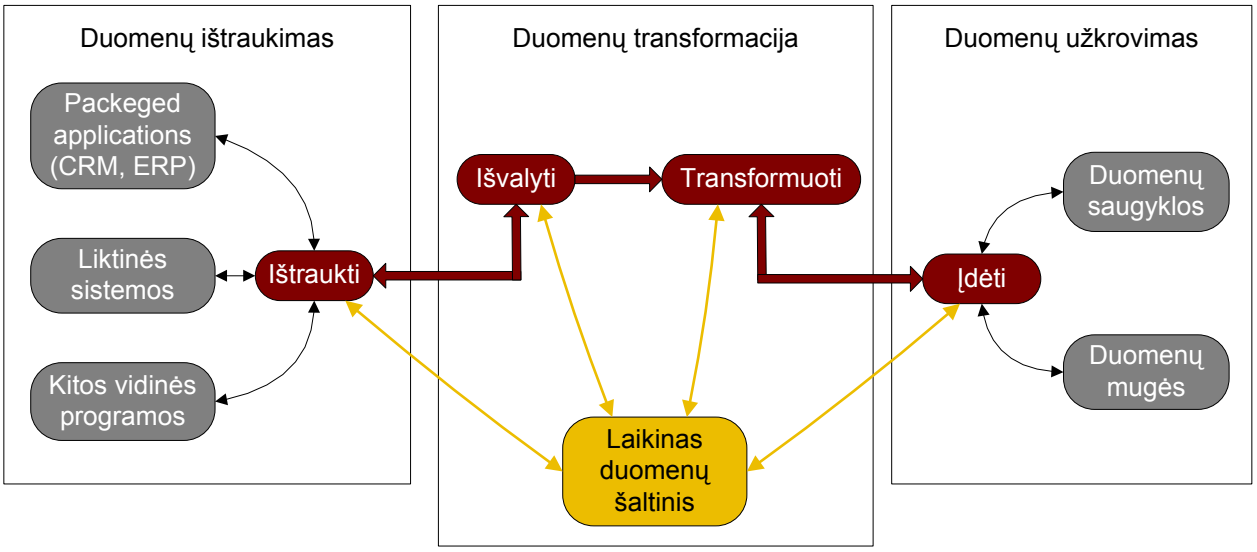
Svarbiausios duomenų saugyklos kūrimo transformacijos – tai reliacinių struktūrų pertvarkymas į saugyklos dimensijų ir faktų struktūras bei laiko dimensijų sukūrimas (jei naudojama laiko dimensija, bet taip būna labai dažnai) (12 pav.). Norint realizuoti šias transformacijas, reikia panaudoti kai kurias duomenų transformavimo operacijas.



12 pav. Duomenų saugyklos kūrimui reikalingos transformacijos

2.8.*ETL procesų ir programinių priemonių analizė*

ETL (ištrauk/ transformuok/ įdėk) – tai procesai leidžiantys ištraukti duomenis iš skirtingų šaltinių, juos išvalyti bei transformuoti ir užkrauti į kitas duomenų bazes: duomenų saugyklas ar duomenų muges (angl. data marts), duomenų analizei atlikti arba kitas operacines sistemas, biznio procesams palaikyti. Principinė ETL procesų vykdymo schema pateikta 13 paveiksle.



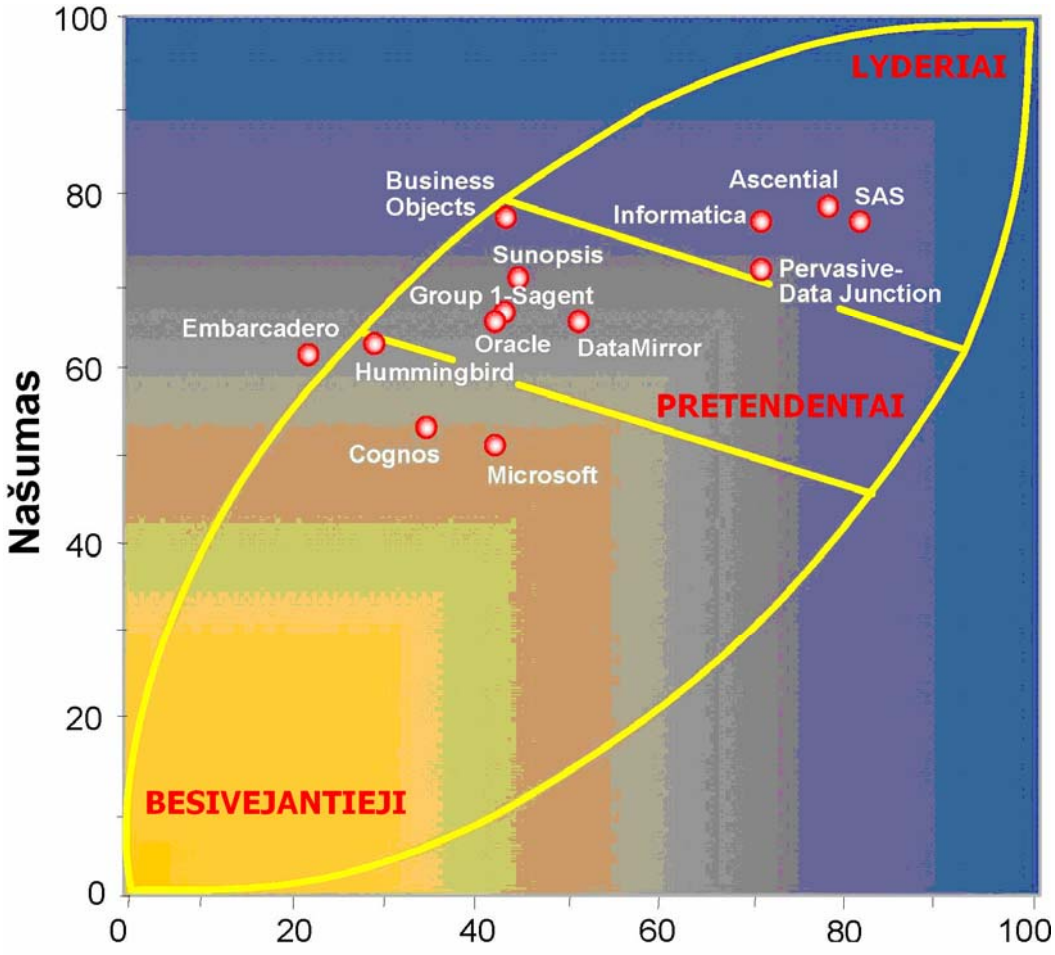
13 pav. ETL procesų vykdymo schema

Programinės priemonės realizuojančios ETL procesus vadinamos ETL priemonėmis.

Didžiosios duomenų bazių valdymo sistemos (DBVS) turi integruotas ETL priemones: MS SQL Server – DTS (*angl. Data Transformation Services*), Oracle – Warehouse Builder. Yra ir atskirai platinamų ETL priemonių: SAS Data Warehouse (SAS Institute), Power Center (Informatica), Data Integrator (Pervasive), Data Integrator (Business Objects) ir kiti.

2.8.1. ETL programinių priemonių palyginimas

ETL programinių priemonių įvairovė labai didelė, tad atsirinkti, kuri priemonė tinkamiausia vartotojui labai sunku. 14 paveiksle pateikta ETL priemonių lyginamoji schema, atlikta METAspectrum 2004 m. balandžio 7 dieną.



14 pav. ETL priemonių palyginimas

Vykdant šią ETL priemonių lyginamąją analizę buvo atsižvelgta į:

- **platformų palaikymą**: nustatant šį kriterijų remtasi duomenų šaltinių, duomenų saugyklų ir darbinės aplinkos skaičiaus palaikymu.
- **transformaciją**: šį kriterijų lėmė realizuota aibė siūlomų duomenų transformacijų ir lanksti vystymo ir integravimo savitos biznio logikos metodika.
- **duomenų valdymo įrankius**: šį kriterijų lėmė patogus įrankis, skirtas transformacijų, periodinių darbų valdymui
- **našumo charakteristiką**: nustatant ši kriterijų remtasi kuo greitesniu duomenų apdorojimu ir jo lygiagretinimo savybe.

ETL priemonių rinkos lyderiu kol kas laikoma Ascential firma. Deja, jinai informaciją apie savo produktus laiko didelėje paslaptyje. Todėl konkretesnės informacijos apie jos kuriamą ETL priemonę „DataStage“ nepavyko rasti. Pagal kitų firmų: SAS, Informatica, Microsoft, Oracle ir kt., oficialiose internetinėse svetainėse pateiktą informaciją sudaryta palyginamąją lentelę (1 lentelė), kurioje pateikti duomenys apie duomenų transformavimo operacijų realizaciją įvairių firmų ETL priemonėse.

1 lentelė

Duomenų transformavimo operacijų realizacija ETL priemonėse

ETL priemonė	**Organizacija**	**Duomenų transformavimo operacijos**	**Transformacijų naudojimui skirta vartotojo sąsaja (*angl. GUI*)**
DataStage	Ascential	Nepateikta informacijos	Nepateikta informacijos
	SAS	Integruota apie 300 standartinių transformavimo operacijų. Vartotojas gali sukurti savo operacijas.	Turi
Power Center	Informatica	Integruota daug transformavimo operacijų	Panaudojimas „drag and drop“ principu
Data Integrator	Pervasive – Data Junction	Integruota daug transformavimo.	Turi
Data Integrator	Business Objects	Integruota daug ir gana sudėtingų transformavimo operacijų.	Panaudojimui užtenka vieno vartotojo „žingsnio“.
Real – time Data Integrator	Sunopsis	Integruotos visos transformavimo operacijos: lentelių sujungimas, skaičiuojami laukai ir t.t.	Turi
Warehouse Builder	Oracle	Neturi integruotų transformavimo operacijų. Transformavimo operacijos rašomos PL/SQL kalba.	Neturi
DTS (*angl. Data Transformation Services*)	Microsoft	Neturi integruotų transformavimo operacijų. Transformavimo operacijos rašomos SQL ar VB kalbomis.	Neturi

Kaip matyti iš pateikto ETL priemonių lyginimo (14 pav.), ETL priemonės, kuriamos atskirai nuo duomenų bazių valdymo sistemų yra lyderės. Tokių kompanijų, kaip Oracle, Microsoft kartu su duomenų bazių valdymo sistema siūlomos ETL priemonės yra priskiriamos tik prie besivejančių ar pretendentų.

Pagal 1 lentelėje pateiktą informaciją irgi galime daryti panašią išvadą. Visos lyderiaujančios ETL priemonės (pagal 14 pav.) turi savyje integruotas duomenų transformavimo operacijas, kurių panaudojimui daugelyje sukurta vartotojo sąsaja. Tuo tarpu Microsoft ir Oracle siūlomose ETL priemonėse, duomenų transformacijos yra rašomos paties vartotojo, panaudojant SQL, PL/SQL programavimo kalbas, kas žymiai sulėtina duomenų saugyklos kūrimo procesą.

2.9. *ETL priemonės duomenų saugyklos kūrimo proceso tobulinimui pasirinkimas*

ETL priemonių analizė atskleidė, kad patogiausios vartotojui yra atskirai nuo DBVS kuriamos ETL priemonės: Ascential, Informatica ar kitų firmų. Tuo tarpu, Oracle, Microsoft firmų platinama ETL priemonė negali pasigirti nei integruotomis transformavimo operacijomis, nei jų panaudojimui skirta vartotojo sąsaja. Norint pagreitinti ir pagerinti duomenų saugyklos kūrimą plačiai naudojamose DBVS, darbui pasirinktas Microsoft firmos MS SQL DTS paketas.

2.9.1. Microsotf SQL Server DTS

Microsoft SQL duomenų bazių valdymo sistemoje duomenų transformacijoms, importavimui ir eksportavimui naudojami DTS (*angl. data transformation services*) (duomenų transformavimo servisai).

DTS pateikia priemonių rinkinį, kuris leidžia išgauti, transformuoti ir susieti duomenis esančiuose skirtinguose šaltiniuose, prisijungimą prie kurių palaiko DTS prisijungimų posistemė.

DTS paketas tai prisijungimų prie duomenų bazės, DTS užduočių (*angl. tasks*) ir DTS transformacijų rinkinys. Taip pat jame nurodoma DTS darbų seka (*angl. workflow*). DTS paketas dažniausiai saugomas MSSQL duomenų bazės metaduomenų saugykloje (*angl. repository*).

DTS užduotis (*angl. task*) yra diskreti funkcionalumo aibė, vykdoma kaip vienas žingsnis DTS pakete. Kiekviena užduotis apibrėžia bendro duomenų perkėlimo darbo sudėtinę dalį, kurios metu perkeliama ar transformuojama kokia nors duomenų dalis arba duomenų bazėje atliekamas koks nors darbas. Duomenų perkėlimo į duomenų saugyklą metu dažniausiai naudojami šie DTS užduočių tipai :

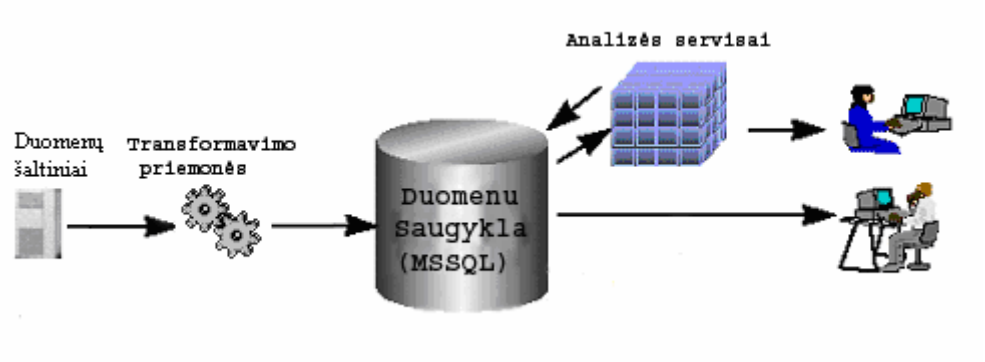
duomenų importas ir eksportas – DTS gali importuoti/eksportuoti duomenis esančius bet kokiuose OLE DB prieiga pasiekiamuose šaltiniuose, nutolusiuose ir lokaliuose SQL serveriuose bei tekstiniuose failuose.

duomenų transformavimas – leidžia paimti bet kokiame duomenų rinkinyje esančius duomenis, arba kombinuoti keliuose rinkiniuose esančius duomenis SQL užklausų pagalba. Duomenys gali būti padedami į kita duomenų rinkinį. Tarp rinkinių galima sudaryti ryšius. Perkeliamų duomenų įrašus galima papildomai transformuoti panaudojant įvairias duomenų konvertavimo bei agregavimo funkcijas.

duomenų bazės objektų kopijavimas – Leidžia perkelti iš vienos duomenų bazės į kitą tokius duomenų bazėse saugomus objektus kaip procedūros, vaizdai (*angl. views*) ir pan.

2.9.2. MS SQL duomenų analizės architektūra

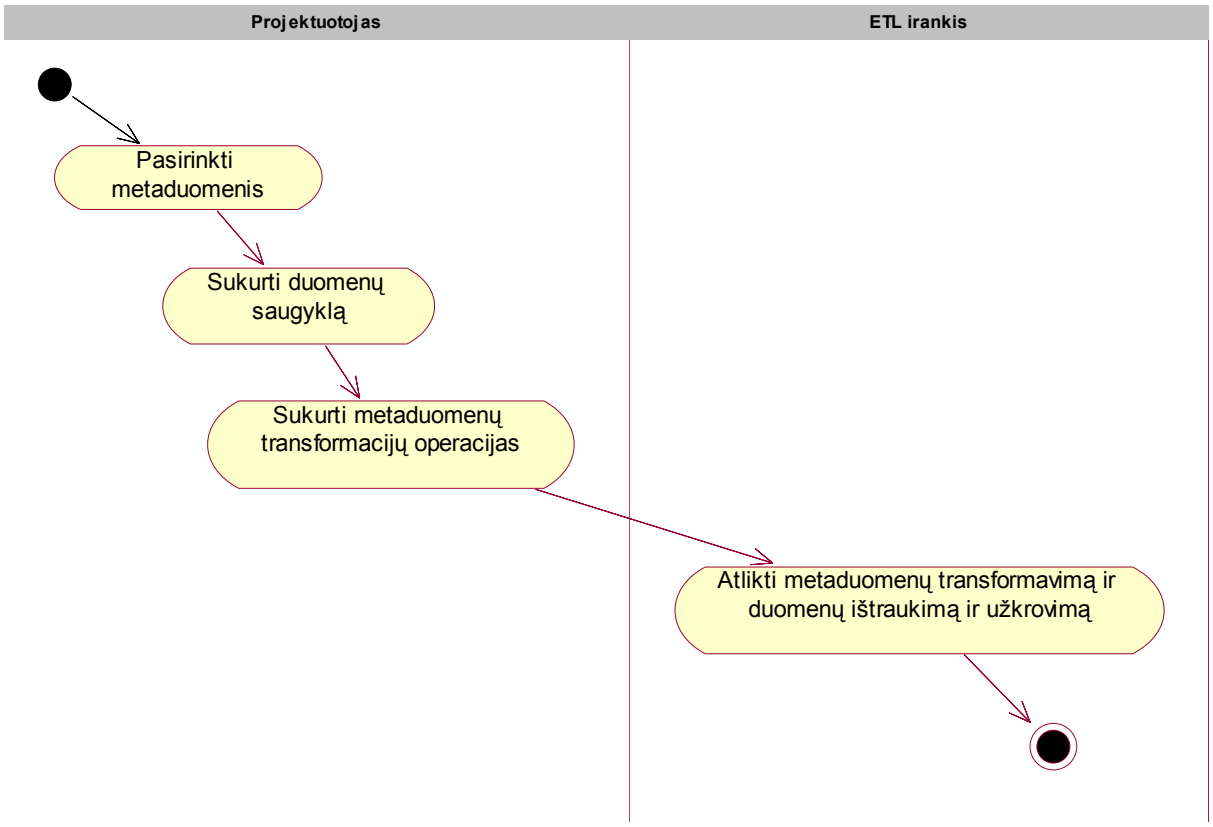
15 paveiksle pateikta bendra duomenų analizės sistemos priemonių architektūra, naudojant MS SQL priemones.



15 pav. Duomenų analizės sistemos architektūra naudojant MS SQL priemones

2.10. *Siekiamos ETL priemonės apibrėžimas*

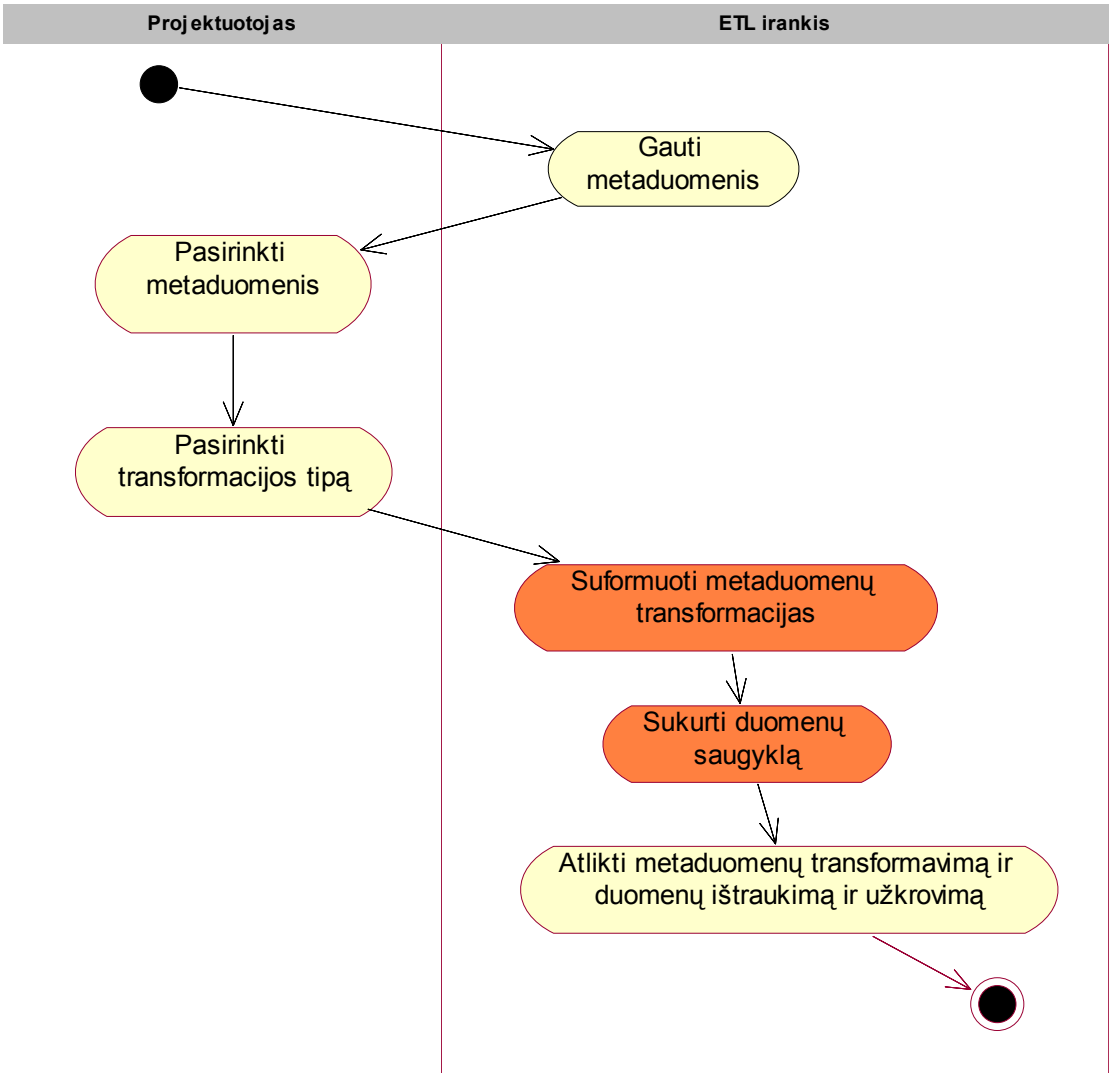
Šiuo metu kuriant duomenų saugyklą MS SQL priemonėmis projektuotojui viską reikia rankomis: iš reliacinės duomenų bazės atsirinkti reikiamus metaduomenis, sukurti metaduomenų transformavimo operacijas (16 pav.).



16 pav. Esamas saugyklos kūrimo procesas

Duomenų saugyklos projektuotojas metaduomenis turi atsirinkti žiūrėdamas į duomenų bazę. SQL kalba surašomos metaduomenų transformacijų procedūros. . Sistema įvykdo parašytas SQL procedūras: sukuria duomenų saugyklą, transformuoja metaduomenis ir perkelia duomenis.

Siekiama sukurti tokią MS SQL pritaikytą ETL priemonę, kuri pagreitintų duomenų transformacijų kūrimą (17 pav.).



17 pav. Siekiamos sistemos veiklos diagrama

Duomenų saugyklos projektuotojas nurodo prisijungimo prie šaltinio duomenų bazės duomenis, sistema ištraukia ir pateikia pagrindinius duomenų bazės metaduomenis. Pasirinkus reikiamus metaduomenis, sistema sukuria nurodytą transformaciją: dimensijos arba fakto (17 pav.). Remiantis MS SQL DTS vykdomas metaduomenų transformavimas ir duomenų perkėlimas iš reliacinės duomenų bazės į kuriamą duomenų saugyklą.

2.11. *Darbo tikslas ir siekiami kokybiniai kriterijai*

Atlikus duomenų saugyklos proceso, duomenų transformacijų tipų ir ETL priemonių rinkos analizę buvo suformuotas darbo tikslas.

Darbo tikslas: Sukurti ETL priemonę, pritaikytą automatizuotam duomenų saugyklos kūrimui MS SQL priemonėmis.

Norint pasiekti išsikeltą darbo tikslą, reikia įgyvendinti tokius **uždavinius**:

- sukurti duomenų saugyklą, esamomis MS SQL priemonėmis,
- sukurti duomenų transformavimo operacijas,
- parengti eksperimentinius duomenų bazės duomenis,
- perkelti duomenis iš duomenų bazės į duomenų saugyklą,
- sukurti kubą ir išbandyti analizės galimybes,
- išanalizuoti sukurtas duomenų transformavimo operacijas, jas apibendrinti,
- parengti duomenų transformavimo operacijų šablonus, kuriuos būtų galima pritaikyti kuriant ETL priemonę.
- suprojektuoti ir realizuoti ETL priemonę,
- išbandyti ETL priemonę įvairioms reliacinėms duomenų bazėms, saugomoms MS SQL Server, transformuoti į duomenų saugyklas.

Siekiamos ETL priemonės privalumai pateikti 2 lentelėje.

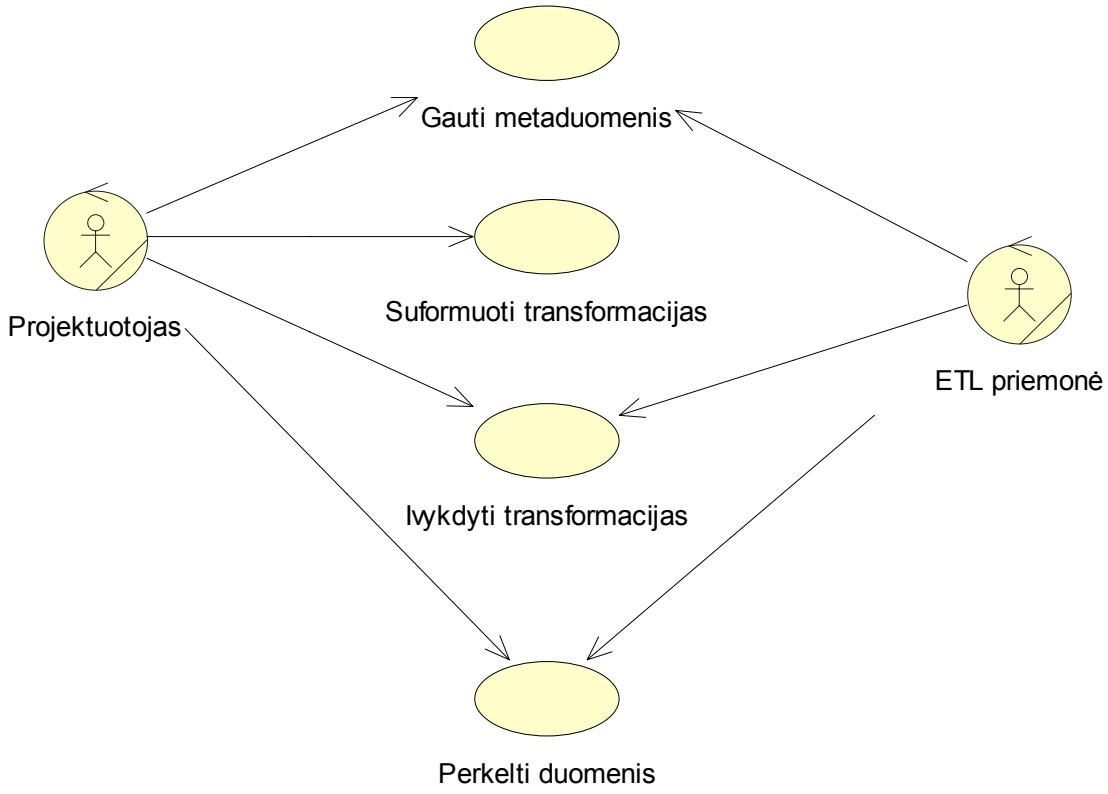
2 lentelė

Siekiami sistemos privalumai

Siekiamos ETL sistemos privalumai	**Esama padėtis**
Duomenų transformacijų operacijos bus realizuotos kuriamoje ETL priemonėje ir kuriant duomenų saugyklą reiks pasirinkti norimą transformaciją.	Kiekvienąkart kuriant saugyklą, duomenų transformacijų operacijas reikia aprašyti programiniu kodu.
Bus sukurta grafinė vartotojo sąsaja (*angl. GUI*) transformacijų panaudojimui.	Grafinės sąsajos ETL priemonė neturi.
Pagreitės duomenų saugyklos kūrimo procesas.	Duomenų transformacijų rašymas užima 80% laiko, skirto duomenų saugyklos kūrimui.

2.12. *Siekiamos ETL priemonės funkcijos*

Siekiamos ETL priemonės funkcijos pateiktos 18 paveiksle.



18 pav. Siekiamos ETL sistemos funkcijos

Prieš pradedant kurti duomenų saugyklą, turite išanalizuoti duomenų šaltinio (reliacinės duomenų bazės saugomos MS SQL Server) metaduomenis. Norint gauti duomenų šaltinio metaduomenis, duomenų saugyklos projektuotojas turi nurodyti šaltinio informaciją: MS SQL serverio vardą, duomenų bazės pavadinimą. Nurodžius visą reikiamą informaciją, sistema ištraukia metaduomenis reikalingus duomenų saugyklos kūrimui.

Sistemos projektuotojas pasinaudodamas vartotojo sąsaja išsirenka reikalingą transformaciją ir metaduomenis ir sukuria transformaciją: dimensijos arba fakto.

Sukūręs visas dimensijas ir faktą, duomenų saugyklos projektuotojas gali įvykdyti transformacijas: sukurti MS SQL Server duomenų saugyklą ir dimensijas bei faktą.

Sukūręs duomenų saugyklą, duomenų saugyklos projektuotojas iš karto arba vėliau gali perkelti duomenis iš duomenų šaltinio į duomenų saugyklą.

2.13. *Reikalavimai duomenų saugyklos šaltinių duomenų bazėms*

Paprastai organizacijos duomenys yra kaupiami įvairiuose formatuose: operacinių duomenų transakcijų apdorojimo sistemose OLTP (*angl. On-line Transaction Processing*), liktinėse sistemose, tekstiniuose failuose, MS Excel failuose ir t.t.

Darbe analizuojami duomenys saugomi OLPT sistemose, konkrečiai MS SQL Server duomenų bazėje. Visi reikalingi duomenys turi būti saugomi vienoje (ne keliose) MS SQL normalizuotoje duomenų bazėje. Normalizuota duomenų bazė turi tenkinti tokius reikalavimus:

- Atributų reikšmės yra atominės, t.y. neskaidomos į dalis tam, kad susieti jas su kitos ar tos pačios lentelės duomenimis.
- Nėra atributo, kuris priklausytų nuo pirminio rakto dalies.
- Neegzistuoja nepirminis lentelės atributas, kuris yra tranzityviojoje funkcinėje priklausomybėje nuo lentelės rakto.

Dalykinė duomenų sritis neturės įtakos duomenų saugyklos kūrimui. Lentelės duomenų bazėje privalo turėti ryšius. Iš lentelių, neturinčių tarpusavio ryšio nebus įmanoma ištraukti duomenų.

2.14. *ETL priemonės kūrimo proceso rizikos faktorių analizė*

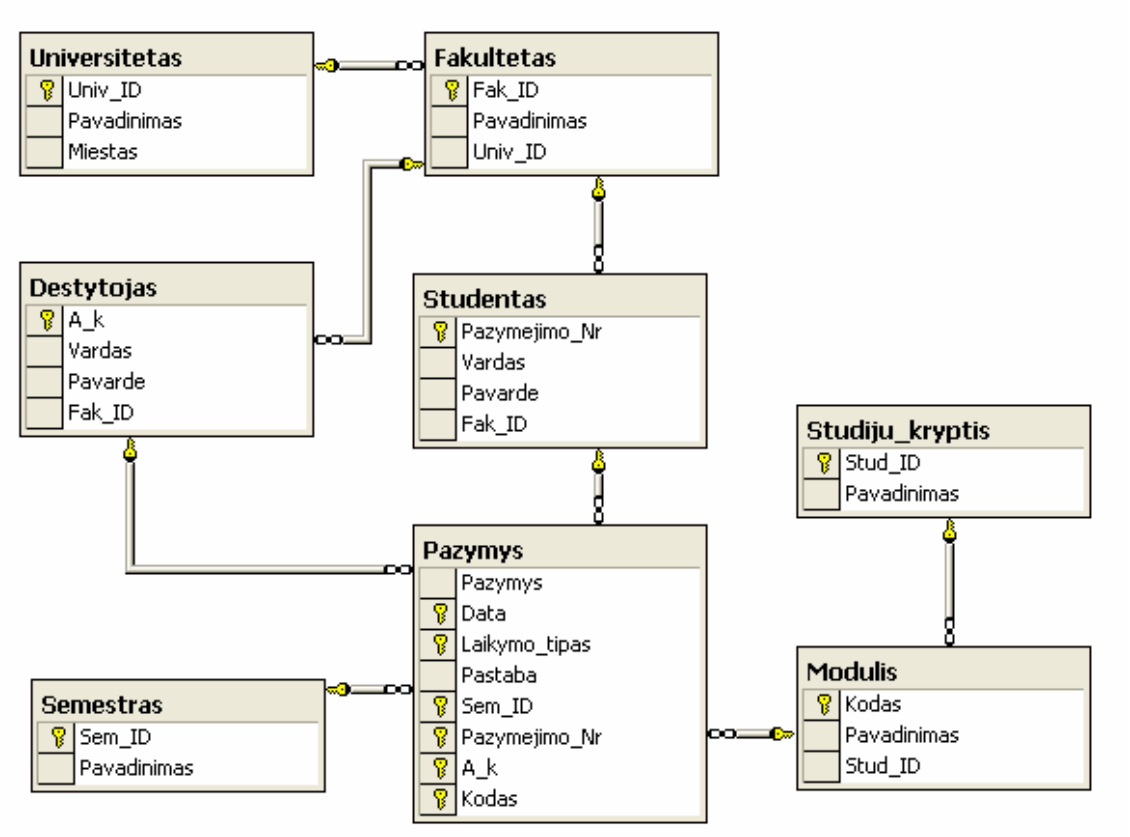
Siūlomas duomenų transformacijų kūrimo sprendimas yra naujas, todėl rasti panašių pavyzdžių ar literatūros nepavyko. Kuriant naują sistemą nėra galimybės remtis naudotojų pastabomis dėl sistemos veikimo, todėl visada išlieka didelė rizika, sukurti probleminę sritį ne visai tenkinantį produktą.

2.15. *Analizės išvados*

1. Sprendžiant duomenų saugyklos kūrimo proceso sutrumpinimo problemą, buvo išanalizuoti saugyklų kūrimo procesas ir egzistuojantys kūrimo įrankiai.
2. Saugyklų kūrime galima išskirti dvi kraštutines tendencijas: kurti saugyklą, apimančią visus įmonės procesus ir potencialiai galinčią patenkinti visus analizės poreikius; arba naudoti standartinius analizės paketus, turinčius fiksuotų analizės priemonių rinkinius
3. Šiame darbe siekiama sukurti galimybes pagreitinti saugyklos kūrimo procesą universaliose DBVS.
4. ETL priemonių analizė atskleidė, kad patogiausios vartotojui yra tos ETL priemonės, kurios yra kuriamos atskirai nuo duomenų bazių valdymo sistemų (SAS, Informatica ir kt, firmų). Tuo tarpu, Oracle, Microsoft firmų platinama ETL priemonė negali pasigirti nei integruotomis transformacijomis, nei jų panaudojimui skirta vartotojo sąsaja.
5. Išnagrinėti duomenų saugyklos kūrimo proceso etapai: duomenų šaltinių identifikavimas, analizės poreikių nustatymas, saugyklos schemos ir duomenų transformacijų projektavimas, duomenų transformavimas ir perkėlimas. Padaryta išvada, kad didžiausias galimybes sutrumpinti saugyklos kūrimo laiką turi schemos ir yra duomenų transformacijų projektavimas.
6. Atlikta detali duomenų transformacijų tipų analizė, atrinktos svarbiausios transformacijos, kurias tikslinga automatizuoti.
7. Suformuluotas tyrimo uždavinys: sukurti dažniausiai naudojamų transformacijų šablonus ir jų panaudojimui skirtą grafinę vartotojo sąsają, kurie padėtų žymiai greičiau sukurti duomenų saugyklas. Transformacijos bus kuriamos MS SQL DBVS, kadangi kiekviena DBVS turi savo specifines kūrimo priemones ir SQL dialektą.

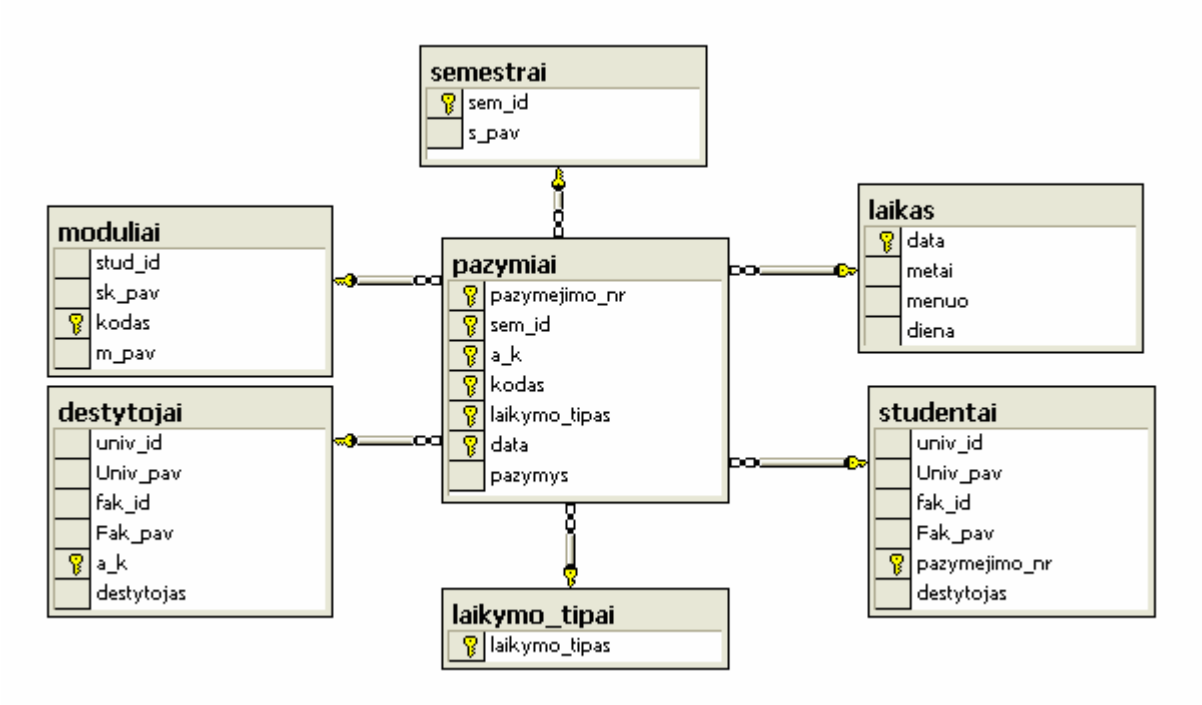
3. Duomenų saugyklos kūrimas, panaudojant egzistuojančias MS SQL priemones

Duomenų saugyklos kūrimui pasirinkta reliacinė duomenų bazė, kurioje kaupiami duomenys apie visų Lietuvos universitetų dėstytojus, studentus, dėstomus dalykus bei studentų įvertinimus (19 pav.). Duomenų bazė saugoma MS SQL Server.



19 pav. Reliacinė studentų duomenų bazė

Kuriama duomenų saugykla, skirta studentų pažangumo analizei. Duomenų saugyklos schema pasirinkta žvaigždės schema (20 pav.). Duomenų saugyklos duomenų bazė irgi saugoma MS SQL Server.



20 pav. Studentų pažangumo analizės „žvaigždės“ schemos modelis

Norėdami sukurti duomenų saugyklą, turime atlikti tokias duomenų transformacijas:

- Iš lentelės „Pažymys“išskirti laiko ir laikymo tipo dimensijas.
- Iš lentelės „Pažymys“ išmesti nereikalingus atributą – „Pastaba”.
- Lenteles „Universitetas“, „Fakultetas“, „Dėstytojas“ sujungti į Dėstytojo dimensiją.
- Lenteles „Universitetas“, „Fakultetas“, „Studentas“ sujungti į Studento dimensiją.
- Lenteles „Studijų kryptis“, „Modulis“ sujungti į Modulio dimensiją.
- Sukurti faktą.

3.1.*Duomenų transformacijų realizacija*

3.1.1. Laiko dimensijos sukūrimas

```
use Pazymiai_WH
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[laikas]') and OBJECTPROPERTY(id, N'IsUserTable')
= 1)
        drop table.dbo.laikas

use Pazymiai
select
        distinct data,                  -- data
                year(data) metai,  -- metai
                month(data) menuo,      -- menuo
```

```
        day(data) diena          -- diena
into Pazymiai_WH.dbo.laikas
from pazymys (nolock)

use Pazymiai_WH
alter table dbo.laikas
    alter column data datetime not null
GO
ALTER TABLE dbo.laikas ADD CONSTRAINT
    PK_laikas PRIMARY KEY CLUSTERED
    (data) ON [PRIMARY]
```

3.1.2. Laikymo tipo dimensijos sukūrimas

```
use Pazymiai_WH
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[laikymo_tipai]') and OBJECTPROPERTY(id, N'IsUserTable') = 1)
    drop table.dbo.laikymo_tipai

use Pazymiai
select
    distinct laikymo_tipas                  -- semestras
into Pazymiai_WH.dbo.laikymo_tipai
from pazymys

use Pazymiai_WH
ALTER TABLE dbo.laikymo_tipai ADD CONSTRAINT
    PK_laikymo_tipai PRIMARY KEY CLUSTERED
    (laikymo_tipas) ON [PRIMARY]
```

3.1.3. Dėstytojo dimensijos sukūrimas

```
use Pazymiai_WH
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[Destytojai]') and OBJECTPROPERTY(id, N'IsUserTable') = 1)
    drop table.dbo.destytojai

use Pazymiai
select
    u.univ_id,  u.pavadinimas Univ_pav,          -- universitetas
    f.fak_id, f.pavadinimas Fak_pav,             -- fakultetas
    d.a_k, d.vardas + ' ' + d.pavarde destytojas -- destytojas
```

```
into Pazymiai_WH.dbo.destytojai
from destytojas d
left join fakultetas f
        on f.fak_id = d.fak_id
left join universitetas u
        on u.Univ_id = f.Univ_id

use Pazymiai_WH
ALTER TABLE dbo.destytojai ADD CONSTRAINT
        PK_destytojai PRIMARY KEY CLUSTERED
        (a_k) ON [PRIMARY]
```

3.1.4. Studento dimensijos sukūrimas

```
use Pazymiai_WH
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[Studentai]') and OBJECTPROPERTY(id, N'IsUserTable') = 1)
        drop table.dbo.studentai

use Pazymiai
select
        u.univ_id,  u.pavadinimas Univ_pav,                    -- universitetas
        f.fak_id, f.pavadinimas Fak_pav,              -- fakultetas
        s.pazymejimo_nr, s.vardas + ' ' + s.pavarde destytojas -- destytojas
into Pazymiai_WH.dbo.studentai
from studentas s
left join fakultetas f
        on f.fak_id = s.fak_id
left join universitetas u
        on u.Univ_id = f.Univ_id

use Pazymiai_WH
ALTER TABLE dbo.studentai ADD CONSTRAINT
        PK_studentai PRIMARY KEY CLUSTERED
        (pazymejimo_nr) ON [PRIMARY]
```

3.1.5. Modulio dimensijos sukūrimas

```
use Pazymiai_WH
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[moduliai]') and OBJECTPROPERTY(id, N'IsUserTable') = 1)
        drop table.dbo.moduliai
```

```
use Pazymiai
select
        sk.stud_id, sk.pavadinimas sk_pav,          -- studiju kryptis
        m.kodas, m.pavadinimas m_pav                -- modulis
into Pazymiai_WH.dbo.moduliai
from modulis m
left join studiju_kryptis sk
        on m.stud_id = sk.stud_id

use Pazymiai_WH
ALTER TABLE dbo.moduliai ADD CONSTRAINT
        PK_moduliai PRIMARY KEY CLUSTERED
        (kodas) ON [PRIMARY]
```

3.1.6. Fakto sukūrimas

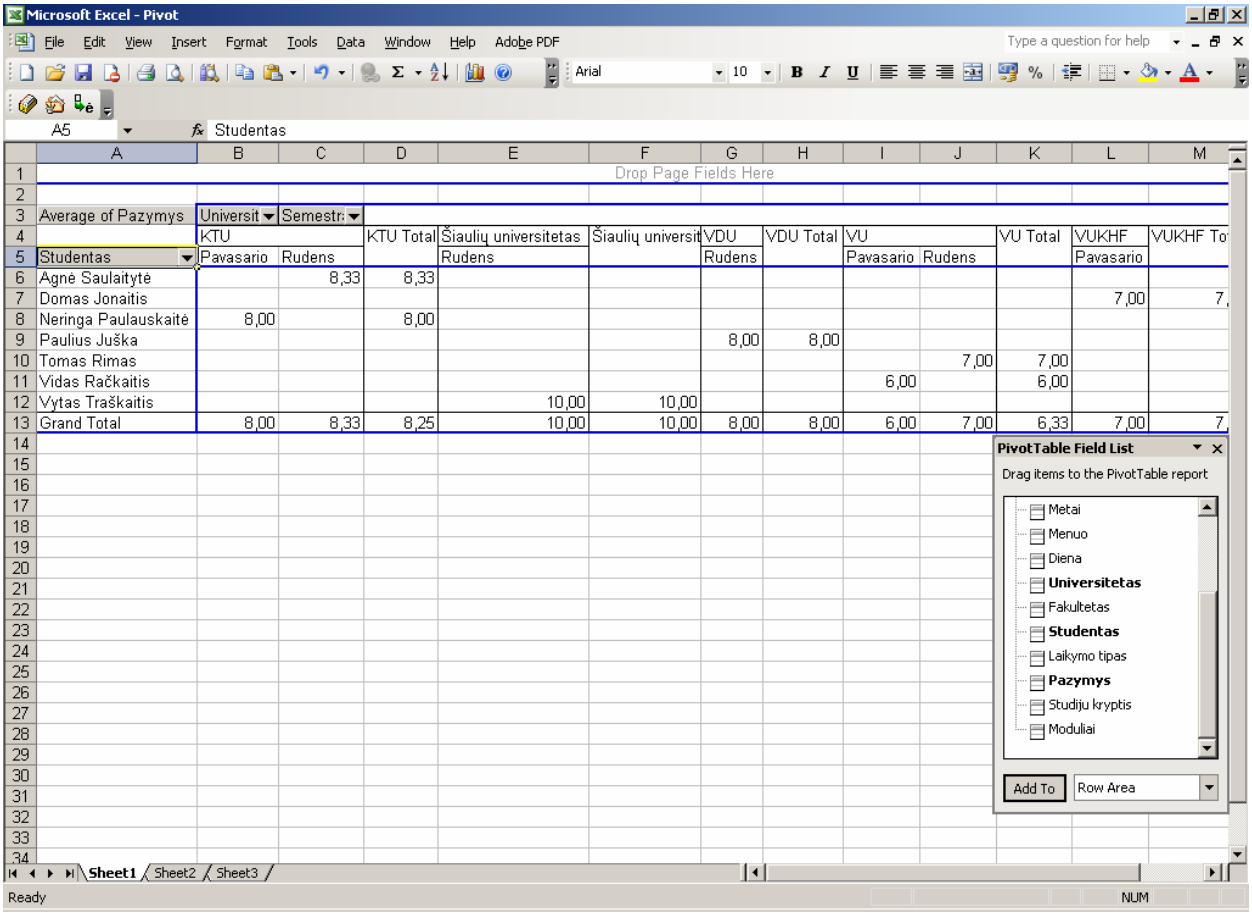
```
use Pazymiai_WH
if exists (select * from dbo.sysobjects where id = object_id(N'[dbo].[pazymiai]') and OBJECTPROPERTY(id, N'IsUserTable') = 1)
        drop table.dbo.pazymiai

use Pazymiai
select
        pazymejimo_nr,          -- studentas
        sem_id,                 -- semestras
        a_k,                    -- destytojas
        kodas,                  -- modulis
        laikymo_tipas,          -- laikymo tipas
        data,                   -- laikas
        pazymys         -- faktiniai duomenys
into pazymiai_WH.dbo.pazymiai
from pazymys

use Pazymiai_WH
alter table dbo.Pazymiai
        alter column pazymejimo_nr int not null
alter table dbo.Pazymiai
        alter column sem_id int not null
alter table dbo.Pazymiai
```

```
        alter column a_k int not null
alter table dbo.Pazymiai
        alter column kodas varchar(8) not null
alter table dbo.Pazymiai
        alter column laikymo_tipas varchar(50) not null
alter table dbo.Pazymiai
        alter column data datetime not null

GO
ALTER TABLE dbo.Pazymiai ADD CONSTRAINT
        PK_Pazymiai PRIMARY KEY CLUSTERED
        (pazymejimo_nr,
        sem_id,
        a_k,
        kodas,
        laikymo_tipas,
        data) ON [PRIMARY]

GO
ALTER TABLE dbo.pazymiai ADD CONSTRAINT
        FK_pazymiai_destytojai FOREIGN KEY
        (a_k) REFERENCES dbo.destytojai
        (a_k)

ALTER TABLE dbo.pazymiai ADD CONSTRAINT
        FK_pazymiai_studentai FOREIGN KEY
        (pazymejimo_nr) REFERENCES dbo.studentai
        (pazymejimo_nr)

ALTER TABLE dbo.pazymiai ADD CONSTRAINT
        FK_pazymiai_moduliai FOREIGN KEY
        (kodas) REFERENCES dbo.moduliai
        (kodas)

ALTER TABLE dbo.pazymiai ADD CONSTRAINT
        FK_pazymiai_laikymo_tipas FOREIGN KEY
        (laikymo_tipas) REFERENCES dbo.laikymo_tipai
        (laikymo_tipas)
```

```
ALTER TABLE dbo.pazymiai ADD CONSTRAINT
    FK_pazymiai_laikas FOREIGN KEY
    (data) REFERENCES dbo.laikas
    (data)

ALTER TABLE dbo.pazymiai ADD CONSTRAINT
    FK_pazymiai_semestrai FOREIGN KEY
    (sem_id) REFERENCES dbo.semestrai
    (sem_id)
```

3.2. *Duomenų analizė, naudojant sukurtą saugyklą*

Sukurtos saugyklos duomenų peržiūrai sukurta ataskaita, panaudojant MS Excel Pivot Table (21 pav.).



21 pav. Studentų pažangumo analizės ataskaita

3.3. *Duomenų saugyklos kūrimo esamomis MS SQL priemonėmis apibendrinimas*

Analizės dalyje buvo išskirti trys pagrindiniai duomenų transformacijos tipai:

- Dimensijos transformacija,

- Fakto transformacija,
- Laiko dimensijos transformacija.

Eksperimentinis duomenų saugyklos kūrimas patikslino analizės rezultatus.

Dimensijos transformacija – tai visos arba dalies reliacinės duomenų lentelės perkėlimas arba kelių reliacinių duomenų lentelių atributų apjungimas į vieną lentelę, dviejų atributų apjungimas.

Fakto transformacija – tai reikalingų duomenų ištraukimas iš vienos arba kelių reliacinių lentelių ir ryšių su dimensijų lentelėmis sukūrimas.

Laiko dimensijos transformacija – reliacinio laiko elemento skaidymas į laiko dimensiją. Laiko dimensija gali būti įvairių tipų:

- Metai – mėnuo – diena,
- Metai – pusmetis – mėnuo – diena.

Priklausomai nuo srities, kuriai yra kuriama duomenų saugyklą, laiko dimensijų gali būti dar įvairesnių.

Kuriamos duomenų saugyklos schema buvo pasirinkta – žvaigždės schema.

3.4.*Duomenų transformacijų šablonai*

Šiame skyriuje pateikiami duomenų transformacijų kūrimo šablonai, sukurti remiantis duomenų transformavimo operacijų tipų analize ir bandomuoju duomenų saugyklos kūrimu.

Pagrindiniai žymėjimai, naudojami aprašant duomenų transformacijų šablonus;

- ***WH_database*** – duomenų saugykla.
- ***DB_database*** – duomenų šaltinio bazė,
- ***DWLentelė*** – duomenų saugyklos lentelė,
- ***DBLentelė*** – duomenų šaltinio lentelė,
- ***PK_atributai*** – pirminio rakto atributas / atributai.

3.4.1. Dimensijos transformacija

Vienos lentelės arba dalies lentelės atributų perkėlimas:

SELECT DISTINCT ***Nurodyti_atributai***	
INTO ***WH_database.dbo.DWLentelė***	Duomenų saugyklos lentelė
FROM ***DB_database.dbo.DBLentelė***	Duomenų šaltinio lentelė
ALTER TABLE ***dbo.DWLentelė*** ADD CONSTRAINT ***PK_DWLentelė*** PRIMARY KEY CLUSTERED	Pirminių raktų sukūrimas duomenų saugyklos lentelėje.

(***Atributas***) ON [PRIMARY]	

Kelių reliacinių lentelių jungimas į vieną duomenų saugyklos lentelę:

SELECT DISTINCT ***A.Nurodyti_atributai, B.Nurodyti_atributai***	
INTO ***WH_database.dbo.DWLentelė***	Duomenų saugyklos lentelė
FROM ***DB_database.dbo.DBLentelė*** A	Duomenų šaltinio lentelė
INNER JOIN ***DB_database.dbo.DBLentelė*** B	Kita duomenų šaltinio lentelė
on A.ID = B.ID	Duomenų šaltinio lentelių sujungimas
ALTER TABLE ***dbo.DWLentelė*** ADD CONSTRAINT	
PK_DWLentelė PRIMARY KEY CLUSTERED (***Atributas***) ON [PRIMARY]	Pirminių raktų sukūrimas duomenų saugyklos lentelėje.

3.4.2. Fakto transformacija

SELECT ***Nurodyti_atributai***	
INTO ***WH_database.dbo.DWLentelė***	Duomenų saugyklos lentelė
FROM ***DB_database.dbo.DBLentelė***	Duomenų šaltinio lentelė
ALTER TABLE ***WH_database.dbo.DWLentelė*** ADD CONSTRAINT ***PK_DWLentelė*** PRIMARY KEY CLUSTERED (***PK_atributai***) ON [PRIMARY]	Pirminių raktų sukūrimas duomenų saugyklos lentelėje.
ALTER TABLE ***dbo.DWLentelė*** ADD CONSTRAINT ***FK_DWLentelė_DWLentelėKita*** FOREIGN KEY (***Atributas***) REFERENCES ***dbo. DWLentelėDimensija*** (***Atributas***)	CIKLAS: fakto lentelės ryšiams su dimensijomis sukurti. Vykdoma tiek kartų, kiek yra dimensijų lentelių.

3.4.3. Laiko dimensijos transformacija

SELECT DISTINCT ***DateTime_atributas***,	
year(***DateTime_atributas***) metai,	Išskiriami metai.
month(***DateTime_atributas***) menuo,	Išskiriamas mėnuo.
day(***DateTime_atributas***) diena	Išskiriama diena.
INTO ***WH_database.dbo.DWLentelė***	Duomenų saugyklos lentelė

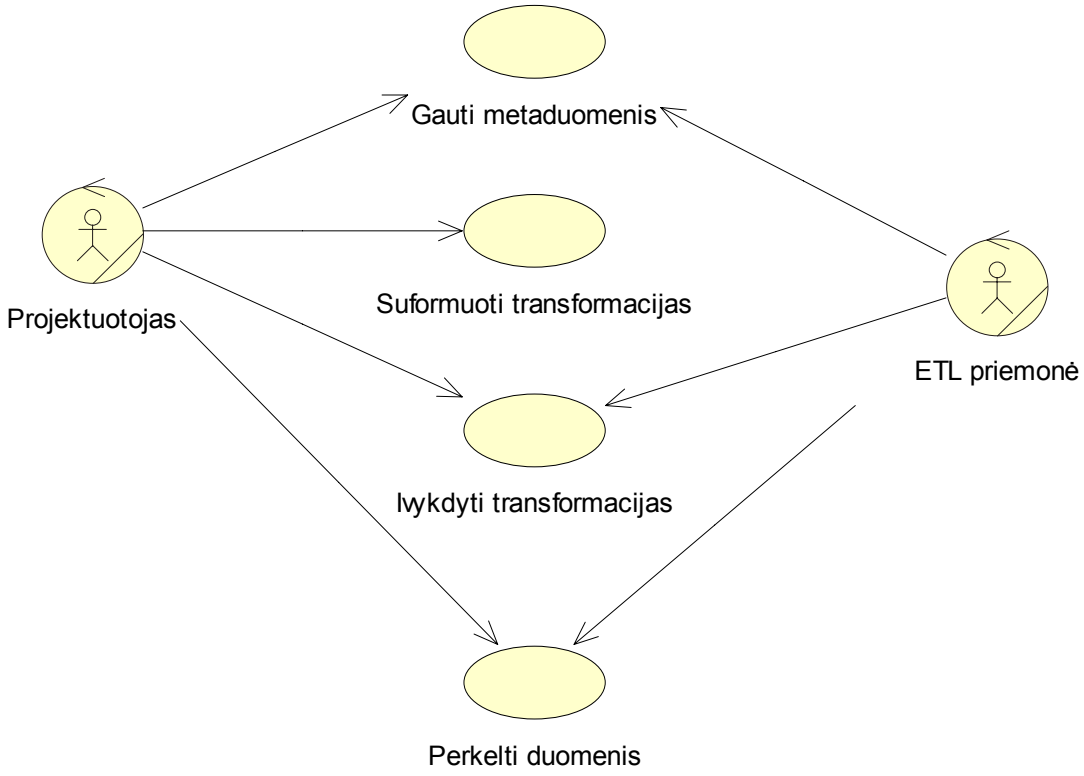
FROM ***DB_database.dbo.DB.Lentelė***	Duomenų šaltinio lentelė
ALTER TABLE ***dbo.DWLentelė*** ADD CONSTRAINT ***PK_DWLentelė*** PRIMARY KEY CLUSTERED (***DateTime_atributas***) ON [PRIMARY]	Pirminių raktų sukūrimas duomenų saugyklos lentelėje.

4. Transformacijų šablonais grindžiamos ETL priemonės reikalavimai

Šiame skyriuje pateikiami reikalavimai kuriamai ETL priemonei.

4.1. *Transformacijų šablonais grindžiamos ETL priemonės reikalavimų specifikacija*

Saugyklos projektuotojo ir ETL priemonės atliekamos funkcijos pateiktos 22 paveiksle.



22 pav. Saugyklos projektuotojo ir ETL priemonės atliekamos funkcijos

Aktoriai:

- *Projektuotojas* – atlieka metaduomenų reikalingų perkelti į duomenų saugyklą atrinkimą, duomenų transformacijų pritaikymą.
- *ETL priemonė* – prisijungia prie nurodyto duomenų šaltinio ir ištraukia metaduomenis, atlieka sukurtas transformacijas, ištraukia iš duomenų šaltinio ir perkelia į duomenų saugyklą duomenis.

Panaudojimo atvejai:

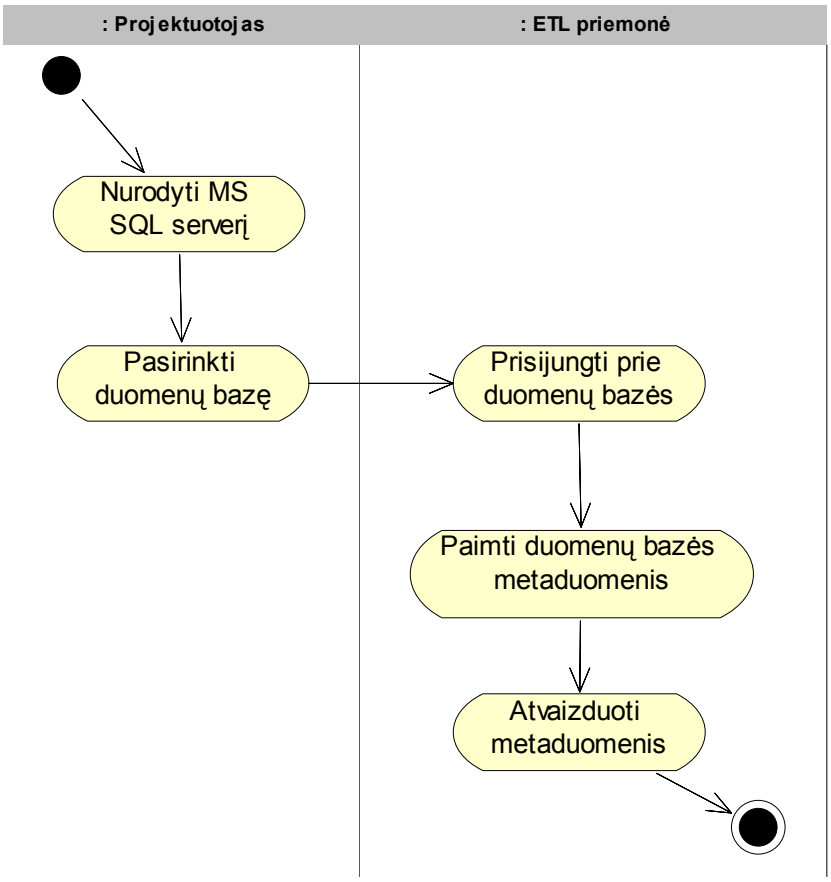
- *Gauti metaduomenis* – sukuriamas prisijungimas prie pasirinktos duomenų bazės, bei gaunami duomenų bazės metaduomenys: lentelės, lentelių atributai ir ryšiai tarp lentelių.

- *Suformuoti transformacijas* – kiekvienai transformacijai atskirai atrenkami reikalingi metaduomenys ir sukuriamos transformacijos.
- *Įvykdyti transformacijas* –įvykdomos visos sukurtos duomenų transformacijos ir metaduomenys išsaugomi duomenų saugykloje.
- *Perkelti duomenis* – iš duomenų šaltinio pagal atliktas metaduomenų transformacijas paimami reikalingi duomenys ir išsaugomi duomenų saugykloje.

3 lentelė

Panaudojimo atvejo „Gauti metaduomenis" specifikacija

Panaudojimo atvejis	Gauti metaduomenis
Nr.	PA1
Aktoriai	Projektuotojas
Sistema	ETL priemonė
Prieš-sąlyga	Turi būti pasirinkta duomenų šaltinio duomenų bazė
Pagrindinis įvykių srautas	*Sistemos reakcija ir sprendimai*
1. Nurodyti MS SQL Server. 2. Pasirinkti duomenų bazę.	1.1 Prisijungti prie nurodyto MS SQL Server. 2.1 Prisijungti prie nurodytos duomenų bazės 2.2 Gauti visų DB lentelių pavadinimus. 2.3 Gauti ryšius tarp lentelių. 2.4 Gauti kiekvienos lentelės atributų sąrašą. 2.5 Gauti kiekvienos lentelės PK.
Po-sąlyga	Ištraukti duomenų bazės metaduomenys ir pateikti vartotojui.
Alternatyvos	1.1.a Nepavyko prisijungti prie nurodyto MS SQL Server. 1.1.b Nurodytas MS SQL Server nerastas. 2.1.a Nurodyta duomenų bazė nerasta. 2.2.a – 2.5.a. Nepavyko ištraukti nurodytų metaduomenų
Vykdymo variantai	Projektuotojas pasirenka duomenų bazę.
Veiklos taisyklės	-
Specialūs (nefunkciniai) reikalavimai	-
Kitos sistemos su kuriomis sąveikauja sistema	MS SQL Server duomenų bazė.
Pastabos	-
Neišspręstos problemos	-
Sudarė	Kristina Paulavičiūtė
Data	2005-05-16

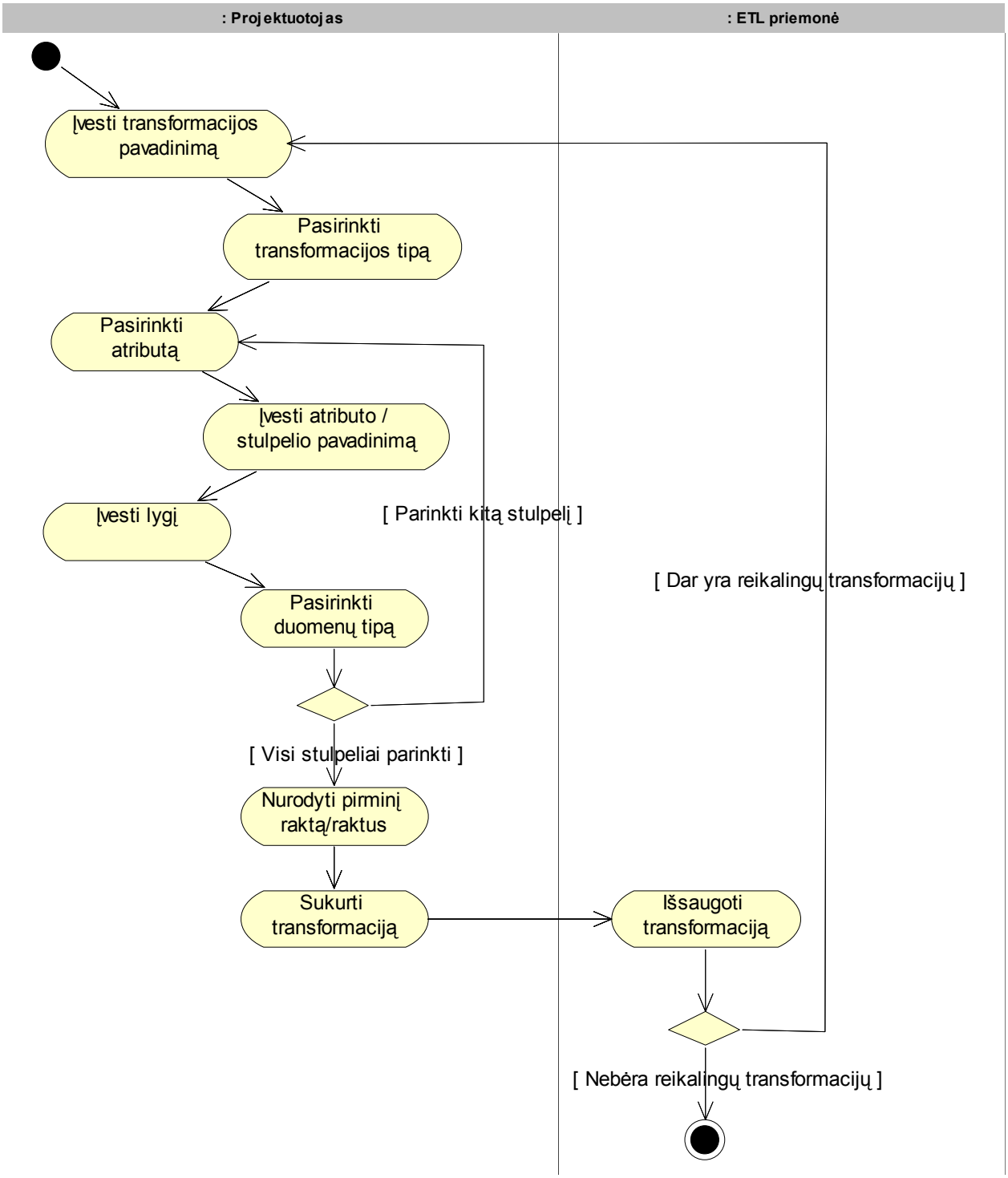


23 pav. Metaduomenų gavimo procesas

4 lentelė

Panaudojimo atvejo „Suformuoti transformacijas" specifikacija

Panaudojimo atvejis	Suformuoti transformacijas
Nr.	PA2
Aktoriai	Projektuotojas
Sistema	ETL priemonė
Prieš-sąlyga	
Pagrindinis įvykių srautas	*Sistemos reakcija ir sprendimai*
1. Įvesti transformacijos (duomenų saugyklos lentelės) pavadinimą, 2. Pasirinkti transformacijos tipą. 3. Pasirinkti atributus. 4. Įvesti stulpelio pavadinimą. 5. Įvesti lygį. 6. Pasirinkti duomenų tipą. 7. Nurodyti pirminį raktą / raktus.	7.1. Suformuojama transformacija
Po-sąlyga	Suformuotos visos duomenų saugyklos sukūrimui reikalingos transformacijos.
Alternatyvos	Neįvesti visi reikalingi duomenys.
Vykdymo variantai	Lentelės, lentelių atributai, pirminiai raktai yra pasirenkami iš pateikto sąrašo.
Veiklos taisyklės	-
Specialūs (nefunkciniai) reikalavimai	Metaduomenys turi būti pateikti vartotojui suprantamu variantu (atvaizduoti ne tik lentelių atributai, jų tipai bet ir ryšiai tarp lentelių)
Kitos sistemos su kuriomis sąveikauja sistema	-
Pastabos	-
Neišspręstos problemos	-
Sudarė	Kristina Paulavičiūtė
Data	2005-05-16

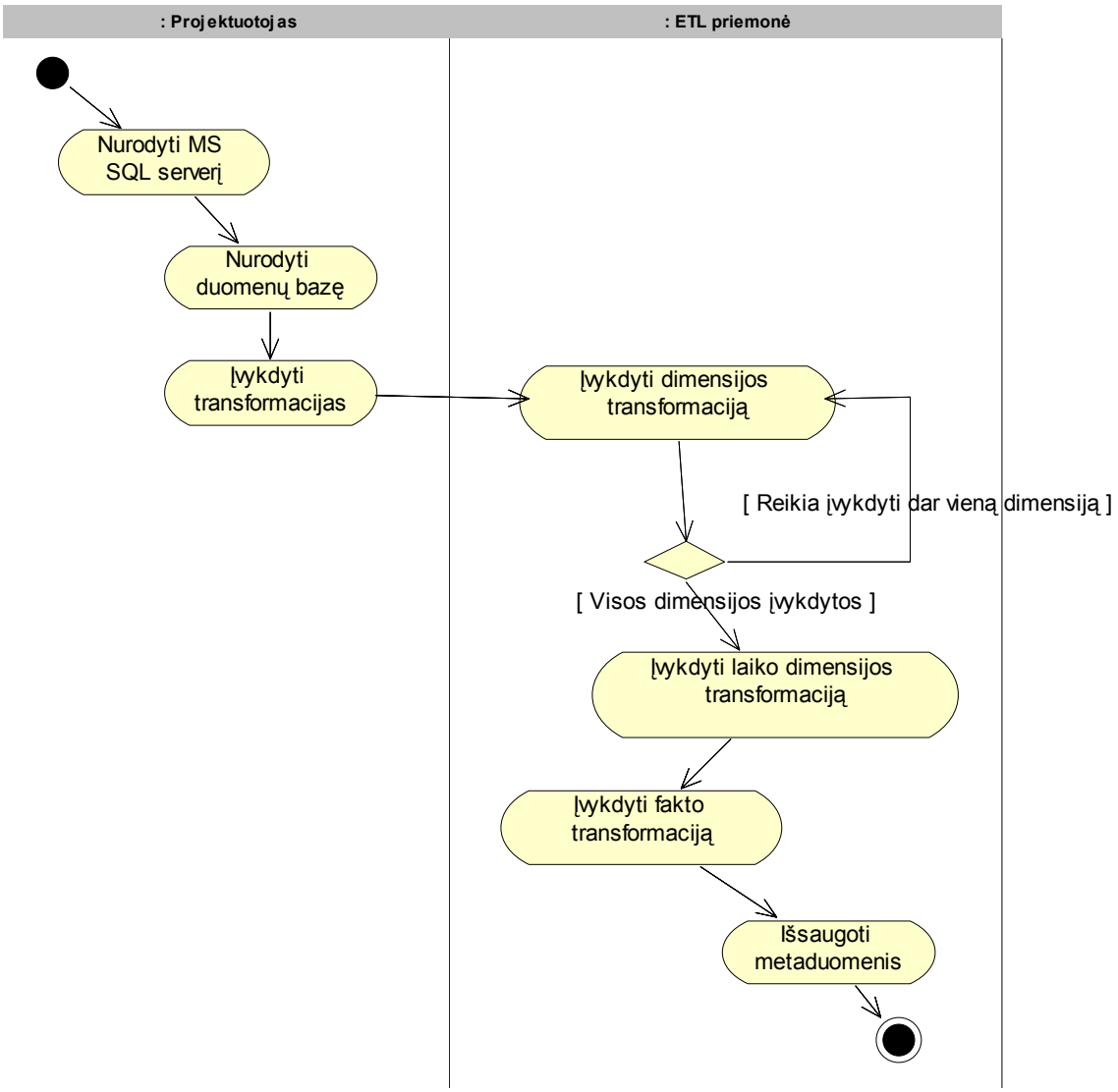


24 pav. Transformacijų suformavimo procesas

5 lentelė

Panaudojimo atvejo „Įvykdyti transformacijas" specifikacija

Panaudojimo atvejis	Įvykdyti transformacijas
Nr.	PA3
Aktoriai	Projektuotojas
Sistema	ETL priemonė
Prieš-sąlyga	
Pagrindinis įvykių srautas	*Sistemos reakcija ir sprendimai*
1. Nurodyti duomenų saugyklos MS SQL Server. 2. Nurodyti duomenų bazę. 3. Įvykdyti transformacijas	3.1. Įvykdyti visų dimensijų transformacijas. 3.2. Laiko dimensijos transformaciją. 3.3. Įvykdyti fakto transformaciją.
Po-sąlyga	Įvykdytos visos metaduomenų transformacijos.
Alternatyvos	-
Vykdymo variantai	-
Veiklos taisyklės	-
Specialūs (nefunkciniai) reikalavimai	Įvykdytos transformacijos turi būti išsaugotos duomenų bazėje.
Kitos sistemos su kuriomis sąveikauja sistema	-
Pastabos	-
Neišspręstos problemos	-
Sudarė	Kristina Paulavičiūtė
Data	2005-05-16

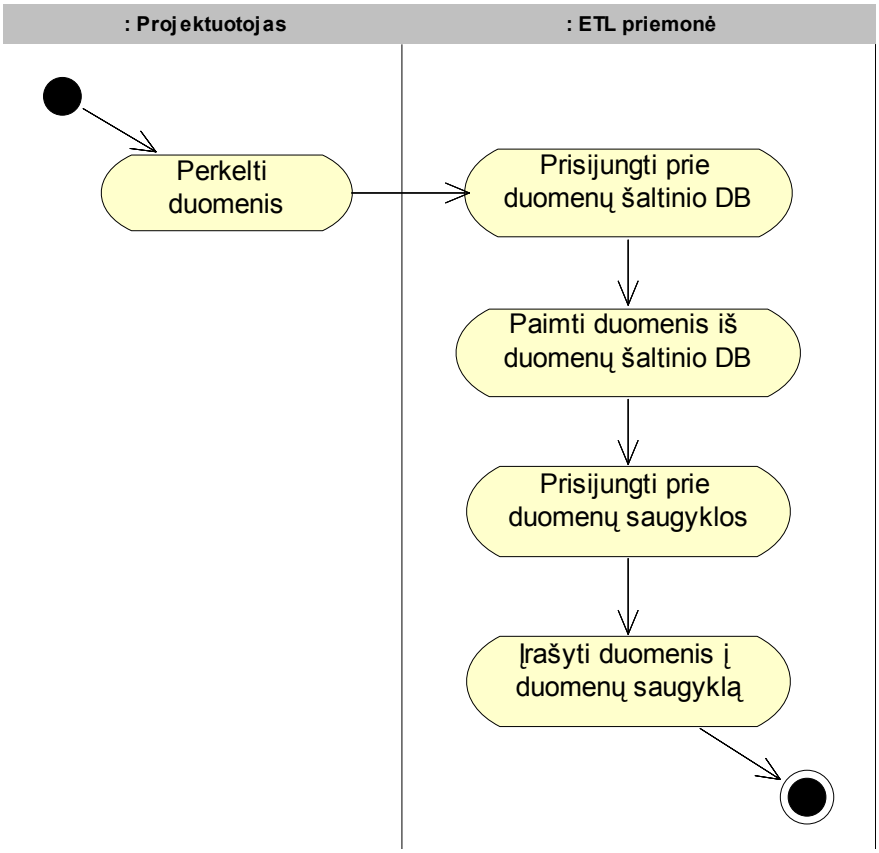


25 pav. Transformacijų įvykdymo procesas

6 lentelė

Panaudojimo atvejo „Perkelti duomenis" specifikacija

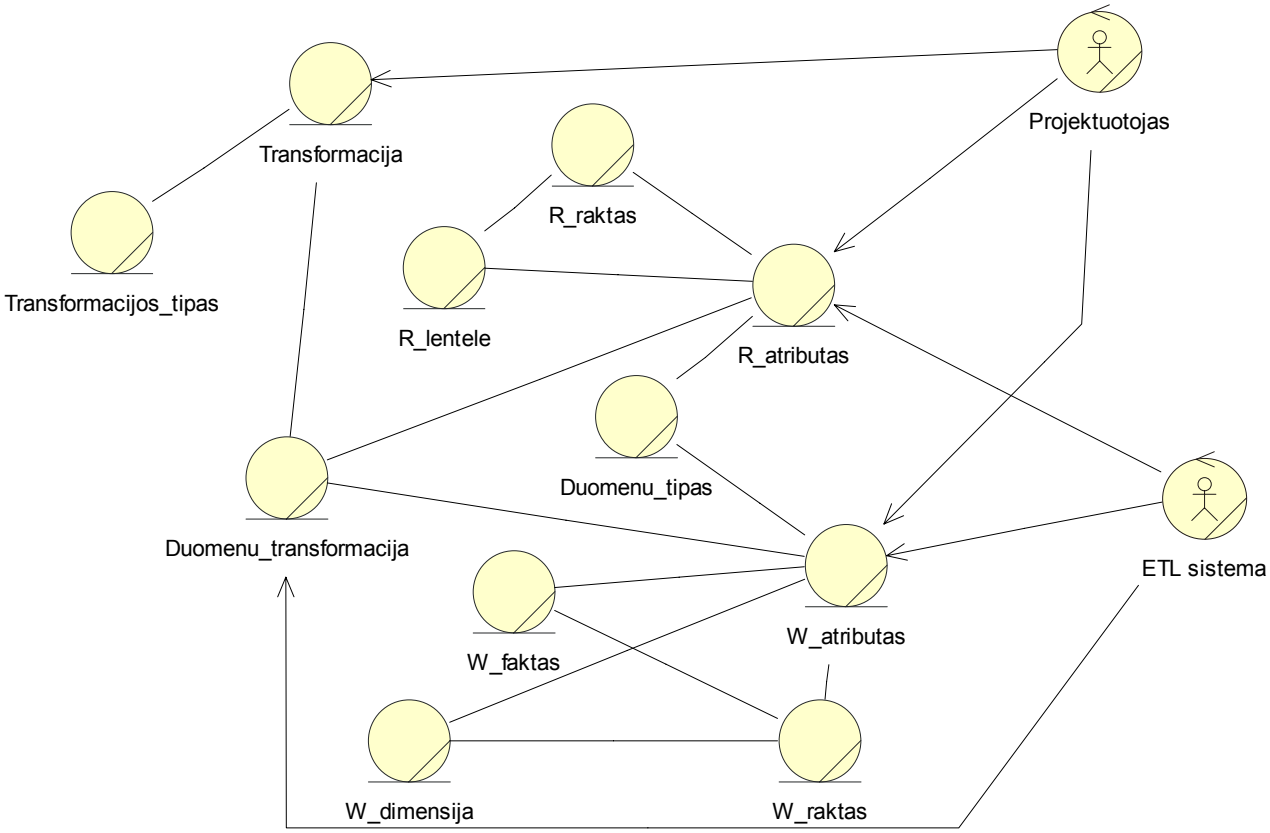
Panaudojimo atvejis	Sukurti saugyklą
Nr.	PA4
Aktoriai	Projektuotojas
Sistema	ETL priemonė
Prieš-sąlyga	
Pagrindinis įvykių srautas	*Sistemos reakcija ir sprendimai*
1. Perkelti duomenis	1.1. Prisijungti prie duomenų šaltinio. 1.2. Ištraukti duomenis. 1.3. Prisijungti prie duomenų saugyklos. 1.4. Įrašyti duomenis.
Po-sąlyga	Duomenys perkelti į duomenų saugyklą.
Alternatyvos	1.1.a. Nepavyko prisijungti prie duomenų šaltinio. 1.2.a. Nepavyko ištraukti duomenų. 1.3.a. Nepavyko prisijungti prie duomenų saugyklos. 1.4.a.Nepavyko įrašyti duomenų.
Vykdymo variantai	-
Veiklos taisyklės	-
Specialūs (nefunkciniai) reikalavimai	-
Kitos sistemos su kuriomis sąveikauja sistema	MS SQL Server duomenų bazė.
Pastabos	-
Neišspręstos problemos	-
Sudarė	Kristina Paulavičiūtė
Data	2005-05-16



26 pav. Duomenų perkėlimo procesas

4.2. *Transformavimo metamodelis*

Pagrindinės dalykinės srities esybės pateiktos 27 paveiksle.

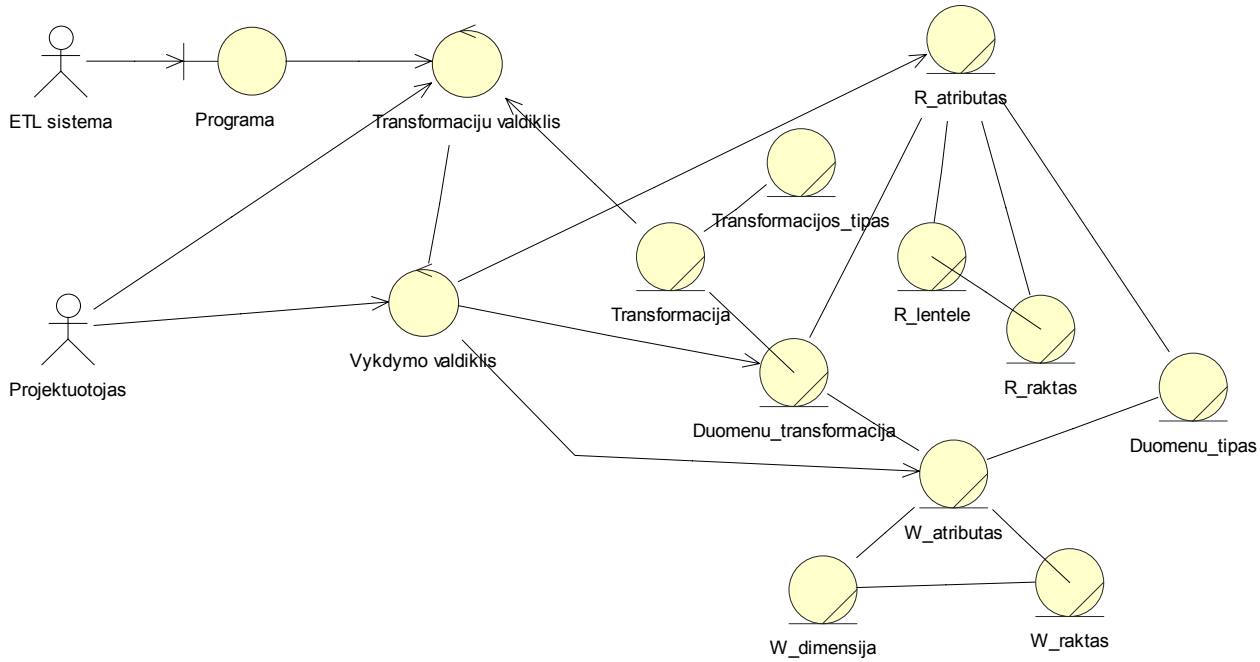


27 pav. Transformacijų metamodelio konceptai

Esybės pavadinime su „W“ rodo kuriamos duomenų saugyklos metaduomenų esybes, pavadinime su „R“ rodo duomenų šaltinio metaduomenų esybes. O esybės „Transformacija“, „Transformacijos tipas“ ir „Duomenų transformacija“ susijusios su pagrindinėmis kuriamos ETL sistemos esybėmis – duomenų transformacijomis.

4.3. *Transformacijų šablonais grindžiamos ETL priemonės analizės klasių diagrama*

Visos sistemos analizės klasių diagrama pateikta 28 paveiksle.



28 pav. Kuriamos ETL priemonės analizės klasių diagrama

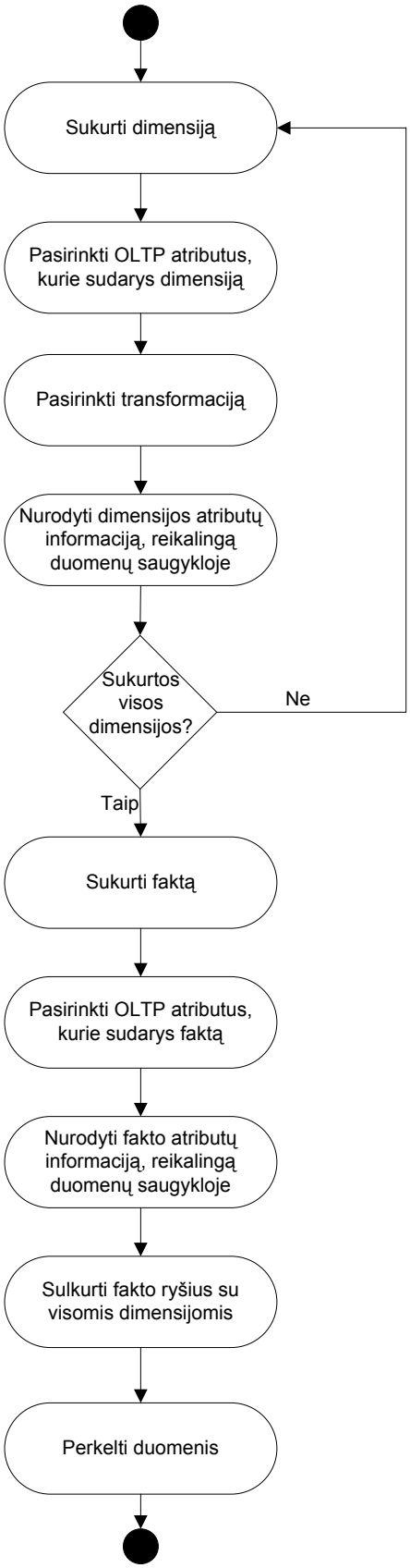
5. Transformacijų šablonais grindžiamos ETL priemonės projektas

5.1. *Transformacijų šablonais grindžiamos ETL priemonės pagrindimas ir esmės išdėstymas*

Kuriant duomenų saugyklą esamomis MS SQL priemonėmis visas transformacijas reikia rašyti rankomis programiniu kodu. Tai užima pakankamai daug laiko. Norint sutrumpint tą laiką, siekiama sukurti priemonę, kurioje būtų realizuotos duomenų transformavimo operacijos. O pats duomenų transformacijų kūrimas vykdytų „drag and drop“ principu, ypač patogiu sistemos naudotojui.

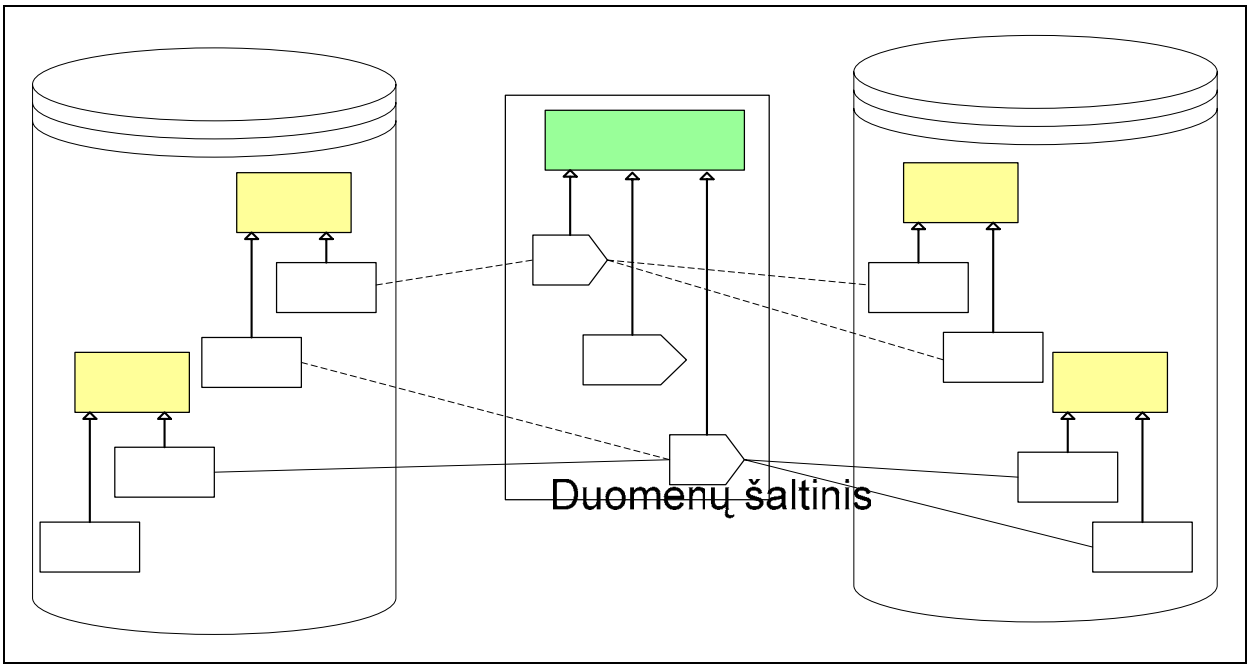
Taip pat sistemoje būtų saugomi visų sukurtų duomenų saugyklų bei įvykdytų transformacijų duomenys, kad esant poreikiui būtų galima pakoreguoti arba įvykdyti iš naujo.

Kuriamos sistemos pagrindinis procesas – duomenų saugyklos kūrimo procesas. Principinė proceso schema (29 pav.) nusako, kad kiekvienai dimensijai arba faktui turi būti parinkti OLTP atributai ir nustatytos tų atributų transformacijos. Kiekvienoje saugykloje turi būti daugiau nei viena dimensija ir vienas faktas, turintis ryšius su visomis dimensijomis.



29 pav. Duomenų saugyklos kūrimo procesas

Kuriamos ETL priemonės veikimo principas (30 pav.): pasirenkame duomenų šaltinio duomenų atributą (stulpelį), pasirinktam atributui pasirenkame transformacijos tipą. Pasirenkame tiek atributų, kiek dimensijoje turi būti atributų. Sistema pati sukuria dimensiją arba faktą.



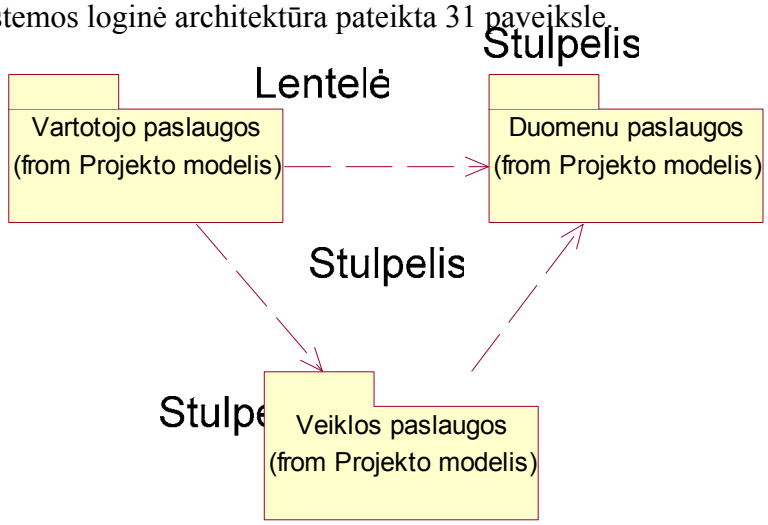
30 pav. Kuriamos ETL priemonės veikimo principas

5.2. *Transformacijų šablonais grindžiamos ETL priemonės architektūra*

5.2.1. Transformacijų šablonais grindžiamos ETL priemonės architektūra

Pereinant nuo kuriamos sistemos analizės prie projekto, sukuriami projekto elementai, kurių atitikmuo su analizės elementais pavaizduojami trasų diagramose (31 – 33 pav.)

Projektuojamos sistemos loginė architektūra pateikta 31 paveiksle.



31 pav. Sistemos loginė architektūra

Sistemos architektūra sudaryta pagal išskirtas veiklos sritis:

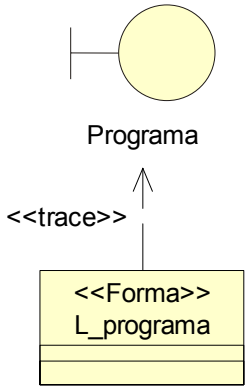
Vartotojo paslaugos – vartotojo sąsajos projektavimas;

Veiklos paslaugos – programos logikos projektavimas;

Duomenų paslaugų – duomenų bazės ir jos prieigos projektavimas.

5.2.2. Transformacijų šablonais grindžiamos ETL priemonės vartotojo paslaugos

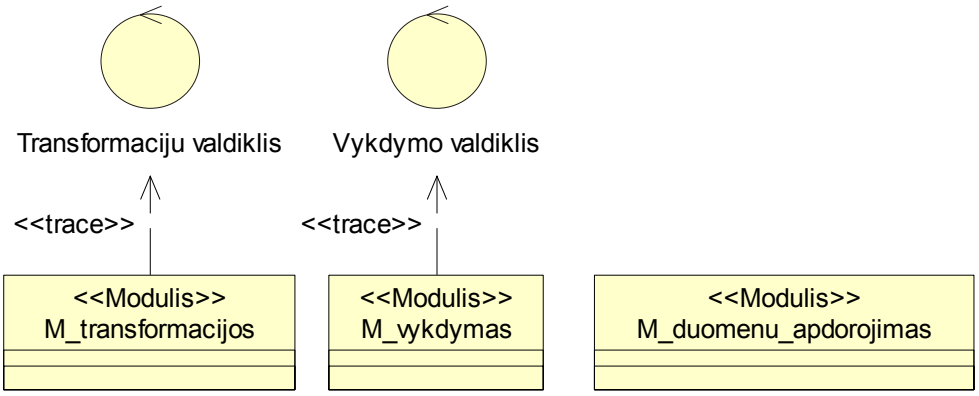
Kol kas planuojamos minimalistines vartotojo paslaugos: sistemą sudarys tik vienas pagrindinis langas (dar keli langai bus skirti sistemos administravimui). Vartotojo paslaugų trasų diagrama pateikta 32 paveiksle.



32 pav. Vartotojo paslaugų trasų diagrama

5.2.3. Transformacijų šablonais grindžiamos ETL priemonės veiklos paslaugos

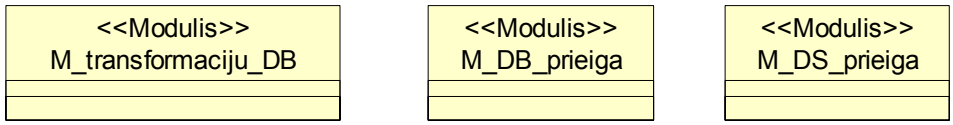
Veiklos paslaugų trasų diagrama pateikta 33 paveiksle. Modulis „M_Transformacijos“ atsakingas už transformacijų ištraukimą iš duomenų bazės, bei jų suformavimą konkrečiam atvejui. Modulis „M_Vykdymas“ atsakingas už transformacijų įvykdymą, bei visų reikalingų duomenų išsaugojimą duomenų bazėje. Sukuriamas naujas modulis „M_duomenu_apdorojimas“, atsakingas už metaduomenų ištraukimą iš duomenų šaltinio, duomenų perkėlimą iš duomenų šaltinio į duomenų saugyklą.



33 pav. Transformavimo paslaugų trasų diagrama

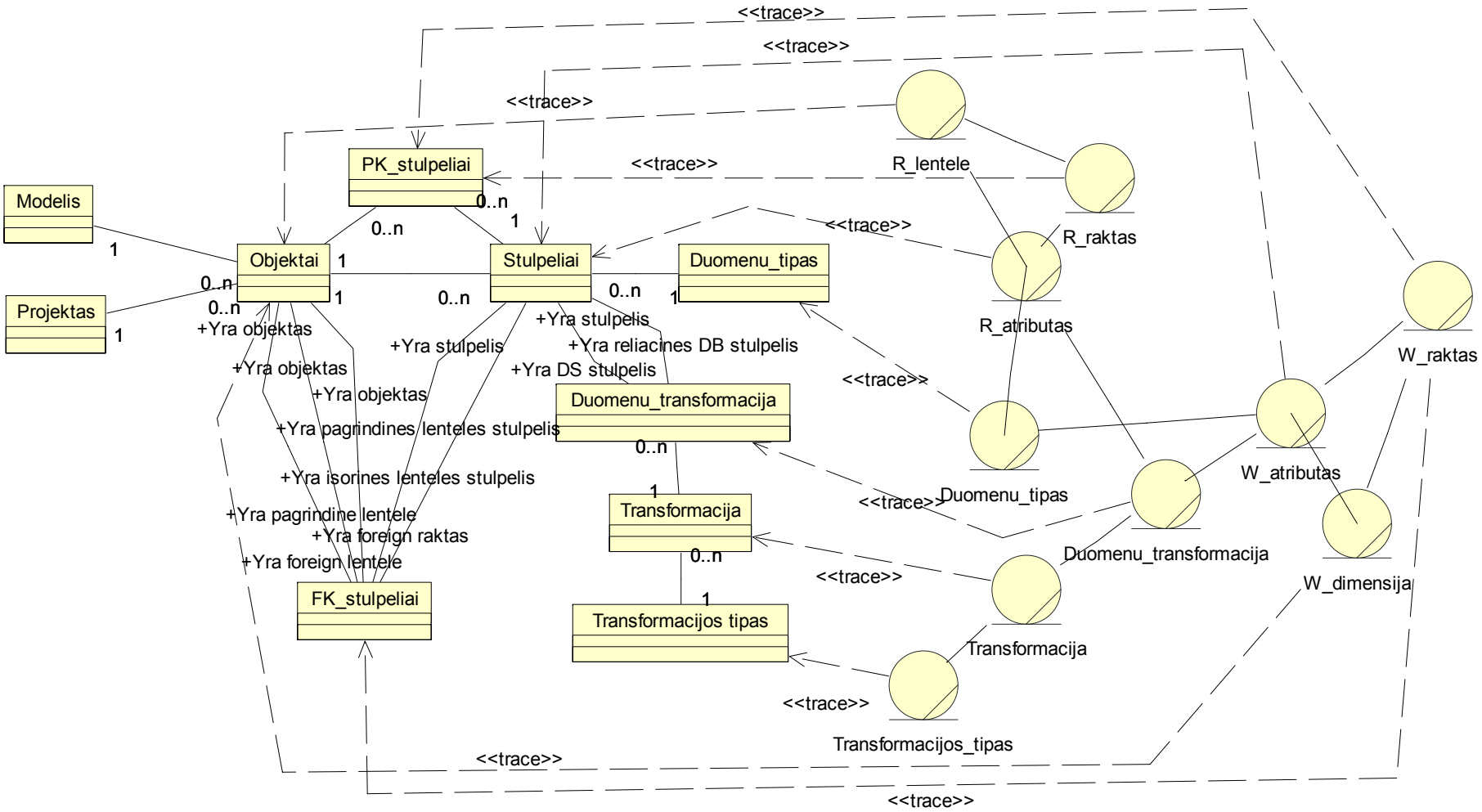
5.2.4. Transformacijų šablonais grindžiamos ETL priemonės duomenų paslaugos

Sistema dirba su dviem duomenų bazėmis: duomenų šaliniu ir duomenų saugojimo duomenų baze. Šias dvi duomenų bazes realizuoja atitinkamai klasės: M_DB_prieiga ir M_DS_prieiga. Dar viena duomenų bazė reikalinga kuriamai sistemai: duomenų šaltinio ir duomenų saugyklos metaduomenų saugojimui bei duomenų transformacijų šablonus, transformacijų pritaikymą konkretiems atvejams (klasė M_transformaciju_DB) (34 pav.).

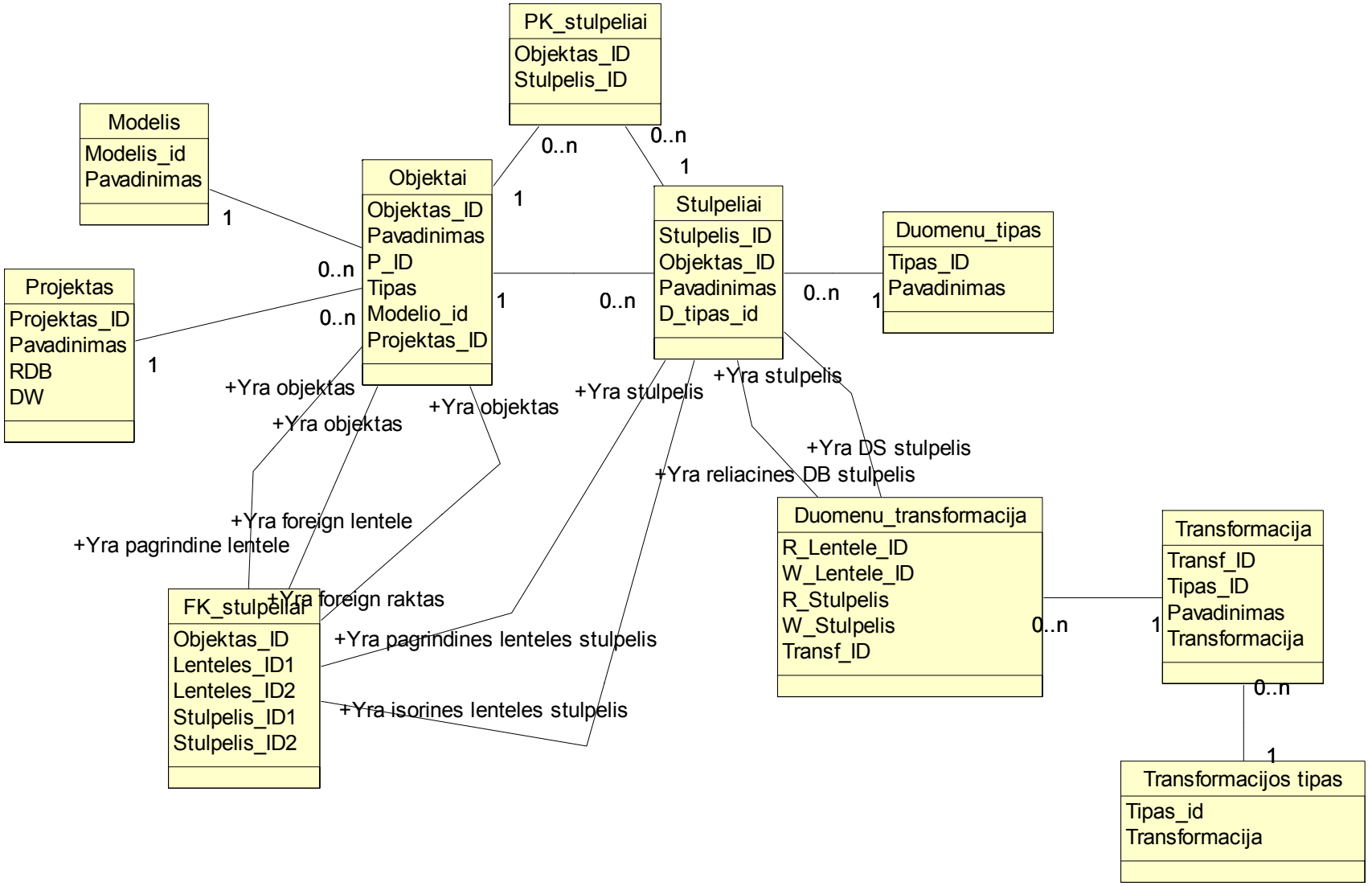


34 pav. Duomenų paslaugų klasės

Dalykinės srities klasių trasų diagrama pateikta 35 paveiksle. Kadangi reliacinės duomenų bazės ir duomenų saugyklos metaduomenų saugojimo struktūra tokia pati, nuspręsta duomenis saugoti bendrose lentelėse, įvedant reliacinės arba duomenų saugyklos požymį (modelis). Toje pačioje duomenų saugykloje bus saugomi ne vieno projekto duomenys, todėl įvesta papildoma klasė (projektas).



35 pav. Dalykinės srities klasių trasų diagrama

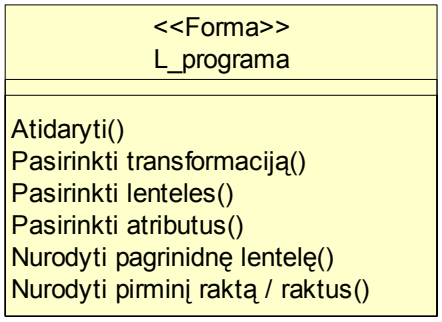


36 pav. Transformacijų metamodelis

5.3. *Transformacijų šablonais grindžiamos ETL priemonės detalus projektas*

Detalizuotos vartotojo, veiklos ir duomenų paslaugų klasių diagramos pateiktos 37 – 39 paveiksluose.

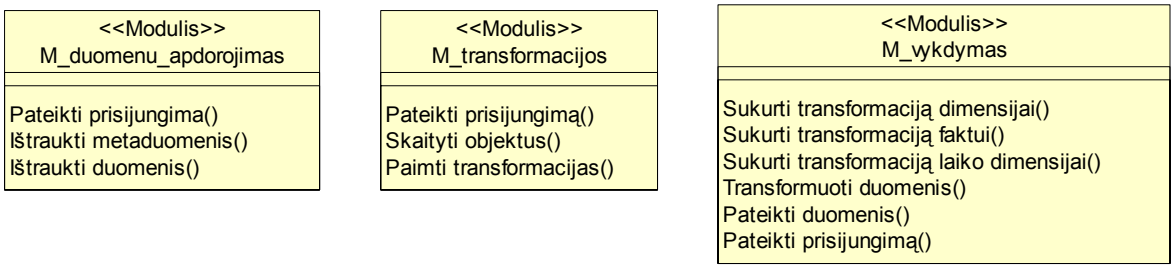
Pagrindinės vartotojo paslaugų operacijos susijusios su dimensijų bei fakto formavimu: pasirinkti transformaciją, atributus, lentelę, nurodyti pirminį raktą (37 pav.).



37 pav. Detali vartotojo paslaugų klasių diagrama

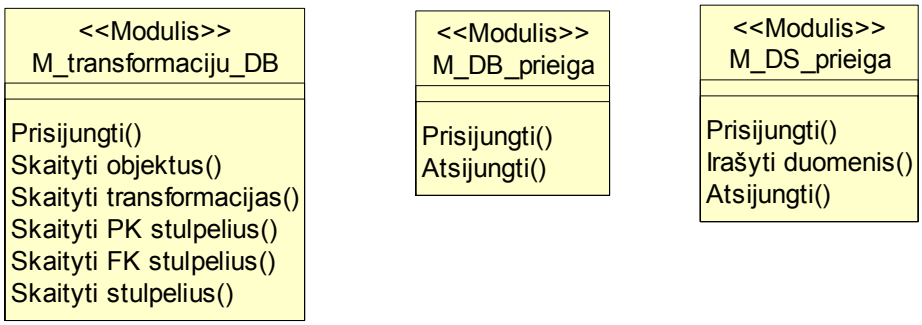
Veiklos paslaugų operacijos skiriasi į tris dalis (38 pav.):

- Duomenų apdorojimas – darbas su duomenimis,
- Transformacijų kūrimas – dimensijų ir fakto kūrimas.
- Transformacijos – transformacijų tipų valdymas.



38 pav. Detali veiklos paslaugų klasių diagrama

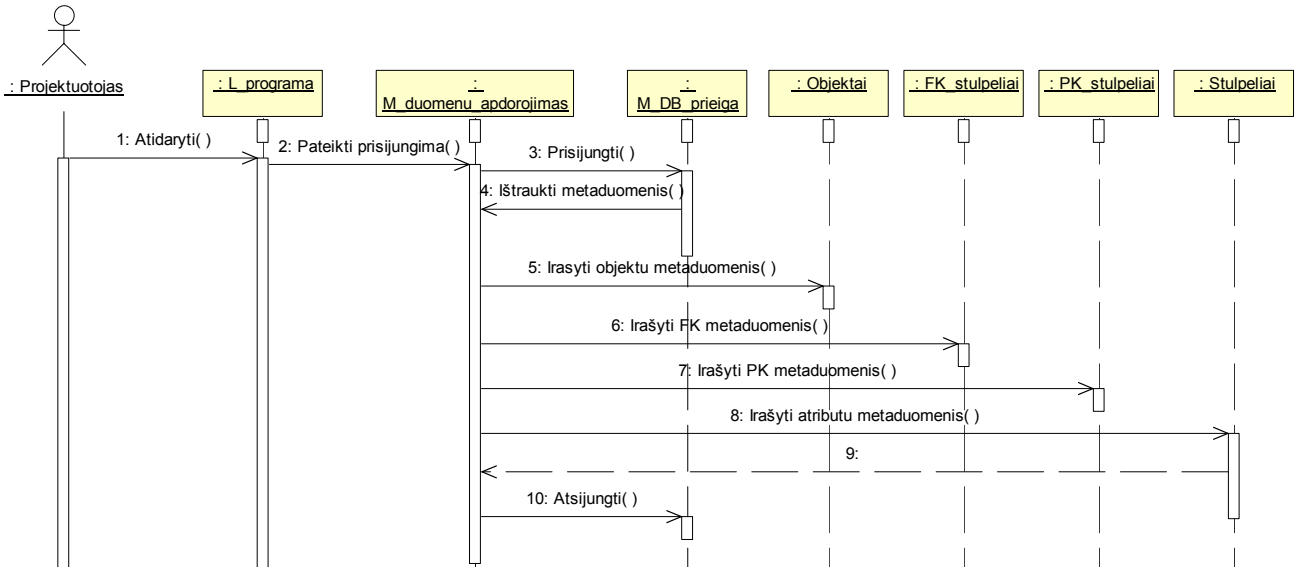
Duomenų paslaugos koordinuoja duomenų prisijungimus prie duomenų šaltinio duomenų bazės, duomenų saugyklos ir kuriamos sistemos duomenų bazės (39 pav.).



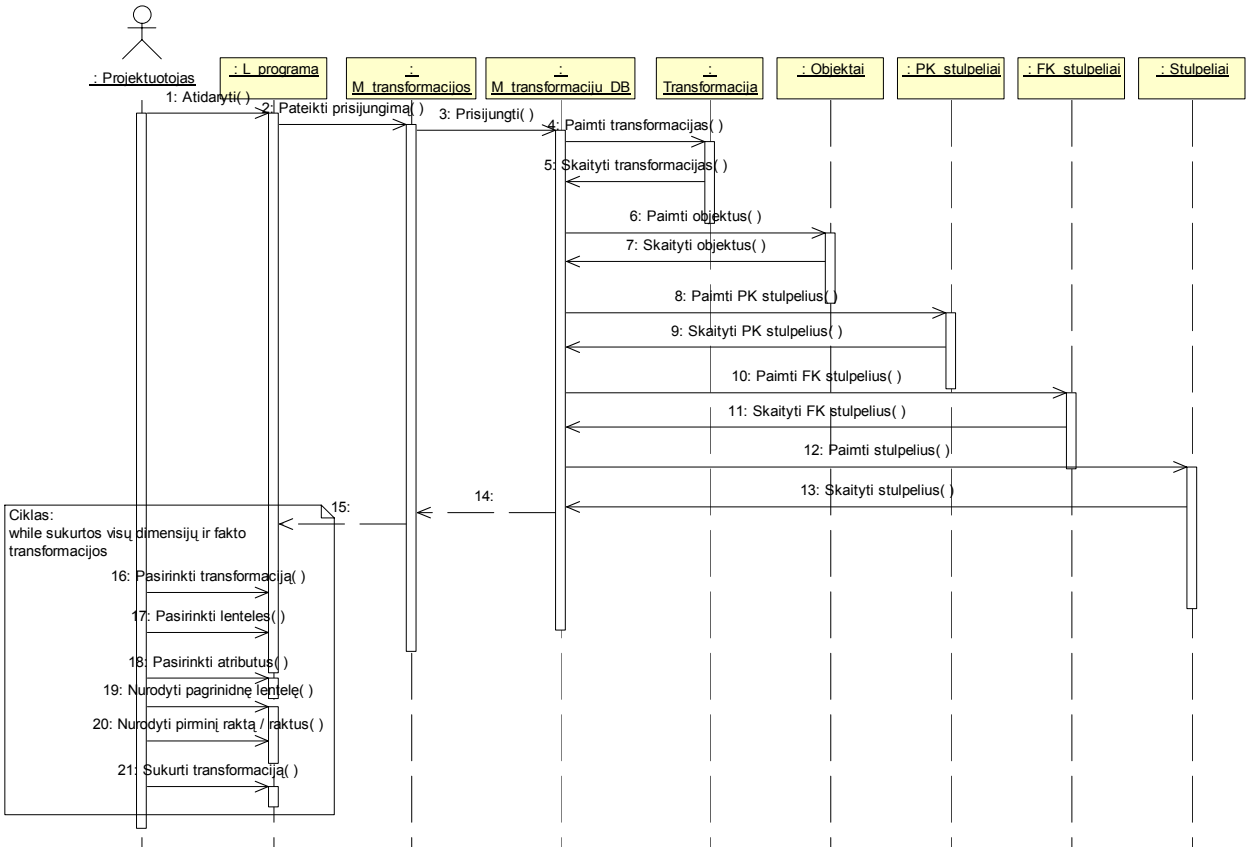
39 pav. Detali duomenų paslaugų klasių diagrama

5.4. *Transformacijų šablonais grindžiamos transformavimo priemonės elgsenos modelis*

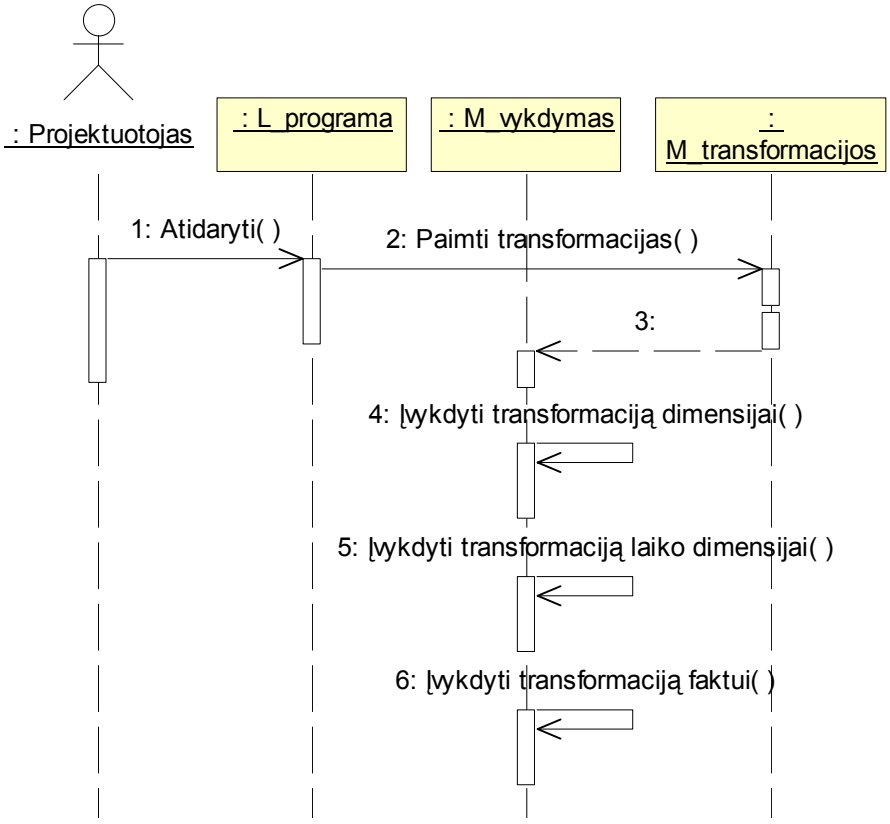
Kuriamoje ETL priemonėje realizuojamų funkcijų sekų diagramos pateiktos 40 – 43 paveiksluose.



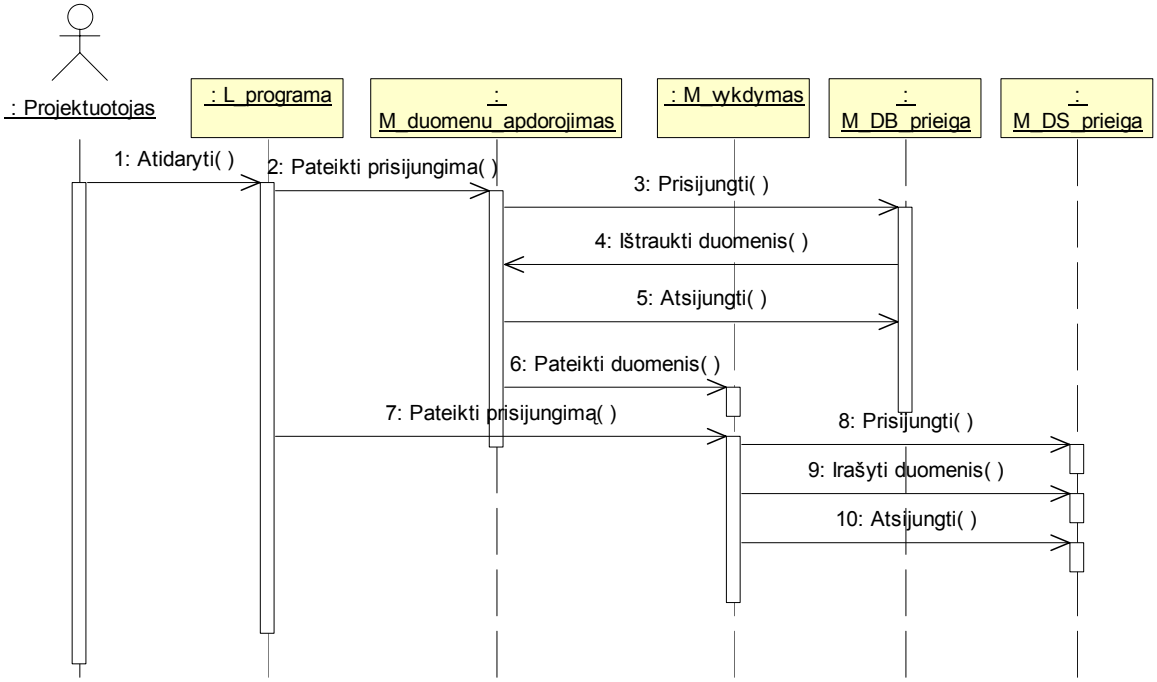
40 pav. Metaduomenų gavimo sekų diagrama



41 pav. Transformacijų kūrimo sekų diagrama



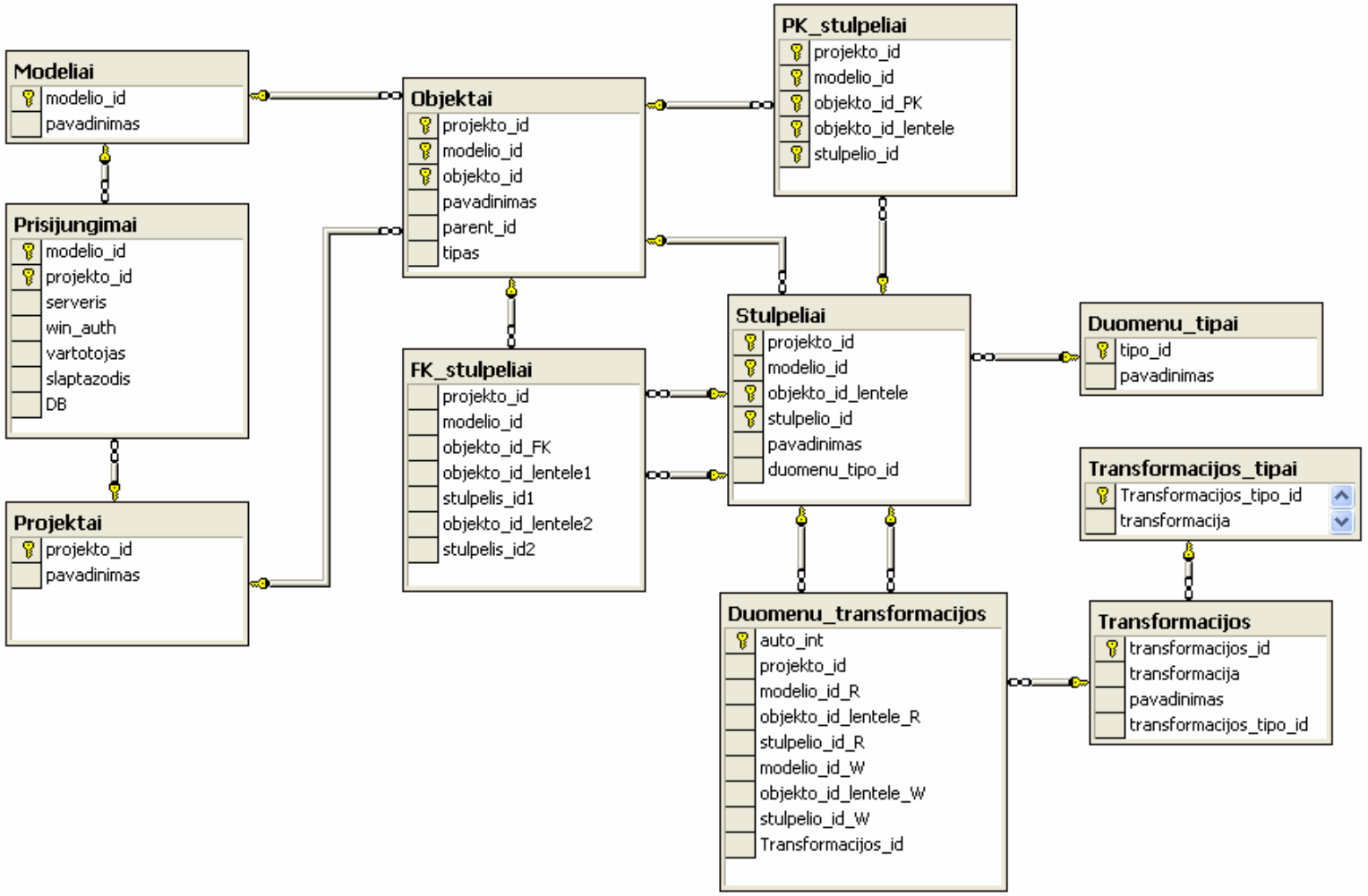
42 pav. Transformacijų vykdymo sekų



43 pav. Duomenų perkėlimo sekų

5.5. *Transformacijų šablonais grindžiamos ETL priemonės duomenų bazės schema*

Kuriamos sistemos duomenų bazės schema 44 paveiksle.

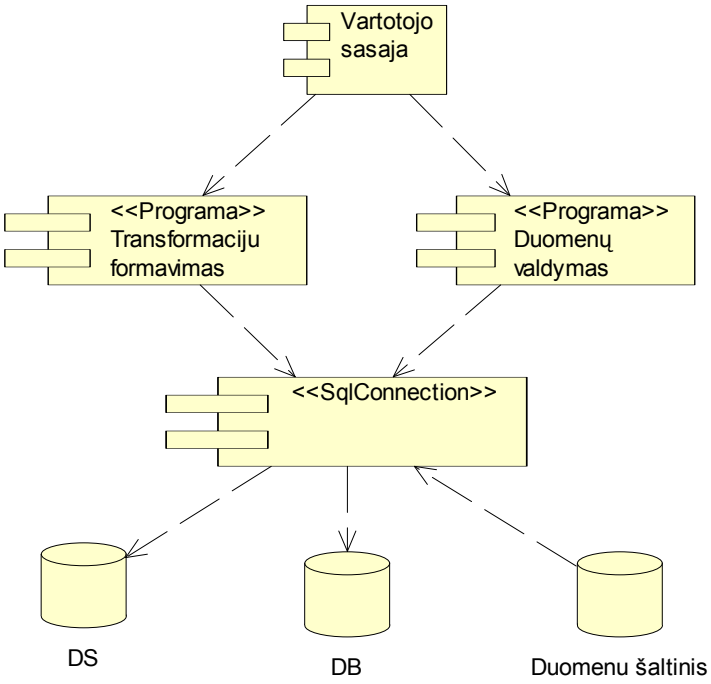


44 pav. Duomenų bazės schema

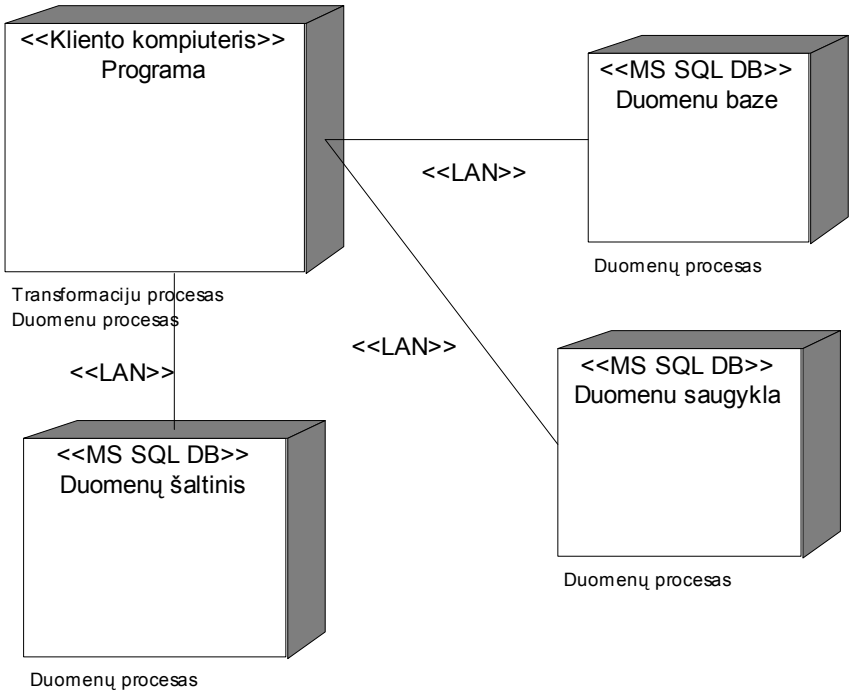
Detalus duomenų bazės lentelių ir jų atributų aprašymas pateiktas dokumento pirmame priede.

5.6. *Transformacijų šablonais grindžiamos ETL priemonės realizacijos modelis*

Sistemos komponentai patiekti 45 paveiksle, o jų fizinis išsidėstymas 46 paveiksle.



45 pav. ETL priemonės komponentų diagrama



46 pav. ETL priemonės įdiegimo diagrama

6. Transformacijų šablonais grindžiamos ETL priemonės realizacija

6.1. *Transformacijų šablonais grindžiamos ETL priemonėje realizuotos funkcijos*

Sukurtoje ETL priemonėje realizuotos projektavimo dalyje apibrėžtos pagrindinės funkcijos:

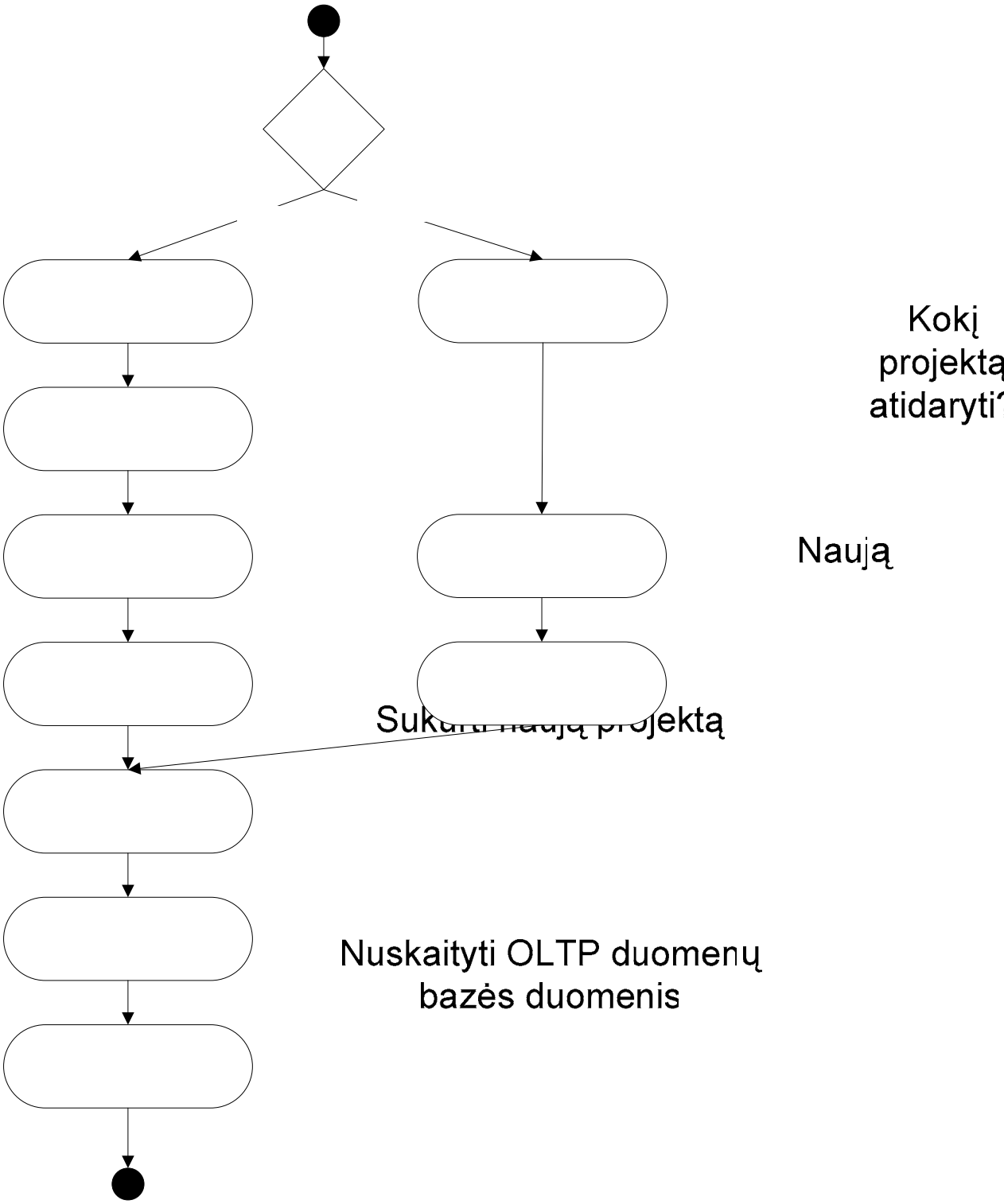
- Gauti metaduomenis,
- Suformuoti transformacijas,
- Įvykdyti transformacijas,
- Perkelti duomenis.

Papildomai realizuotos sistemos administravimo funkcijos:

- Tvarkyti projektų tipus.
- Tvarkyti duomenų transformacijų tipus.

6.2. *Transformacijų šablonais grindžiamos ETL priemonės veikimas*

Principinė naudojimosi paketu schema pateikiama 47 paveiksle.

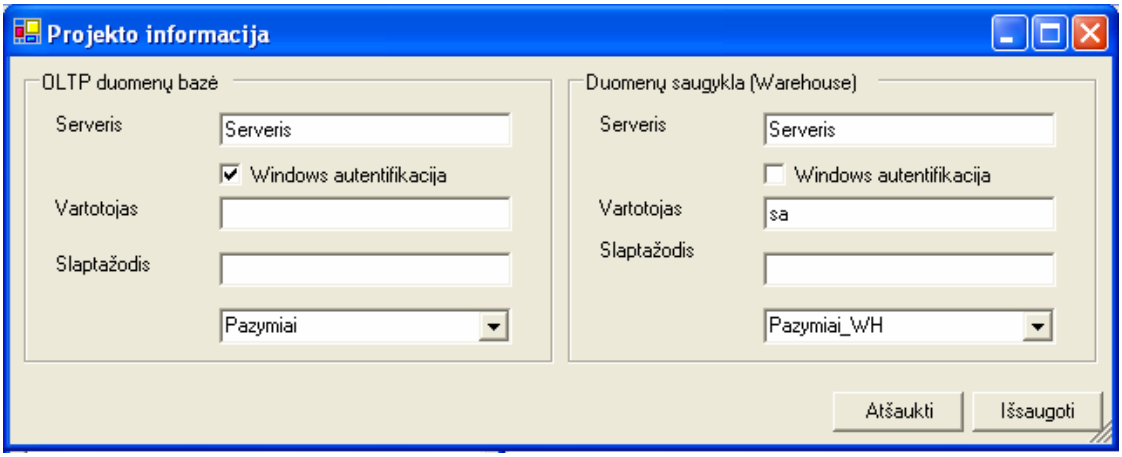


47 pav. Principinė saugyklos kūrimo struktūrinė programinė schema

Kaip atlikti sukurti naują duomenų saugyklos kūrimo projektą ir jį įgyvendinti arba, kaip koreguoti esantį, aprašyta šio skyriaus poskyriuose.

6.2.1. Naujo duomenų saugyklos projekto kūrimas

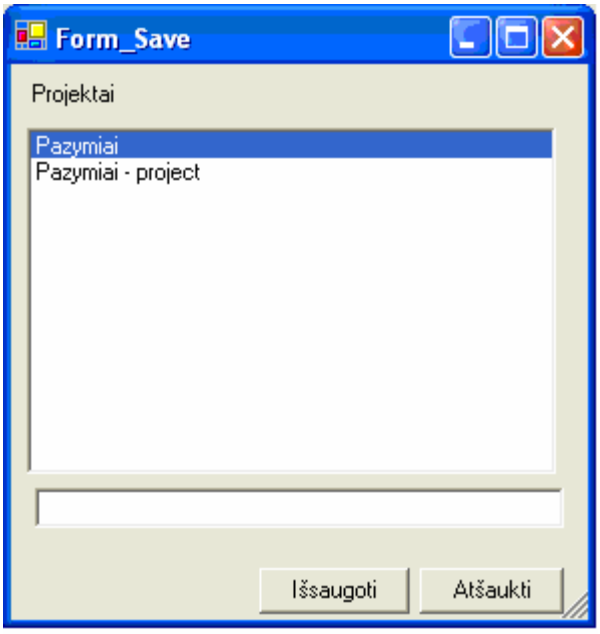
Norėdami sukurti naują projektą, duomenų saugyklos kūrimą turite atverti projekto informacijos langą (iš meniu *Projektas-> Naujas...*).



48 pav. Projekto informacijos suvedimas

Šiame lange (48 pav.) reikia nurodyti duomenų šaltinio ir duomenų saugyklos duomenų bazių duomenis: MS SQL serverio pavadinimą, prisijungimo duomenis, duomenų bazę. Įrašę visus reikiamus duomenis ir paspaudę mygtuką [Išsaugoti], išsaugote projekto nustatymus, bet neišsaugote projekto informacijos ETL sistemos duomenų bazėje.

Norėdami išsaugoti projekto informaciją sistemos duomenų bazėje, iš meniu pasirinkite *Projektas -> Išsaugoti....* (49 pav.)

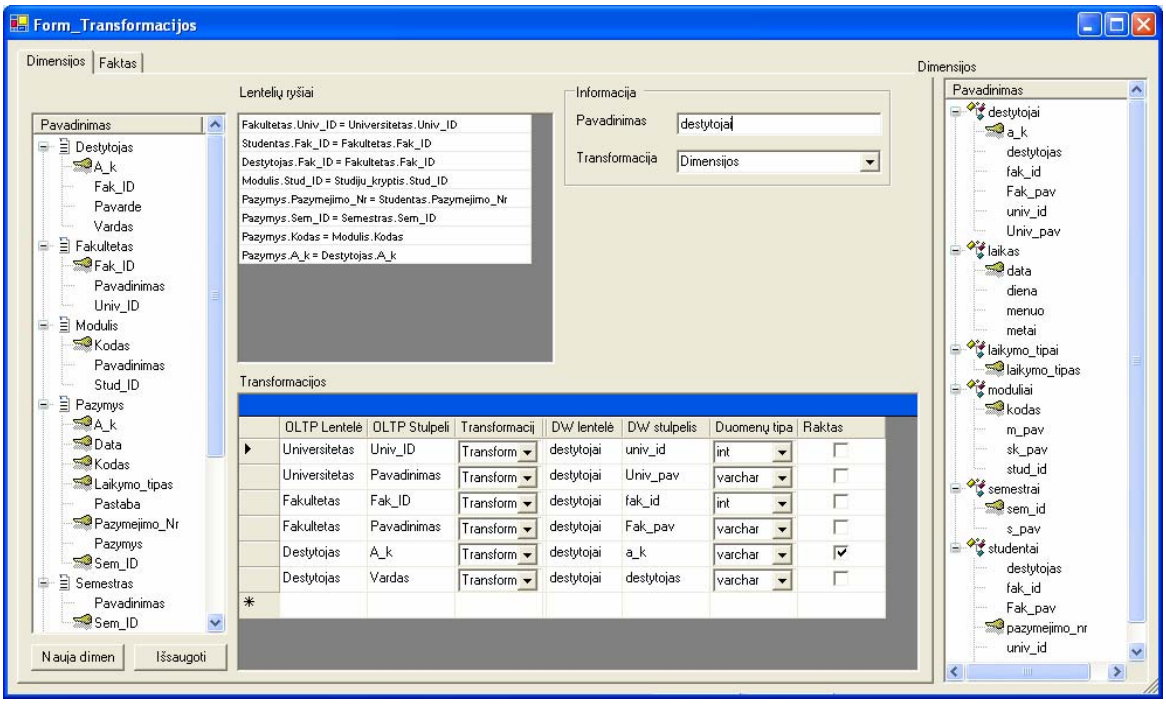


49 pav. Projekto informacijos išsaugojimo langas

Atsivėrusiame lange nurodykite projekto pavadinimą ir spauskite [Išsaugoti].

Prieš pradedant formuoti saugyklos dimensijas ir faktą, turite nuskaityti duomenų šaltinio metaduomenis. Tai padarysite pasirinkę meniu punktą *Projektas -> OLTP nuskaitymas*. Nuskaitytus duomenis sistema pateikia pagrindiniame sistemos lange.

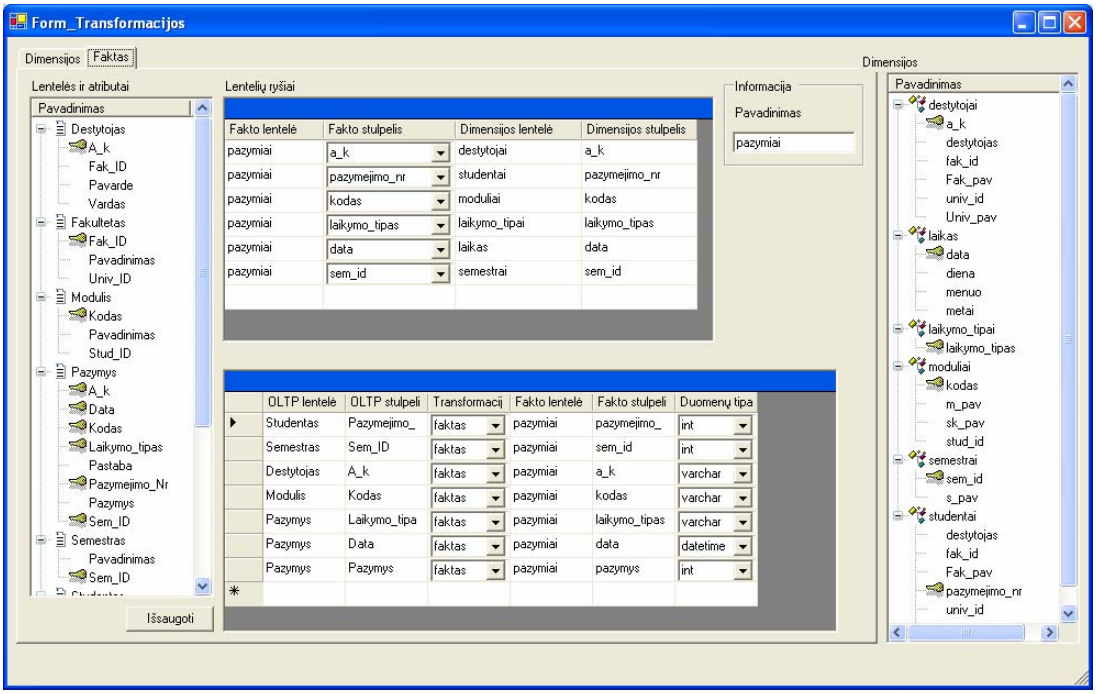
Norėdami formuoti dimensijas arba faktą, iš pagrindinio meniu atverkite pagrindinį sistemos langą *Projektavimas -> Dimensijos*.... (50 pav.)



50 pav. Dimensijų kūrimo langas

Atsivėrusiame lange (50 pav.) sukurkite naują dimensiją, spausdami [Nauja dimen]. Nurodykite dimensijos (duomenų saugyklos dimensijos lentelės) pavadinimą ir pasirinkite dimensijos tipą. Lango kairiame krašte esančiame medyje pasirinkite visus OLTP atributus, kurie sudarys dimensiją: ant pasirinkto atributo spauskite dešinį pelės klavišą ir *Įterpti dimensiją*. Įtraukę visus OLTP atributus, nurodykite informaciją, reikalingą duomenų saugykloje: pavadinimą, duomenų tipą, pirminio rakto požymį, jei atributas bus dimensijos pirminiu raktu. Išsaugokite sukurtą dimensiją, spausdami [Išsaugoti]. Analogiškai sukurkite visas kitas dimensijas.

Sukurti faktą, atverkite kortelę *Faktas* (51 pav.).



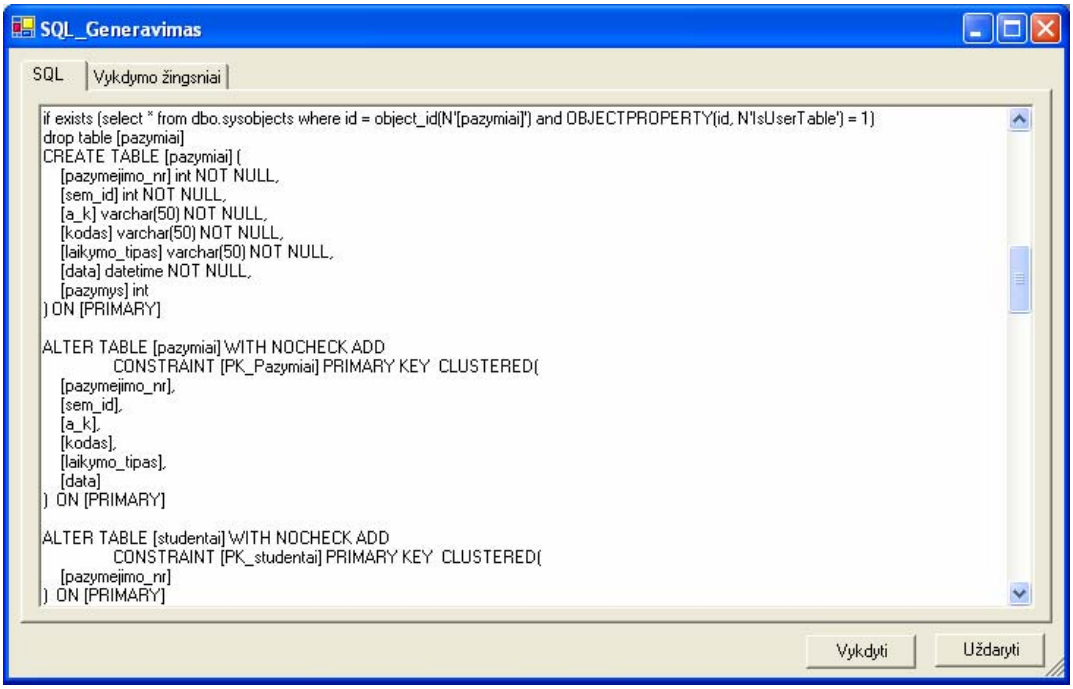
51 pav. Fakto kūrimo langas

Atsivėrusiame lange (51 pav.) nurodykite fakto (duomenų saugyklos dimensijos lentelės) pavadinimą. Lango kairiame krašte esančiame medyje pasirinkite visus OLTP atributus, kurie sudarys faktą: ant pasirinkto atributo spauskite dešinį pelės klavišą ir *Įterpti faktą*. Įtraukę visus OLTP atributus, nurodykite informaciją, reikalingą duomenų saugykloje: pavadinimą, duomenų tipą. Išsaugokite sukurtą faktą, spausdami [Išsaugoti].

Lentelėje *Lentelių ryšiai* pateikiamas visų dimensijų ir dimensijų pirminių raktų sąrašas. Sukurkite fakto ryšius su dimensijomis, t.y. lentelėje *Lentelių ryšiai* parinkite atitinkamą fakto atributą.

Pastaba: Ryšius su faktu privalo turėti visos dimensijos.

Sukūrę visas dimensijas, faktą bei ryšius tarp dimensijų ir fakto, turite sukurti duomenų saugyklą: *Projektavimas -> Generuoti SQL* (52 pav.).

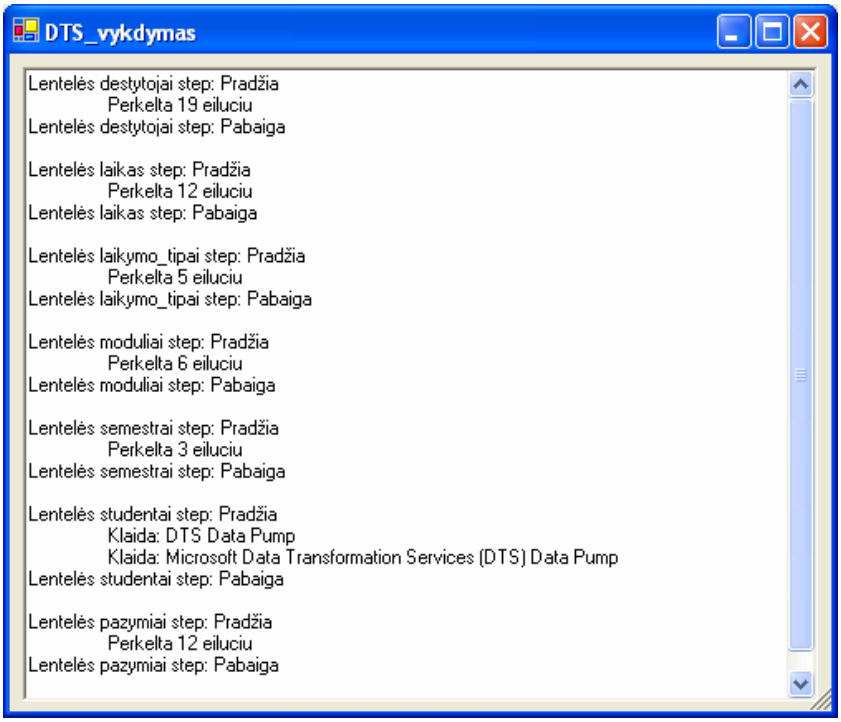


52 pav. SQL kodo generavimo ir saugyklos kūrimo langas

Langas (52 pav.) atveriamas su sugeneruotu duomenų saugyklos kūrimo SQL kodu, kurį, jeigu norite galite, nusikopijuoti. Sukurti duomenų saugyklą, spauskite mygtuką [Vykdyti]. Sistema MS SQL serveryje sukuria duomenų saugyklą. Saugyklos kūrimo vykdymo žingsnius galite peržiūrėti kortelėje *Vykdymo žingsniai*.

Sukūrę duomenų saugyklą, perkelkite duomenis iš duomenų šaltinio į duomenų saugyklą, paspausdami *Projektavimas -> Perkelti duomenis*.

Duomenys perkeliami į duomenų saugyklą. Duomenų perkėlimo eigą, rezultatus ir, jeigu buvo, klaidas galite peržiūrėti formoje (53 pav.).



53 pav. Duomenų iš duomenų šaltinio į duomenų saugyklą perkėlimo eiga ir rezultatai

6.2.2. Egzistuojančio duomenų saugyklos projekto koregavimas

Jeigu norite peržiūrėti ar pakoreguoti egzistuojančio projekto duomenis, atsidarykite egzistuojantį projektą *Projektas -> Atidaryti...*. (54 pav.)

54 pav. Egzistuojančio projekto atidarymo langas

Atvertame lange pažymėkite norimą projektą ir spauskite [Atidaryti]. Atveriant projektą, į pagrindinį sistemos langą užkraunama visa projekto informacija: dimensijos, faktas (50, 51 pav.).

Jeigu norite koreguoti dimensiją, pasirinkimo medyje *Dimensijos* ant pasirinktos transformacijos (dimensijos) spauskite dešinį pelės klavišą ir *Koreguoti*. Lentelėje *Transformacijos* galite koreguoti dimensijos transformaciją, įtraukti arba išmesti atributus.

Jeigu norite pašalinti dimensiją iš saugyklos, pasirinkimo medyje *Dimensijos* ant pasirinktos transformacijos (dimensijos) spauskite dešinį pelės klavišą ir *Šalinti*.

Pakoregavę duomenų saugyklos duomenis, turite iš naujo sukurti duomenų saugyklą ir perkelti duomenis *Projektavimas -> Generuoti SQL* (52 pav.).

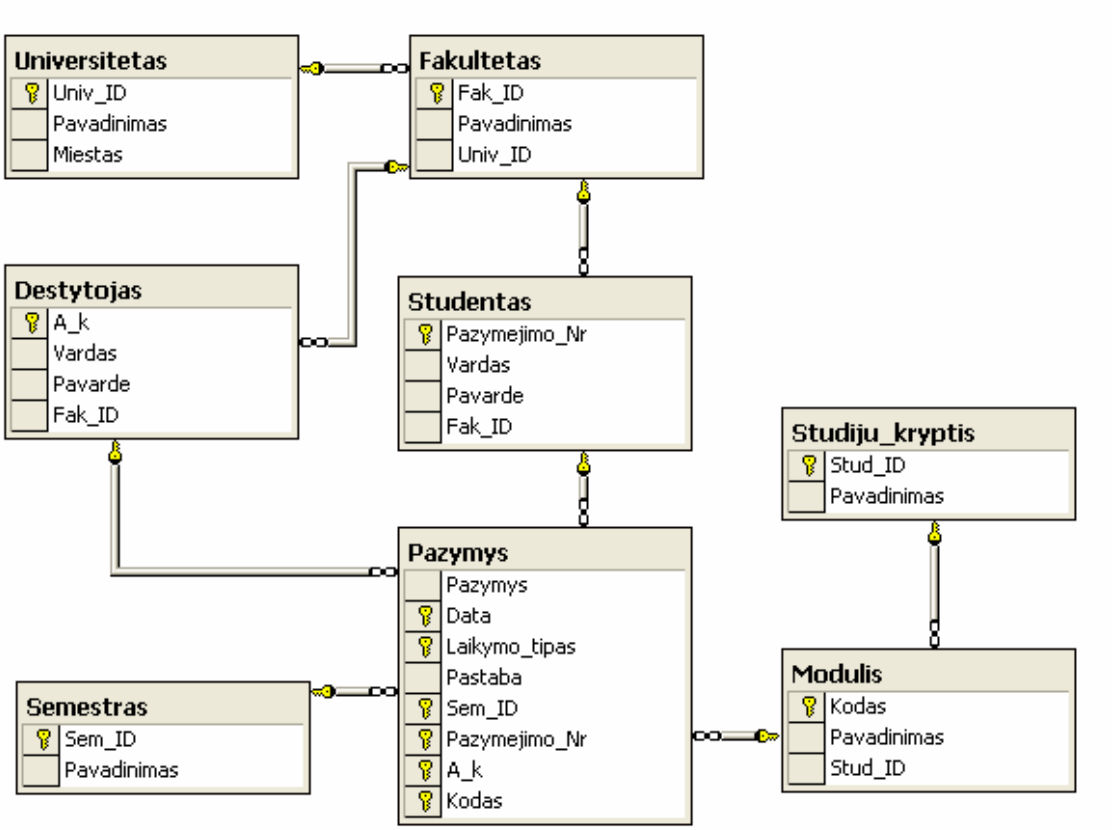
Langas (52 pav.) atveriamas su sugeneruotu duomenų saugyklos kūrimo SQL kodu, kurį, jeigu norite galite, nusikopijuoti. Sukurti duomenų saugyklą, spauskite mygtuką [Vykdyti]. Prieš kurdama duomenų saugyklą, sistema išmeta visus MS SQL serveryje buvusios ankstesnės saugyklos duomenis ir saugyklą sukuria iš naujo. Saugyklos kūrimo vykdymo žingsnius galite peržiūrėti kortelėje *Vykdymo žingsniai*.

Sukūrę duomenų saugyklą, perkelkite duomenis iš duomenų šaltinio į duomenų saugyklą, paspausdami *Projektavimas -> Perkelti duomenis*.

Duomenys perkeliami į duomenų saugyklą. Duomenų perkėlimo eigą, rezultatus ir, jeigu buvo, klaidas galite peržiūrėti formoje (53 pav.).

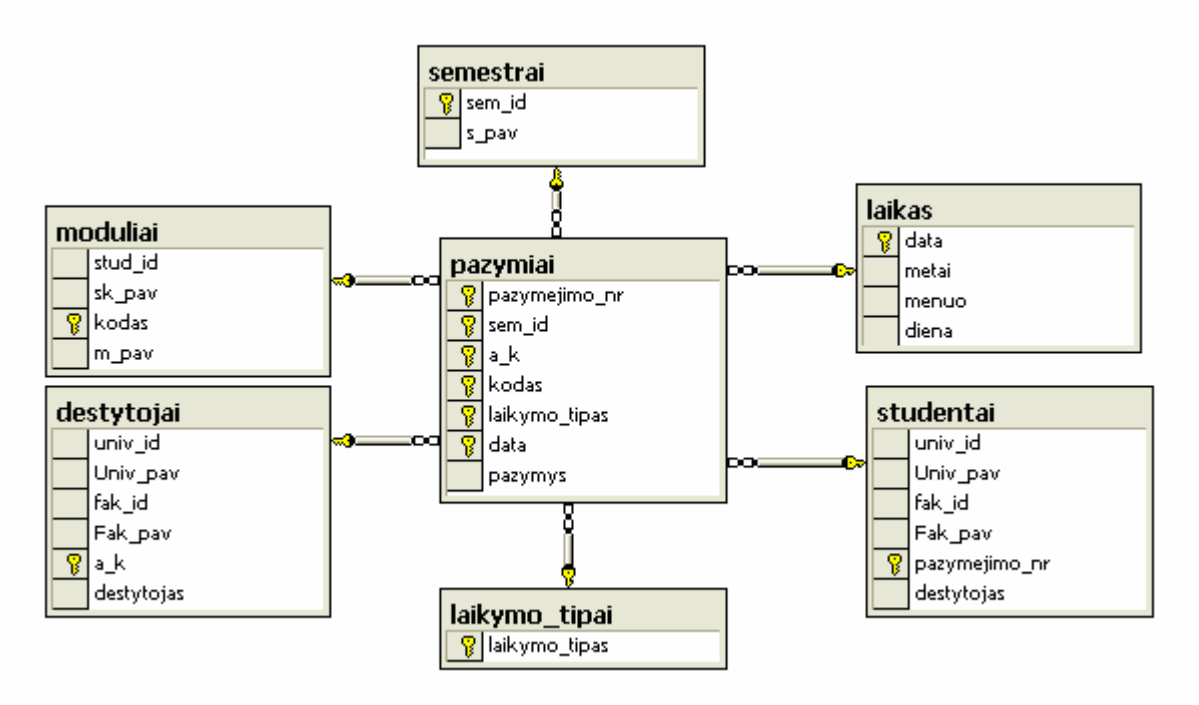
7. Duomenų saugyklos kūrimas, panaudojant sukurtą ETL priemonę

Duomenų saugyklos kūrimui pasirinkta ta pati duomenų bazė, kuri panaudota kuriant duomenų saugyklą esamomis MS SQL priemonėmis (žr. skyriuje *Duomenų saugyklos kūrimas, panaudojant egzistuojančias MS SQL priemones* (*3*)). Reliacinė duomenų bazėje kaupiami duomenys apie visų Lietuvos universitetų dėstytojus, studentus, dėstomus dalykus bei studentų įvertinimus (55 pav.). Duomenų bazė saugoma MS SQL Server.



55 pav. Reliacinė studentų duomenų bazė

Kuriama duomenų saugykla, skirta studentų pažangumo analizei. Duomenų saugyklos schema pasirinkta žvaigždės schema (56 pav.). Duomenų saugyklos duomenų bazė irgi saugoma MS SQL Server.



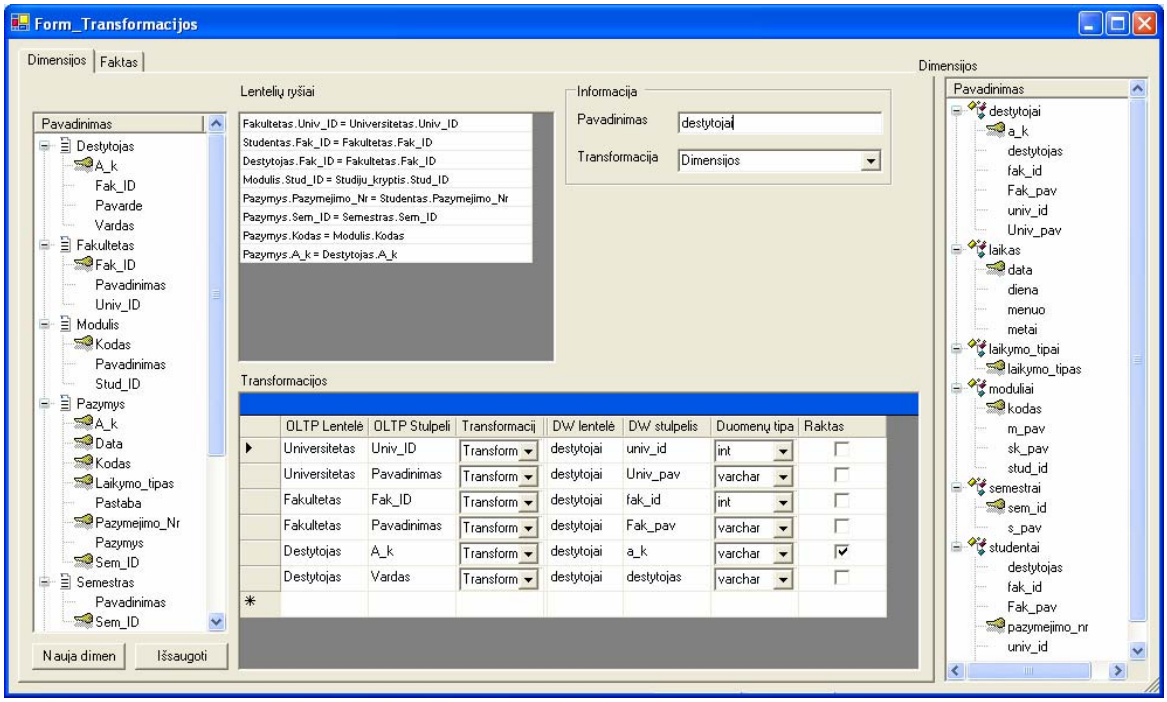
56 pav. Studentų pažangumo analizės „žvaigždės“ schemos modelis

Norėdami sukurti duomenų saugyklą, turime atlikti tokias duomenų transformacijas:

- Iš lentelės „Pažymys“išskirti laiko ir laikymo tipo dimensijas.
- Iš lentelės „Pažymys“ išmesti nereikalingus atributą – „Pastaba”.
- Lenteles „Universitetas“, „Fakultetas“, „Dėstytojas“ sujungti į Dėstytojo dimensiją.
- Lenteles „Universitetas“, „Fakultetas“, „Studentas“ sujungti į Studento dimensiją.
- Lenteles „Studijų kryptis“, „Modulis“ sujungti į Modulio dimensiją.
- Sukurti faktą.

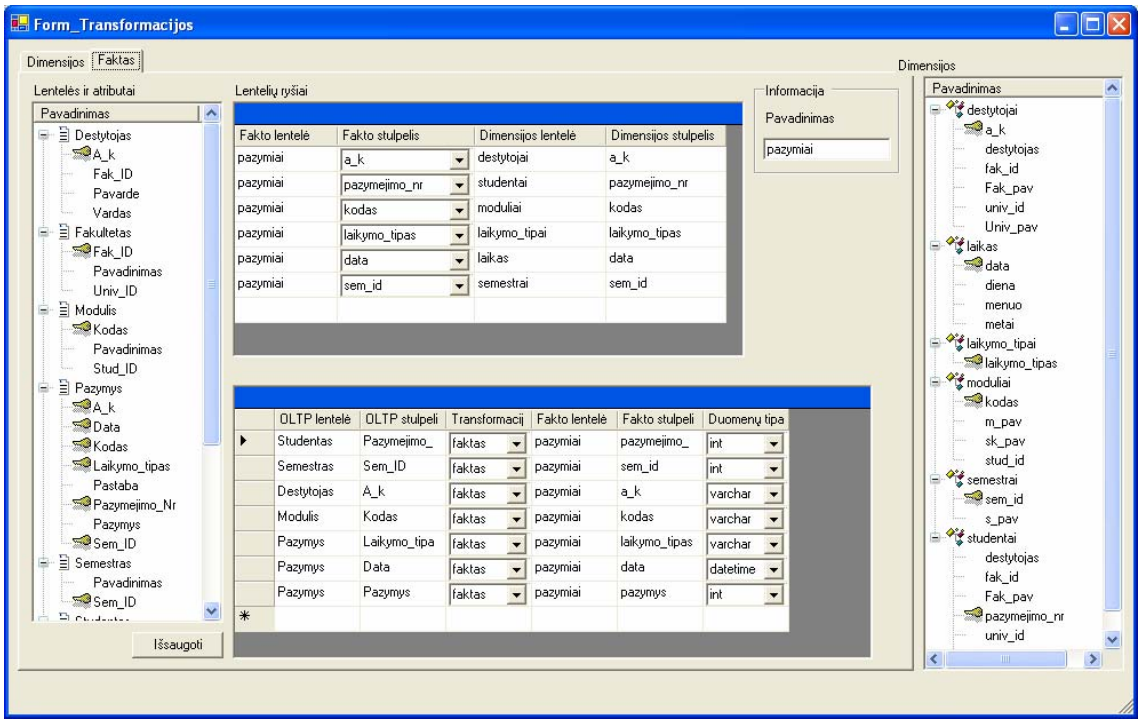
7.1.*Dimensijų ir fakto kūrimo duomenų transformacijos*

Sukurtos visos reikalingos dimensijos pateikiamos 57 paveiksle.



57 pav. Duomenų saugyklos dimensijos

Sukurtas duomenų saugyklos faktas ir ryšiai tarp dimensijų ir fakto pateikiami 58 paveiksle.



58 pav. Duomenų saugyklos faktas

Sukūrę duomenų saugyklos dimensijas ir faktą, galite sukurti duomenų saugyklą ir į ją perkelti duomenis iš duomenų šaltinio.

Duomenų saugyklos analizės galimybės analogiškos 4 skyriuje aptartoms analizės galimybėms.

7.2.*Duomenų saugyklos kūrimo rezultatai ir išvados*

Atliekant duomenų saugyklos kūrimą, esamomis MS SQL priemonėmis, visas transformacijas reikėjo rašyti programiniu kodu. Tuo tarpu sukurtoje ETL priemonėje, transformacijų tipus ir reikalingus duomenų šaltinio metaduomenis galite pasirinkti naudodamiesi patogia grafine vartotojo sąsaja.

8. Transformacijų šablonais grindžiamos ETL priemonės realizacijos įvertinimas

8.1. *Transformacijų šablonais grindžiamos ETL priemonės palyginimas su MS SQL*

Analizės dalyje buvo išanalizuotos MS SQL duomenų saugyklos kūrimo galimybės. Realizuodami ETL priemonę, skirtą MS SQL, sukurta priemonė žymiai pagreitinanti duomenų saugyklos kūrimo procesą (7 lentelė).

7 lentelė

Sukurtos ETL priemonės palyginimas su MS SQL

	Sukurta ETL sistema	**MS SQL**
Duomenų transformavimo operacijos	Realizuotos pagrindinės duomenų transformavimo operacijos, reikalingos duomenų saugyklos kūrimui.	Kiekvienąkart kuriant saugyklą, duomenų transformacijų operacijas reikia aprašyti programiniu kodu.
Vartotojo sąsaja (*angl. GUI*), transformavimo operacijoms panaudoti	Turi.	Neturi.

Analizės metu buvo du svarbiausi ETL priemonių vertinimo kriterijai: ar realizuotos transformavimo operacijos ir, ar priemonė turi vartotojo sąsają. MS SQL neturi nei vieno nei kito. Sukurtoji ETL priemonė turi pagrindinės operacijas, reikalingas duomenų saugyklos kūrimui ir vartotojo sąsają joms panaudoti. Todėl sukurtoji ETL priemonė smarkiai lenkia MS SQL ir priartėja prie atskirai platinamų ETL priemonių tokių kaip Power Center (Informatica), Data Integrator (Business Objects) ir pan., galimybių.

8.2. *Duomenų saugyklos kūrimo laiko įvertinimas*

Šitas palyginimas turi daug subjektyvių pusių (projektuotojo SQL žinios ir darbo greitis), tačiau bendrą tendenciją galima nustatyti. Lentelėje (8 lentelė) pateiktas duomenų transformacijos sudėtingumas – tai atributų, sudarančių dimensiją skaičius. Duomenų saugyklos kūrimo laikas pateiktas minutėmis.

8 lentelė

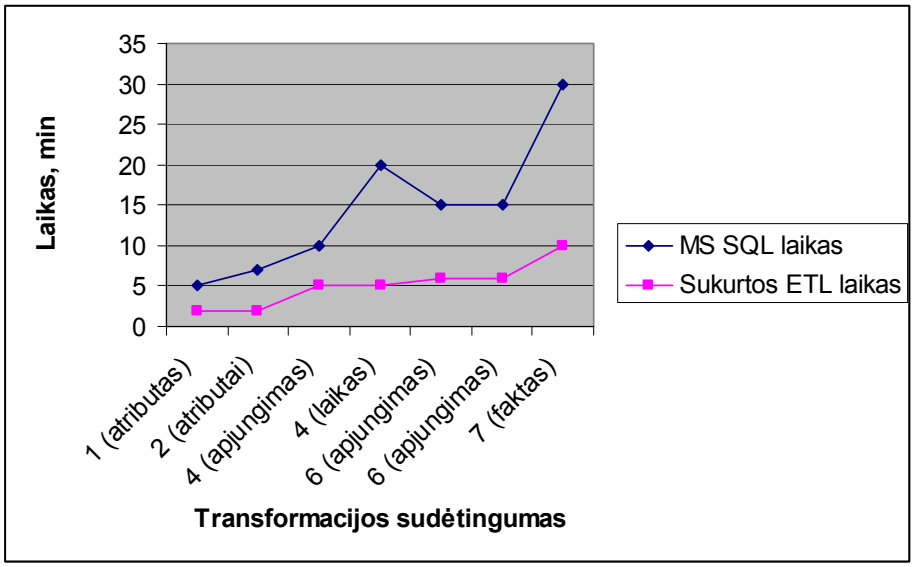
Sukurtos ETL priemonės palyginimas su kitomis ETL priemonėmis

Transformacija	**Sudėtingumas**	**MS SQL laikas**	**Sukurtos ETL laikas**
Dėstytojai – kelių lentelių atributų jungimas	6	15	6
Studentai – kelių lentelių atributų jungimas.	6	15	6
Moduliai – kelių lentelių atributų jungimas.	4	10	5
Laikas – laiko atributų išskyrimas.	4	20	5
Laikymo tipas – vienas atributas.	1	5	2
Semestrai – du tos pačios lentelės atributai	2	7	2
Pažymiai (faktas)	7	30	10

* lentelėje pateikiamas laikas, kai projektuotojas išmano SQL vidutiniškai ir iš anksto yra numatęs, kokias dimensijas kurs.

Kuriant duomenų transformavimo operacijas prie vienos operacijos sugaištas priklauso nuo operacijos sudėtingumo (pvz. laiko transformacijos kūrimas). Tuo tarpu dirbant su sukurta ETL priemone laikas kuriant dimensijas, priklauso tik nuo dimensiją sudarančių eilučių skaičiaus. Kuriant faktą, dar reikia įvertinti ryšių su dimensijomis kūrimą.

Matome (59 paveikslas), kad jei transformavimo operacijos yra nedidelės ir paprastos (laikymo tipas ir semestrai), ETL priemonės laikas nedaug skiriasi nuo MS SQL priemonių, tačiau kuriant transformacijas, turinčias daugiau eilučių, arba tokias, kurių realizacija programiniu kodu yra sudėtingesnė, sukurtos ETL priemonės laikas yra daugiau negu dvigubai trumpesnis.



59 pav. Darbo MS SQL ir sukurta ETL palyginimas

8.3. *Transformacijų šablonais grindžiamos ETL priemonės perspektyvos ir plėtimo galimybės*

Sukurta ETL priemonė paspartino duomenų saugyklos kūrimo procesą. Tačiau kol kas ši priemonė yra palyginti ribota ir norint aprėpti visapusišką duomenų saugyklos kūrimo procesą galima sukurtą priemonę patobulinti:

1. Naudojant ETL priemonę, galima sukurti tik žvaigždės schemos saugyklą. Tai ne vienintelis saugyklos schemos tipas. Dar galimi tokie schemų tipai: snaigė, plokštė, terasa žvaigždžių spiečius. Realizavus visas schemas būtų galima kurti įvairias duomenų saugyklas.
2. Papildyti fakto transformaciją, taip, kad būtų galima nurodyti agregavimo funkcijas.
3. Šiuo metu duomenų šaltiniu gali būti tik MS SQL duomenų bazė. Nėra sudėtinga įvesti ir kitas duomenų bazes. Tam tikslui reikia parašyti kitas šaltinio duomenų bazės metaduomenų nuskaitymo procedūras, kurios priklauso nuo duomenų bazės.
4. Šiuo metu duomenų šaltiniu gali būti tik duomenų bazės lentelės, bet lygiai taip pat duomenų šaltiniu gali būti – duomenų bazės lentelės vaizdas (*angl. view*).
5. ETL priemonėje įvesti galimybę konstruoti formules: pvz., dviejų laukų sandauga, dviejų atributų apjungimas į vieną.
6. Duomenų saugykla kuriama tik MS SQL Server. Nesunku padaryti, kad saugyklą būtų kuriama kitoje duomenų bazėje – skirtingose duomenų bazių valdymo sistemose kuriasi duomenų tipai.
7. Laiko transformacijai automatizuoti būtų galima suformuoti vedlį (*angl. wizard*), kuris leistų pasirinkti kokio tipo laiko dimensiją norite formuoti (pvz., metai – mėnuo – diena arba metai – ketvirtis – mėnuo – diena ir t.t.). Taip pat turėtų atsirasti galimybė suformuoti laiko dimensiją, jei duomenų šaltinyje nenaudojamas joks laiko elementas.

9. Išvados

1. Sprendžiant ilgo duomenų saugyklų kūrimo proceso sutrumpinimo problemą, buvo atlikta literatūroje aprašyto saugyklų kūrimo proceso bei praktikoje naudojamų įrankių analizė.
2. Įrankių analizės metu nustatyta, kad geresnėmis savybėmis pasižymi specialūs ETL įrankiai, kurie turi iš anksto sudarytas schemų ir duomenų transformavimo operacijas. Universalios DBVS turi tik universalias transformavimo priemones, kurios reikalauja daug kūrėjo pastangų.
3. Literatūros analizės metu nustatyta, kad didžiąją duomenų saugyklos kūrimo proceso dalį sudaro duomenų transformavimo operacijų kūrimas, tačiau šios operacijos yra tipinės ir jų kūrimą galima automatizuoti.
4. Buvo nuspręsta patobulinti duomenų saugyklos kūrimo procesą, sukuriant tipinių transformacijų šablonus ir išbandyti juos MS SQL priemonei.
5. Pradžioje buvo atliktas duomenų saugyklos kūrimas esamomis MS SQL priemonėmis. Saugyklos kūrimas padėjo įsigilinti į patį kūrimo procesą ir sukurti pagrindinių duomenų transformavimo operacijų tipų šablonus.
6. Remiantis išskirtu duomenų saugyklos kūrimo procesu buvo suprojektuota ETL sistema, sukurtas transformacijų metamodelis ir suprojektuota MS SQL duomenų bazė, skirta saugoti duomenų šaltinių, duomenų saugyklų ir transformacijų tipų metaduomenis.
7. Realizavus ETL priemonę ir atlikus duomenų saugyklos kūrimo tyrimą, nustatyta, kad laikas reikalingas saugyklos kūrimui sutrumpėjo beveik per pusę.
8. Sukurta ETL priemonė savyje turi tik pagrindines duomenų saugyklos kūrimui reikalingas transformavimo operacijas: dimensijų ir fakto kūrimą, ir skirta žvaigždės tipo schemų kūrimui, tačiau panašiu principu sistemą galima praplėsti ir kitomis reikiamomis transformavimo operacijomis bei schemomis.
9. Pagrindinis darbo rezultatas ir jo naujumas yra tipinių saugyklų schemų ir duomenų transformacijų metamodelis ir eksperimentinis saugyklos kūrimo pagreitinimo galimybių tyrimas, kuris patvirtina, kad tokiu būdu galima patobulinti esamus saugyklų kūrimo procesus.
10. Darbo analizė ir numatomi rezultatai buvo pristatyti 10 – ojoje tarpuniversitetinėje magistrantų ir doktorantų konferencijoje „Informacinės technologijos“, straipsnis išspausdintas konferencijos leidinyje.

10. Literatūra

1. Inmon W.H. Building the Data Warehouse. John Wiley & Sons, 2002.
2. Kimball R. The Data Webhouse Toolkit: Building the Web-Enabled Data Warehouse, John Wiley & Sons, 2000.
3. Kimball R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley & Sons, 2002.
4. Silverston L. A library of universal data models for all enterprises. John Wiley & Sons, 2001.
5. Moody D., Kortink M.. From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design // CEUR Workshop Proceeeding, 2000. Žiūrėta [2004 – 10 – 27]. Prieiga per Internetą: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/paper5.pdf.
6. Common Warehouse Metamodel (CWM) Specification. OMG document Version 1.1 Formal/03-03-02, 2003.
7. DTS Overwiev. Žiūrėta [2004 – 11– 02]. Prieiga per Internetą: http://msdn.microsoft.com.
8. ETL Tools. Žiūrėta [2004 – 11– 15]. Prieiga per Internetą: http://www.sas.com/news/analysts/meta_etl_0404.pdf
9. Ascential Software Corporation. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.ascential.com
10. Informatika Corporation. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.informatica.com
11. SAS. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.sas.com
12. Pervasive Software. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.pervasive.com
13. Business Objects. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.businessobjects.com
14. Sunopsis. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.sunopsis.com
15. DataMirror. Žiūrėta [2004 – 12 – 27]. Prieiga per Internetą: www.datamirror.com
16. Oracle Corporation. Žiūrėta [2005 – 01 – 12]. Prieiga per Internetą: www.oracle.com
17. Paulavičiūtė K., Nemuraitė L. Duomenų transformacijos duomenų saugyklos kūrime // Informacinės technologijos (10 – oji tarpuniversitetinė magistrantų ir doktorantų konferencija), 2005. p. 138 – 141.

11. Terminai

DB – duomenų bazė

DBVS – duomenų bazių valdymo sistema.

DTS – duomenų transformacijos servisai (*angl. Data Transformation Services*)

ETL – išgauk/ transformuok/ įdėk (*angl. extract/ transform/ load*)

OLTP - On-line Transaction Processing

OLAP - On-Line Analytical Processing

MS SQL – Microsoft SQL

GUI – grafinė vartotojo sąsaja (*angl. grafic user interface*)

Priedai

1. ETL sistemos duomenų bazės aprašas

Duomenu_tipai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
tipo_id	LONG	☐	☑	Duomenų tipo identifikacijos numeris.
pavadinimas	TEXT(50)	☐	☐	Duomenų tipo pavadinimas.

Duomenu_transformacijos

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
projekto_id	LONG	☐	☐	Projekto identifikacijos numeris.
modelio_id_R	LONG	☐	☐	Reliacinio dumenų modelio identifikacijos numeris.
objekto_id_lentele_R	LONG	☐	☐	Reliacinės duomenų bazės lentelės identifikacijos numeris.
stulpelio_id_R	LONG	☐	☐	Reliacinės duomenų bazės atributo identifikacijos numeris.
modelio_id_W	LONG	☐	☐	Duomenų saugyklos modelio identifikacijos numeris.
objekto_id_lentele_W	LONG	☐	☐	Duomenų saugyklos lentelės identifikacijos numeris.
stulpelio_id_W	LONG	☐	☐	Duomenų saugyklos atributo identifikacijos numeris.
Transformacijos_id	LONG	☐	☐	Duomenų transformacijos identifikacijos numeris.

FK_stulpeliai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
projekto_id	LONG	☐	☐	Projekto identifikacijos numeris.
modelio_id	LONG	☐	☐	Modelio identifikacijos numeris.
objekto_id_FK	LONG	☐	☐	Išorinio rakto identifikacijos numeris.
objekto_id_lentele1	LONG	☐	☐	Išorinio rakto tėvinės lentelės identifikacijos numeris.
stulpelis_id1	LONG	☐	☐	Išorinio rakto tėvinės lentelės atributo identifikacijos numeris.
objekto_id_lentele2	LONG	☐	☐	Išorinio rakto vaikinės lentelės identifikacijos numeris.
stulpelis_id2	LONG	☐	☐	Išorinio rakto vaikinės lentelės atributo identifikacijos numeris.

Modeliai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
modelio_id	LONG	☐	☑	Modelio identifikacijos numeris.
pavadinimas	TEXT(50)	☐	☐	Modelio pavadinimas.

Objektai

Stulpeliai	Duomenų	Tuščia	Pirminis	Paaiškinimas

	tipas	reikšmė	raktas	
projekto_id	LONG	☐	☑	Projekto identifikacijos numeris.
modelio_id	LONG	☐	☑	Modelio identifikacijos numeris.
objekto_id	LONG	☐	☑	Objekto (lentelės, pirminio arba išorinio rakto) identifikacijos numeris.
pavadinimas	TEXT(127)	☐	☐	Objekto pavadinimas.
parent_id	LONG	☑	☐	"Tėvinio" objekto, kuriam priklauso objektas, identifikacijos numeris. Šis numeris nurodomas pirminiams arba išoriniams raktas.
tipas	TEXT(50)	☐	☐	Objekto tipas: U (lentelė), PK (pirminis raktas), F (išorinis raktas).

PK_stulpeliai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
projekto_id	LONG	☐	☑	Projekto identifikacijos numeris.
modelio_id	LONG	☐	☑	Modelio identifikacijos numeris.
objekto_id_PK	LONG	☐	☑	Pirminio rakto identifikacijos numeris.
objekto_id_lentele	LONG	☐	☑	Lentelės, kuriam priklauso pirminis raktas, identifikacijos numeris.
stulpelio_id	LONG	☐	☑	Pirminį raktą sudarančio atributo identifikacijos numeris.

Prisijungimai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
modelio_id	LONG	☐	☑	Modelio identifikacijos numeris.
projekto_id	LONG	☐	☑	Projekto identifikacijos numeris.
serveris	TEXT(255)	☐	☐	MS SQL serverio pavadinimas.
win_auth	BIT	☑	☐	Požymis, ar prisijungimui prie MS SQL serverio naudojama Windows autentifikacija.
vartotojas	TEXT(255)	☑	☐	Vartotojo prisijungimo vardas.
slaptazodis	TEXT(255)	☑	☐	Slaptažodis.
DB	TEXT(255)	☐	☐	Duomenų bazės pavadinimas.

Projektai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
projekto_id	LONG	☐	☑	Projekto identifikacijos numeris.
pavadinimas	TEXT(50)	☐	☐	Projekto pavadinimas.

Stulpeliai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
projekto_id	LONG	☐	☑	Projekto identifikacijos numeris.
modelio_id	LONG	☐	☑	Modelio identifikacijos numeris.
objekto_id_lentele	LONG	☐	☑	Lentelės, kuriam priklauso stulpelis, identifikacijos numeris.
stulpelio_id	LONG	☐	☑	Stulpelio identifikacijos numeris.
pavadinimas	TEXT(127)	☐	☐	Stulpelio pavadinimas.

duomenu_tipo_id	LONG	☐	☐	Stulpelio duomenų tipo identifikacijos numeris.

Transformacijos

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
transformacijos_id	LONG	☐	☑	Transformacijos identifikacijos numeris.
transformacija	TEXT(255)	☐	☐	Transformacija, užrašyta SQl kalba.
pavadinimas	TEXT(127)	☐	☐	Transformacijos pavadinimas.
transformacijos_tipo_id	LONG	☐	☐	Transformacijos tipo identifikacijos numeris.

Transformacijos_tipai

Stulpeliai	Duomenų tipas	Tuščia reikšmė	Pirminis raktas	Paaiškinimas
Transformacijos_tipo_id	LONG	☐	☑	Transformacijos tipo identifikacijos numeris.
transformacija	TEXT(255)	☐	☐	Transformacijos pavadinimas.
Fakto_pozymis	BIT	☐	☐	Požymis, ar transformacija yra fakto transformacija.

2. K. Paulavičiūtė, L. Nemuraitė „Duomenų transformacijos duomenų saugyklos kūrime“