

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
INFORMACIJOS SISTEMŲ KATEDRA

Vygintas Jakimavičius

**Vaistinių tinklo duomenų saugyklos modelis ir
taikymas**

Magistro darbas

Darbo vadovas
prof. dr. Lina Nemuraitė

Kaunas, 2009

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
INFORMACIJOS SISTEMŲ KATEDRA

Vygintas Jakimavičius

**Vaistinių tinklo duomenų saugyklos modelis ir
taikymas**

Magistro darbas

Recenzentas

2009-01-

dr. Audronė Janavičiūtė

Vadovas

prof. dr. Lina Nemuraitė
2009-01-

Atliko

2009-01-

IFM-3/4 gr. stud.
Vygintas Jakimavičius

Kaunas, 2009

Turinys

1. Įvadas	1
Vaistinių tinklo duomenų saugyklos analizė.....	5
1.1. Analizės tikslas.....	5
1.2. Tyrimo sritis, objektas ir problema	5
1.2.1. Situacija šiandieninėje rinkoje	7
1.2.2. Verslo analitika	8
1.2.2.1. Kuo gali padėti verslo analitika?.....	8
1.2.2.2. Verslo analitikos privalumai	8
1.2.2.3. Pagrindinės kompanijų problemos	8
1.2.2.4. OLAP duomenų bazės.....	10
1.2.2.4.1. Dimensijos ir skaičiuojamosios reikšmės	10
1.3. Vartotojų analizė	13
1.3.1. Vartotojų aibė, tipai ir savybės.....	13
1.3.2. Vartotojų tikslai ir problemos	13
1.4. Saugyklų kūrimo metodikos analizė	14
1.4.1. Skirtingi OLAP duomenų lygių pavadinimai duomenų apdorojime	17
1.4.2. OLAP duomenų šaltiniai.....	18
1.4.3. Bendrasis duomenų saugyklos proceso modelis	19
1.5. Oracle ir MS SQL duomenų analizės kūrimo priemonių lyginamoji analizė.....	21
1.6. Architektūros ir galimų įgyvendinimo priemonių variantų analizė	23
1.7. Darbo tikslas ir siejami privalumai	26
1.8. Reikalavimai duomenims.....	26
1.9. Nefunkciniai reikalavimai ir apribojimai	27
1.9.1. Reikalavimai veikimui	27
1.9.2. Reikalavimai sąveikai/suderinamumui su kitomis sistemomis.....	28
1.9.3. Kiti reikalavimai (patikimumui, saugumui, plėtimui).....	29
1.10. Rizikos faktorių analizė.....	29
1.11. Rezultato kokybės kriterijai	30
1.12. Analizės kubų projektavimo metodikų analizė	31
1.12.1. Verslo poreikių surinkimas ir analizė.....	31
1.12.2. Duomenų modeliavimas.....	31
1.12.3. Duomenų šaltiniai	32
1.12.4. ETL projektavimas.....	32
1.12.5. Pasikeitusių duomenų nustatymo (CDC) metodas.....	33
1.12.6. Duomenų saugyklų modeliai.....	35
1.13. Analizės kubų veikimo ir atnaujinimo optimizavimas.....	38
1.14. Analizės išvados.....	42
2. Vaistinių tinklo duomenų analizės sistemos reikalavimų specifikacija	43
2.1. Sistemos panaudojimo atvejų diagramos ir specifikacijos.....	43
2.2. Analizės kubo specifikacija.....	50
2.3. Detali reikalavimų specifikacija.....	51
2.4. Dalykinės srities modelis	52
3. Klientų analizės sistemos projektas	54
3.1. Sistemos pagrindimas ir esmės išdėstymas.....	54
3.2. Sistemos architektūra	55
3.2.1. Bendra sistemos architektūra	56
3.2.2. Reikalavimai sistemos funkcionavimo palaikymui.....	57
3.2.3. Duomenų transformacijų servisai.....	59
3.2.4. Duomenų atnaujinimo automatizavimas.....	65

3.2.5.	Analizės kubo formavimo procesas	67
3.2.6.	Analizės modelių metaduomenų eksportas naudojant XML standartą	68
3.2.7.	Analizės kubų formavimo procesas	69
3.3.	Duomenų bazės schema	69
4.	Eksperimentinis sistemos tyrimas ir vertinimas	71
4.1.	Eksperimentinio diegimo aprašymas	71
4.2.	Sistemos diegimas	71
4.3.	Sistemos naudojimo instrukcija	73
4.4.	Sistemos veikimo ir savybių analizė, kokybės kriterijų įvertinimas	83
4.5.	Sistemos taikymo rekomendacijos	85
4.6.	Tolimesnio sistemos plėtojimo, tobulinimo galimybės	85
5.	Išvados	87
6.	Literatūra	88
7.	Terminų ir santrumpų žodynelis	90
	Pharmacy Data Warehouse Model and Practice	91
	Summary	91
8.	Priedai	92
	1 priedas. Techninė užduotis	92
	2 priedas. Pažyma iš įmonės apie sistemos įdiegimą	93
	3 priedas. Realizuotos užklaustos, funkcijos	94

1. Įvadas

Šiandieninių įmonių veikla jau nebeįsivaizduojama be kompiuterizuotos apskaitos. Jau bene kiekvienoje, net ir mažiausioje įmonėje rasime bent vieną kompiuterį, kuriame vedama įmonės buhalterinė apskaita, finansai, rengiamos ataskaitos. Tačiau rinka diktuoja savas sąlygas ir tempą. Verslo pasaulyje yra labai svarbu kuo greičiau sugebėti priimti reikalingus sprendimus, gauti norimą informaciją. Yra svarbu apie savo verslą gauti visą informaciją – sekti klientų įpročius, jų poreikius, analizuoti pardavimų informaciją.

Kai duomenų yra nedaug, jų analizę galima atlikti pasitelkus ir tradicines reliacines duomenų bazių valdymo sistemas (RDBVS), kurios kartu naudojamos ir transakcijoms realiuoju laiku vykdyti (Online Transaction Processing, OLTP). OLTP duomenų bazės pritaikytos tam, kad galėtų palaikyti daug vienu metu į jas besikreipiančių vartotojų, kurie intensyviai įterpia bei keičia duomenis. Paprastai tokios duomenų bazės turi sudėtingą struktūrą (didelė normalizacija), o duomenys jose parodo esamą įmonės situaciją, tačiau ne istoriją. Jose esančių duomenų analizė dažniausiai remiasi įprastų užklausų, parašytų SQL kalba, vykdymu.

Tačiau kai duomenų susikaupia daug, ir organizacija mėgina atlikti strateginę analizę naudodama tą pačią OLTP duomenų bazę, kaip ir einamojo duomenų apdorojimo atveju, dažnai susiduriama su keliomis problemomis:

- didelių skaičiavimo pajėgumų reikalaujančios intensyviai vykdomos analitinės užklausos neigiamai veikia darbinės sistemos produktyvumą;
- jeigu darbinėje duomenų bazėje saugoma istorinė informacija, nereikalinga kasdienėms operacijoms atlikti, tokios sistemos produktyvumas mažėja, nes duomenų laikui bėgant vis daugėja.

Siekiant išvengti minėtų problemų ir išsaugoti istorinius duomenis, kuriamos atskiros duomenų saugyklos (Data Warehouses), kuriose telpa jau istorine tapusi informacija iš darbinės duomenų bazės. Istorinių duomenų saugojimą ir analizę perkėlus į duomenų saugyklas, išvengiama padidėjusių darbinių duomenų bazių serverių apkrovų.

Pastaraisiais metais didesnės įmonės vis dažniau siūlo savo klientams nuolaidų korteles, kuriomis pasinaudojus, galima gauti įvairias nuolaidas, specialius pasiūlymus. Ne išimtis ir UAB „Nemuno Vaistinė“. Turinti per 200 filialų ir nuolat besiplečianti įmonė, jau nuo 2003 metų klientams pradėjo siūlyti klientų kortelės, su kuriomis klientai gali gauti iki 10% nuolaidas visoms prekėms. Įmonėje įdiegta verslo valdymo sistema buvo modifikuota, jog kiekvieną kartą, kai pardavimo metu yra panaudojama kliento kortelė, visa informacija apie pardavimą yra užregistruojama informacinėje sistemoje. Tačiau iki šiol tai apčiuopiamos naudos nedavė – buvo

kaupiama tik informacija apie kortelių panaudojimo atvejus, bet šie duomenys toliau niekur nebenaudojami. Tačiau tokiais atvejais yra svarbu atsakyti į tokis klausimus, kaip „kurie klientai yra lojalūs?“, „kaip prekių pardavimai yra pasiskirstę pagal klientų amžių bei gyvenamą vietovę?“ ir pan.

Tad natūraliai iškilo poreikis sukurti sistemą, kurios pagalba galima būtų sisteminti, analizuoti nuolaidų kortelių pagalba kaupiamus duomenis. Iki šiol įmonėje nebuvo tokios sistemos, tad buvo nuolat kaupiama informacija, kuria nebuvo galimybės tinkamai pasinaudoti. Tradicinės priemonės analizuoti duomenis, generuojant reikalingas ataskaitas, gana ženkliai apkraudavo centrinę sistemą ir dėl to mažėjo bendras sistemos našumas. Jeigu tokias ataskaitas paleidžia keletas sistemos vartotojų vienu metu, sistemos apkrovimas taip padidėja, jog likusieji vartotojai tuo metu iš viso negali naudotis sistema, arba jų darbo efektyvumas ženkliai sumažėja. Sistemų analitikai, siekdami kuo efektyviau išnaudoti naktinį sistemos darbą, neretai užklauso paleidžia vykdyti naktiniai. Nakties metu vartotojų skaičius yra minimalus, tad reikalingas ataskaitas generuoti palankiau yra būtent nakties metu. Tačiau ir tai nėra išeitis – pakoregavus ataskaitą, ją tenka generuoti iš naujo. Tad standartinių priemonių klientų duomenų analizei nepakanka.

Žinant visus šias standartinių ataskaitų priemonių trūkumus, natūraliai kyla būtinybė ieškoti geresnių duomenų analizės priemonių. Tokios priemonės keliami reikalavimai tiesiogiai atitinka esamų priemonių trūkumus:

- Duomenų analizės priemonė turėtų būti lanksti – t.y. turėtų būti galimybė peržiūrėti norimus duomenis skirtingais pjūviais be papildomų programavimo darbų. Taip pat svarbi galimybė kurti norimas ataskaitas, diagramas.
- Duomenų analizė turi kuo mažiau įtakoti esamos darbinės sistemos darbą – reikalingi duomenys gali būti saugomi atskiroje duomenų bazėje, tad dirbant su jais, nereikia kreiptis į pagrindinę duomenų bazę.
- Duomenų analizės priemonės turi turėti galimybę išnaudoti naktinį sistemos darbą daugiausiai resursų reikalaujančioms operacijoms atlikti, tuo siekiant kuo mažiau įtakoti pagrindinės sistemos darbą.

Visas išvardintas savybes pilnai tenkina OLAP priemonės. OLAP (angl. On-line Analytical Processing) – tai technologija, leidžianti greitai, realiaame laike ir įvairiais įmanomais pjūviais peržiūrėti informaciją, naudojant duomenų modelį, kuris atspindi realų organizacijos veiklos vaizdą, kaip jį supranta vartotojas, t. y. organizacijos duomenų atvaizdas yra daugiamatis. Sukūrus duomenų kubą, OLAP duomenų bazės leis lengvai atlikti įvairius

veiksmus: perėjimą nuo bendresnių reikšmių iki individualių ir atvirkščiai, t. y. judėti hierarchine pjūvių (dimensijų) struktūra. Žinodami kurio nors produkto pardavimų pokytį per mėnesį visoje apskrityje, esant reikalui galime pažiūrėti išsamiau - kaip produkto pardavimai kito skirtinguose šios apskrities rajonuose ar net atskiruose padaliniuose.

Tad yra iškeltas uždavinys sukurti įmonės klientų analizės sistemą, kuri minimalia įtaką pagrindinės sistemos darbą, būtų lanksti ir efektyvi. Tad yra svarbu sukurti sistemą, kuri efektyviai keistųsi duomenimis su pagrindine sistema, kuo mažiau įtakodama pagrindinės sistemos darbą. Tam reikalinga sukurti metodiką, kuri leistų išvengti dažniausiai pasitaikančių klaidų bei leistų sukurti efektyviai veikiančią sistemą. Kuriant tokią metodiką, yra svarbu atlikti esamų literatūros šaltinių analizę, apžvelgti specialistų komentarus bei pasiūlymus – sistemos poreikiams pritaikyti ETL (ang. Extract, Transform and Load) priemones duomenų atnaujinimui, parinkti nagrinėjamam atvejui tinkamus duomenų saugyklos modelius, optimizuoti sistemos veikimą.

Tad apibendrinant, galima teigti, jog šio tyrimo objektas – tai vaistinių tinklo klientų analizės modelis ir jo realizacija, kuri leidžia analizuoti įmonės klientų įpročius bei poreikius realiuoju laiku ir minimaliai įtakojant pagrindinės sistemos darbą. Norint tai pasiekti, reikia atlikti šiuos uždavinius:

- Apžvelgti esamus verslo analizės sprendimus ir atrinkti geriausias jų savybes bei jas panaudoti kuriamoje sistemoje.
- Atlikti būsimų kuriamos sistemos vartotojų poreikių analizę.
- Pagal iškeltus reikalavimus suprojektuoti sistemos modelį, kuris atitiktų iškeltus vartotojų reikalavimus.
- Suprojektuoti ETL procesus duomenų paėmimui iš darbinės sistemos, parinkti tinkamas duomenų saugyklos schemas kuriamai sistemai.
- Sukurti vartotojų reikalavimus atitinkantį įmonės klientų analizės kubą.
- Ištestuoti suskurtos sistemos veikimą.

Projekto analizės etapas yra labai svarbus, siekiant užtikrinti bendrą projekto sėkmę, siekiant sukurti sistemą, atitinkančią vartotojų poreikius. Tad analizės etapo metu reikalinga apžvelgti esamus ETL sprendimus, sukurti kuriamos saugyklos modelį, apžvelgti įgyvendinimo galimybes. Taip reikia nepamiršti ir duomenų saugumo užtikrinimo, realizuoti sistemą, atsižvelgiant į esamą techninę bazę ir pan.

Siekiant užtikrinti kuriamos sistemos kokybę, būtina numatyti projekto kokybės įvertinimo kriterijus:

- Sistemos našumas – sistema turi veikti efektyviai, geritai ir minimaliai įtakoti pagrindinės sistemos darbą.
- Patikimumas – pateikiami sistemos duomenys turi būti patikimi ir korektiški.
- Saugumas – sistema turi būti apsaugota nuo nesankcionuotų veiksmų bei užtikrinti priėjimą prie informacijos tiems vartotojams, kurie turi teisę.

Darbo metu sukurta duomenų saugykla optimizuota, ištestuotas jos veikimas.. Visas projektas susideda iš kelių etapų:

- Analizės etapas – šio etapo metu išanalizuoti įmonės verslo procesai, jų dalyviai bei reikalavimai būsimai sistemai. Atlikta esamų sprendimų, aprašytų literatūros šaltiniuose, analizė.
- Reikalavimų specifikacija – tai etapas, kurio metu identifikuojai būsimų sistemos vartotojų reikalavimai sistemai, jos veikimui bei funkcionalumui.
- Projektavimas – tai projekto etapas, kurio metu pagal reikalavimų specifikacijos metu gautą informaciją parengiamas sistemos projektas, numatomas sistemos veikimas, duomenų kubo atnaujinimo intervalai ir pan.
- Testavimas – šio etapo metu testuojama sukurtoji sistema, jos veikimas. Sistema optimizuojama, siekiant padidinti jos efektyvumą, pašalinamos pastebėtos klaidos bei trūkumai. Taip pat numatomos tolimesnės sistemos plėtojimo galimybės.
- Galiausiai, apibendrinami projekto rezultatai, suformuojamos rekomendacijos panašių sistemų kūrimu, apžvelgiamos ateities perspektyvos, pateikiamos darbo išvados.

2. Vaistinių tinklo duomenų saugyklos analizė

2.1. Analizės tikslas

Santykių su klientais valdymas (sutr. angl. CRM) – tai įmonės klientų poreikių tenkinimas, panaudojant žinias, gautas iš klientų valdymo sistemų, panaudojant įmonės bei įvairius išorinius duomenis. CRM sistemos pagrįsto paprasta idėja – tas, kuris geriau žino kliento poreikius, tas ir išlaiko ilgalaikius ir naudingus santykius su tuo klientu. Tad CRM sistemos pagrindinė užduotis yra maksimizuoti santykius su klientais laiko atžvilgiu, koncentruojant visus verslo aspektus (marketingo, pardavimų, veiklos ir aptarnavimo), kad įtvirtinti ir palaikyti abipusiai naudingus tarpusavio santykius. Norėdama to pasiekti, kiekviena organizacija turi sugebėti susidaryti vieną, integruotą kiekvieno kliento vaizdą.

Duomenų saugyklos misija - sudaryti nuoseklią ir išbaigtą verslo informaciją visiems įmonės padaliniams, pagrįstą operaciniais, sprendimų priėmimo paramos, ir kitais duomenimis. Norint tai pasiekti, įmonės duomenys turi būti išanalizuoti, suprasti, transformuoti ir pristatyti ten, kur jų reikia. Dėl to duomenų saugyklos administravimas turi apimti visos duomenų saugyklos terpės kūrimą, duomenų nukreipimą, valdymą ir tvarkymą. Vis dėlto, dauguma šiandieninių duomenų saugyklų nėra parengtos klientų valdymo sistemų metams iššūkiams ir nemažai dar turi būti nuveikta, kad tai būtų pasiekta. Tai reikalauja papildomų ar išplėstinių dalykinių sričių, įvairių išorinių duomenų bei didelės apimties duomenų apdorojimo. Šiandien didžiausias duomenų saugyklų iššūkis – tai klientų valdymo sistemų palaikymas. Duomenų saugyklos, kurios tenkina klientų valdymo sistemų reikalavimus, yra pelnytai vadinamos klientų duomenų saugyklomis (sutr. Ang. CDR – Customer Data Repository).

2.2. Tyrimo sritis, objektas ir problema

Šio magistrinio darbo tyrimo sritis – tai įmonės klientų duomenų analizės sistemos modelio sudarymas.

Tyrimo objektais – įmonės klientų duomenų saugyklos projektavimo ir kūrimo procesas.

Kuriant ir projektuojant duomenų saugyklas, susiduriama su įvairiomis problemomis. Reikės išanalizuoti problemas, su kuriomis susiduriama kuriant ir projektuojant duomenų saugyklas ir pasiūlyti tinkamiausią sprendimą, kuris padėtų kurti efektyvias įmonės klientų duomenų saugyklas ir analizės modelius.

Įmonėms kasdien susiduriant su vis didėjančia konkurencija, tokie įmonės uždaviniai kaip marketingo valdymas, pardavimų valdymas, klientų aptarnavimas ir palaikymas iškelia tokias pagrindines problemas:

- ✓ Kaip pagerinti pardavimų darbuotojų sugebėjimus taip, kad tai leistų padidinti pardavimų apimtį?
- ✓ Kaip efektyviai pritraukti daugiau klientų?
- ✓ Kaip greitai ir efektyviai nukreipti produktus ir paslaugas, kuo efektyviau panaudojant turimas lėšas?
- ✓ Kaip efektyviai išlaikyti ir kontroliuoti vertingus klientų resursus taip, kad nebūtų netikėto jų praradimo?
- ✓ Kaip pagerinti klientų aptarnavimo rodiklius taip, kad klientai gautų veiksmingą, greitą ir precizišką aptarnavimą, taip padidinant klientų lojalumą?
- ✓ Kaip pritraukti daugiau vertingų klientų ir atsiimti jų pritraukimo išlaidas?
- ✓ Kaip atrasti tokias klientų patenkinimo taisykles, kad atitinkamai galima būtų pakreipti marketingo sprendimus?

Klientų valdymo sistemos ypač pagelbsti sprendžiant aukščiau išvardintas problemas.

Klientų valdymas – tai nauja veiklos sritis, skirta supaprastinti santykius tarp įmonių ir klientų ir palengvinti su klientais susijusių marketingo, pardavimo sričių veiklas. Visa tai sudaro trys pagrindinės taisyklės:

1. Įkuriant, gerinant ir plečiant įmonių „įmonė – klientas rūpinimosi“ tinklus.

Įmonės gali naudoti tiek tradicinius, tiek elektroninius (pavyzdžiui, internetą), siekiant pritraukti daugiau klientų, naudotis įmonės produktais ar paslaugomis, padarant produktus ir paslaugas pasiekiamus virtualiai bet kada, bet kur, per bet kokį pardavimų ar slaugų kanalą. Taip įmonės gali labiau patenkinti klientų poreikius ir tuos klientus ilgiau išlaikyti.

2. Greitas, tikslus ir patikimas bendravimas su klientais

Jei įmonė yra pilnai pajėgi greitai, tiksliai ir patikimai dirbti su klientų poreikiais susijusiose srityse, tokiose, kaip marketingas, pardavimai, paslaugos ir t.t., tuomet įmonė tampa į klientus orientuota įmone. Pasitelkiant verslo procesų kontrolę, įtraukiant sukauptą klientų, pardavimų ir paslaugų informaciją, įmonės derina savo veiklą, kad būtų patenkinti klientų poreikiai.

3. Gerinami santykiai su klientais, pasitelkiant verslo išvalgos sistemas

Įmonės bendradarbiavimo su klientais metu sukaupiama dideli kiekiai vertingos informacijos, kuria galima apdoroti, analizuoti. Taip galima gauti

vertingos informacijos, kuri gali būti pagrindu įmonės veiklos planavimui ir veiklos sprendimų priėmimui. Elektroninė komercija šiuo metu ypač plėtojasi ir vystosi, tapdama nauju verslo veiklos modeliavimo ir vykdymo instrumentu. Įmonės elektroninės komercijos sistemose, prekyba ir klientai yra pagrindas, o tvirti ir ilgalaikiai santykiai su klientais yra labai svarbūs sėkmingam įmonės vystymuisi. Dėl visų šių priežasčių, į klientus orientuotos, klientų valdymo sistemos pamažu tampa vienomis pagrindinių įmonių elektroninės komercijos sistemomis. Dažniausiai, įmonių klientų valdymo sistemos susideda iš keturių pagrindinių dalių – tai pardavimų, marketingo, produktų platinimo ir paslaugų palaikymo valdymo. Kaip alternatyvą, klientų aptarnavimas telefonu galėtų būti penktoji klientų valdymo sistemos dalis. Daugumoje vakarų šalių įmonių, klientų valdymo sistemos padėjo įmonėms gauti naudą, pritraukiant naujų klientų, taupant laiką, padidinant darbo efektyvumą bei didinant pelną. Tai ypač pastebima tose įmonėse, kurių veikla labiau susijusi su tiesioginiais santykiais su klientais (pavyzdžiui, bankai, draudimo bendrovės, nekilnojamojo turto agentūros, ryšių tarnybos, aviakompanijos, transporto įmonės, saugos tarnybos).

2.2.1. Situacija šiandieninėje rinkoje

Šiuolaikinė rinka kinta labai sparčiai. Globalizacija, įmonių susiliejamai ir įsigijimai, stiprėjanti organizacijų konkurencija, technologinės naujovės skatina kompanijas pergalvoti savo verslo strategijas. Stipri tarpusavio konkurencija iki minimumo sumažino pelno maržas. Be to, Lietuvai tapus Europos Sąjungos nare, rinkoje gali atsirasti dar daugiau stiprių žaidėjų ir konkurentų.

Visa tai kompanijas skatina atrasti naujus pajamų šaltinius, naujas rinkas, padidinti pelningumą, užimamą rinkos dalį. Tačiau tai padaryti reikia mažinant operacines išlaidas. Šis veiksnys tampa prioritetiniu beveik visoms Lietuvoje įsikūrusioms kompanijoms. Todėl mažinami personalo kaštai, automatizuojami brangiai kainuojantys procesai, nuolat ieškoma kitų galimybių.

Tuo pačiu keičiasi ir vartotojų poreikiai. Didėjant konkurencijai tarp įmonių, jie tampa vis geriau informuoti ir daugiau reikalauja. Dėl to dauguma kompanijų pertvarko savo strategijas, kad jos būtų labiau orientuotos į klientus, o ne į siūlomus produktus.

Dėl tokios nuolat kintančios, sudėtingos išorinės aplinkos ir augančių reikalavimų, šiuolaikinės kompanijas yra sudėtinga valdyti. Tiek vadovams, tiek ir visiems darbuotojams

reikalinga vieninga, tiksli, nuolatinė verslo informacija, kurios pagalba būtų galima optimizuoti pinigų srautus, užtikrinti procesų efektyvumą ir sumažinti kaštus. Tad pagrindinė priemonė atremti visus rinkos iššūkius ir priimti efektyvius strateginius sprendimus yra disponavimas pagrindiniu turtu šiuolaikinėje ekonomikoje – informacija.

2.2.2. Verslo analitika

Verslo analitika (*Business Intelligence*) yra terminas, naudojamas apibrėžti operatyvinių įmonės duomenų panaudojimą informacijos ir žinių gavimui, taip reikalingų efektyvių strateginių bei kasdieninių sprendimų priėmimui.

2.2.2.1. Kuo gali padėti verslo analitika?

Per pastaruosius kelerius metus įmonėse sparčiai išaugo prieinamos informacijos kiekiai, ir kompanijos kiekvieną dieną sukaupia milžiniškus duomenų kiekius. Atsižvelgiant į tai, kad kompanijos susiduria su didėjančiais reikalavimais aktyviai bei sparčiai reaguoti į rinkos jėgas ir tendencijas, duomenų panaudojimo efektyvumas tampa labai svarbiu veiksniumi. Visiems organizacijos darbuotojams reikia priėjimo prie informacijos, kuri padėtų jiems priimti pagrįstus sprendimus, imtis koreguojančių priemonių ir patobulinti verslo procesus. Vadovai bei departamentai neturėtų būti priklausomi nuo IT išteklių ir specialistų sugebėjimų padėti. Informacija turi būti prieinama visoje kompanijoje tuo metu, kai tik jos prireikia, užtikrinant, kad kiekvienas turėtų jam reikalingus duomenis.

2.2.2.2. Verslo analitikos privalumai

- Kaštų mažinimas – mažiau laiko sugaištama informacijos apdorojimui – padidėja darbo efektyvumas, IT personalas perkeliamas prie projektų, sukuriančių didesnę vertę organizacijai, automatizuojami daug kainuojantys informacijos srautai.
- Didinamas pelningumas – geriau išskiriami klientų elgesio ir pirkimo įpročiai, pagal tai diferencijuojami produktai ir paslaugos.
- Pagerinama komunikacija – nuolat keičiamasi svarbia valdymo informacija tarp departamentų.
- Mažinami klientų pritraukimo kaštai – nustatomos pelningiausių klientų charakteristikos ir vykdomos orientuotos marketingo kampanijos.
- Gerinamas klientų aptarnavimas – greitai reaguojama į besikeičiančias rinkos tendencijas.

2.2.2.3. Pagrindinės kompanijų problemos

Šiandien bene kiekvienoje įmonėje rasime įdiegtą verslo valdymo ar finansinę sistemą, kuri dažnai yra papildoma ir kitomis IT sistemomis: gamybos valdymo, sandėlių valdymo, santykių su klientais valdymo, personalo valdymo ar kt. Visa įmonės veiklos ir verslo analizė

remiasi į aukščiau paminėtų sistemų ataskaitų posistemas, t.y. ataskaitų rinkinius, kurias siūlo įdiegtos įvairiausios IT sistemos.

Iš čia kyla visa eilė kasdieninių problemų, kurias tenka sutikti dažnoje įmonėje:

- Ataskaitų ribotumas – jei nėra sukurtos ir įdiegtos reikalingos bei kompleksinės ataskaitos, įmonės analizė norimu pjūviu tampa sunkiai įvykdoma arba iš viso negalima.
- Naujos ataskaitos kuriamos neleistinai ilgai, jei iš viso yra sukuriamos. Taip pat kiekvieną kartą tenka mokėti sistemos diegėjams ar prižiūrėtojams už naujų ataskaitų kūrimą ar esamų modifikavimą.
- Ataskaitų generavimas yra kasdien ilgėjantis procesas, kadangi duomenų kiekis sparčiai didėja. Neretai dienomis jau negalima daryti ataskaitų, nes tai itin stabdo finansinės ir kitų sistemų darbą.
- Dėl didelio duomenų kiekio įmonė ir jos verslas yra analizuojamas tik makro-lygiu, nes labiau detalizuoti informacijos neleidžia turimos technologijos, gaminių ir kitų analizės objektų nomenklatūra.
- Vartotojai, kuriems reikalingi įmonės duomenys, juos apdoroja įvairiausiomis priemonėmis. Todėl būtinai atsiranda klaidų, nesutapimų, skirtingo vienodų objektų traktavimo.
- Keičiant turimą finansinę sistemą kita ar apjungiant skirtingas sistemas, yra prarandama dalis įmonės istorinių duomenų bei galimybė analizuoti verslą, nes visų duomenų perkėlimas į naują finansinę sistemą – itin sudėtingas uždavinys.
- Turint kelias IT sistemas nėra matomas vieningas įmonės vaizdas.
- Nutolę padaliniai neturi galimybės interaktyviai analizuoti norimus duomenis.
- Nepakankamas arba jokie duomenų saugumo.
- Nėra galimybės perkelti duomenis į kitas Windows aplikacijas ar aplinkas: Excel, Word, PDF, HTML ir kt. bei keistis duomenimis.

Apibendrinant aukščiau paminėtus faktus, kiekviena įmonė gali susidurti su tokiais pagrindiniais trūkumais: duomenų gavimo greitis, duomenų ir informacijos analizės funkcionalumas, galimybės analizuoti visą įmonę ar įmonių grupę vienoje aplinkoje nebuvimas.

Norint spręsti minėtas problemas, yra **reikalinga verslo duomenų analizės sistema**, kuri pirmiausia apima funkcionalų duomenų ir informacijos analizės bei vizualizavimo įrankį.

2.2.2.4. OLAP duomenų bazės

Daugelį aukščiau minėtų problemų gali išspręsti OLAP duomenų bazės [1], [2]. Tokiu atveju darbinėse reliacinėse duomenų bazėse ar duomenų saugyklose saugomi pradiniai duomenys transformuojami, ir sukuriamos optimizuotos duomenų saugojimo struktūros - OLAP duomenų kubai (data cubes), kurie specialiai pritaikyti greitai duomenų analizei.

OLAP kubuose didelė tarpinių skaičiavimų dalis atliekama dar iki duomenų kubo naudojimo, o vieną kartą atliktų tarpinių skaičiavimų rezultatais gali pasinaudoti visi prieigos teises turintys šios duomenų bazės vartotojai.

Kadangi OLAP duomenų bazės saugo tarpines agregatines reikšmes, pagal duomenų kubo įgyvendinimo būdą duomenų pasikeitimai ar naujų duomenų įkėlimas į OLAP duomenų bazę gali pareikalauti atnaujinti ar pertvarkyti kubo informaciją (turi būti perskaičiuotos kai kurios agregatinės reikšmės).

2.2.2.4.1. Dimensijos ir skaičiuojamosios reikšmės

Aprašant OLAP duomenų struktūras [15] (kubus) naudojama dimensijų sąvoka (dar kitaip tai galima įsivaizduoti kaip duomenų peržiūrą tam tikrais pjūviais). OLAP duomenų bazėje esantys įvykiai (eilutė lentelėje) gali būti analizuojami pagal įvairius parametrus – įvairiomis dimensijomis.

Pamėginsime paaiškinti dimensijų sąvoką naudodami pardavimų modelio pavyzdį. Užregistruotas prekių pardavimo įvykis apibūdinamas keliais parametrais (kada tai įvyko, koks produktas parduotas, kuriame padalinyje), todėl ir įmonės mastu pardavimus galima analizuoti įvairiomis dimensijomis:

- Pagal laiką. Galima sužinoti, kaip prekyba sekėsi šį mėnesį, palyginti su praėjusiu mėnesiu, su praėjusiais metais ir pan.
- Pagal produktą. Parduodamos įvairios produktų grupės, jų tipai ir pagaliau - konkretūs produktai. Galima sužinoti, kokią pajamų dalį sugeneravo ne tik konkretus produktas, bet ir viena ar kita produktų grupė, koks perkamiausias produktas ir pan.
- Pagal vietovę. Didelės organizacijos pardavimus vykdo didelėje teritorijoje, todėl konkrečios parduotuvės gali būti priskirtos prie apskričių, rajonų, rajonuose gali būti numatyti mažesni struktūriniai vienetai, pavyzdžiui, miestai ir pan. Tokiu būdu galima atsakyti, kaip sekėsi parduoti tą ar kitą prekę apskrityje, rajone ar mieste.

Kaip matome iš pavyzdžio, dimensijos gali turėti tam tikrą hierarchinę struktūrą, pavyzdžiui, vietovės ir laiko dimensijos turi net po tris hierarchinius lygius (apskritis > rajonas >

miestas ir metai > ketvirtis > mėnuo). Tokia hierarchinė struktūra gali būti ir labai sudėtinga. Šiuo atveju sukūrėme trimatį duomenų kubą, kurio matmenis apibrėžia laiko, produkto ir vietovės dimensijos. Tačiau loginė duomenų kubo struktūra gali būti ne tik trimatė, bet ir daugiamatė (keturmatė, penkiamatė ir t. t.).

Kita pagrindinė OLAP duomenų bazių sąvoka yra skaičiuojamoji vertė (angl. measure). Galima pasirinkti, kokią su konkrečiu įvykiu susijusią skaitinę reikšmę analizuosime. Pateikiamame pavyzdyje galėtume analizuoti ne tik pajamas iš šiamo mieste per mėnesį parduoto konkretaus produkto, bet ir jo pardavimo pelną ar parduotą prekių vienetų skaičių.

Sukūrus duomenų kubą, OLAP duomenų bazės leis lengvai atlikti įvairius veiksmus:

Perėjimą nuo bendresnių reikšmių iki individualių ir atvirkščiai, t. y. judėti hierarchine dimensijų struktūra. Žinodami kurio nors produkto pardavimų pokytį per mėnesį visoje apskrityje, esant reikalui galime pažiūrėti išsamiau - kaip produkto pardavimai kito skirtinguose šios apskrities rajonuose ar net atskirose parduotuvėse.

Duomenų apibendrinimą pagal įvairius parametrus, t. y. analizę įvairiomis dimensijomis. Pavyzdžiui, ataskaitas apie pelną pardavus tam tikrą produktą galime analizuoti ne laike, bet pagal vietovę.

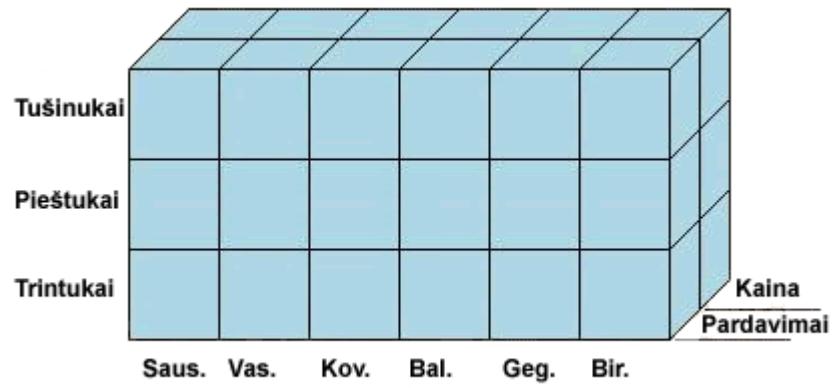
Analizuojamojo matmens pakeitimą, t. y. analizę pagal skirtingas skaičiuojamąsias reikšmes.

Pavyzdžiui, turime duomenų lentelę (1 lentelė).

1 lentelė.

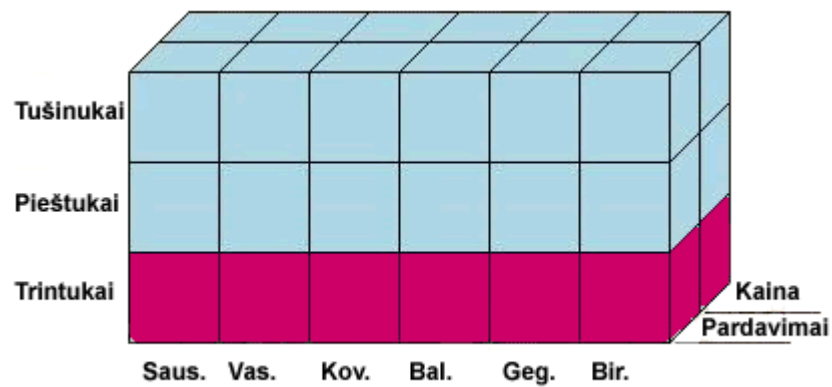
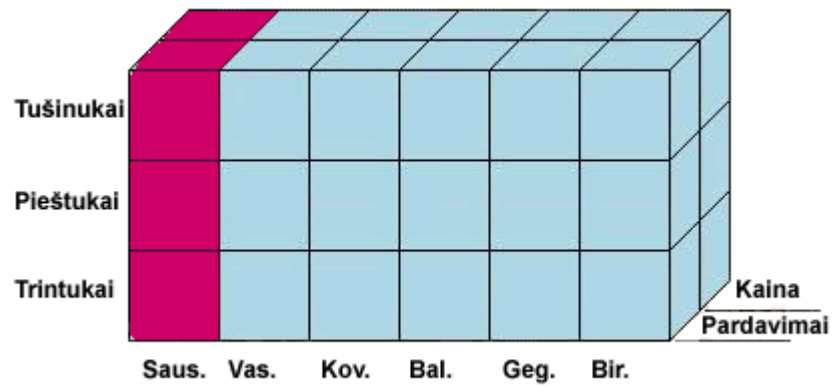
Prekės	Pardavimai	Kaina
Tušinukai	250	191
Pieštukai	322	217
Trintukai	108	83

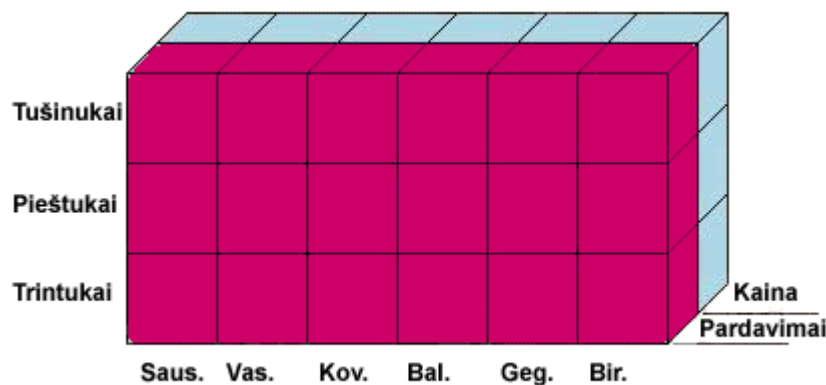
Tokiu atveju, galima sudaryti duomenų kubą (1 paveikslėlis).



1 pav. Duomenų kubas

Turimą duomenų kubą galima peržiūrėti įvairiomis dimensijomis (2 pav.)





2 pav. Duomenų kubo dimensijos

2.3. Vartotojų analizė

2.3.1. Vartotojų aibė, tipai ir savybės

Galima išskirti šias vartotojų grupes:

- Būsiami/esami duomenų saugyklos bei OLAP sistemų vartotojai – jie pateikia reikalavimus būsimai sistemai ir naudoja galutinai realizuotą sistemą;
- Duomenų bazės administratorius – pateikia duomenis, duomenų bazės schemas ir duomenų šaltinius saugyklos projektuotojui;
- Saugyklos projektuotojas –
 - išanalizuoja vartotojo reikalavimus;
 - UML įrankių pagalba sudaro pradinį modelį;
 - realizuoja sistemą.

2.3.2. Vartotojų tikslai ir problemos

Duomenų saugyklų bei OLAP sistemų vartotojams yra gana sunku tiksliai apibrėžti reikalavimus būsimai sistemai ir juos pateikti analitikui.

Duomenų bazių administratoriaus paskirtis – tai reliacinių duomenų bazių priežiūra, veiksmingumo užtikrinimas. Nusprendus kurti daugiamatę saugyklą, reikalinga aukštesnė administratoriaus kvalifikacija bei žinios.

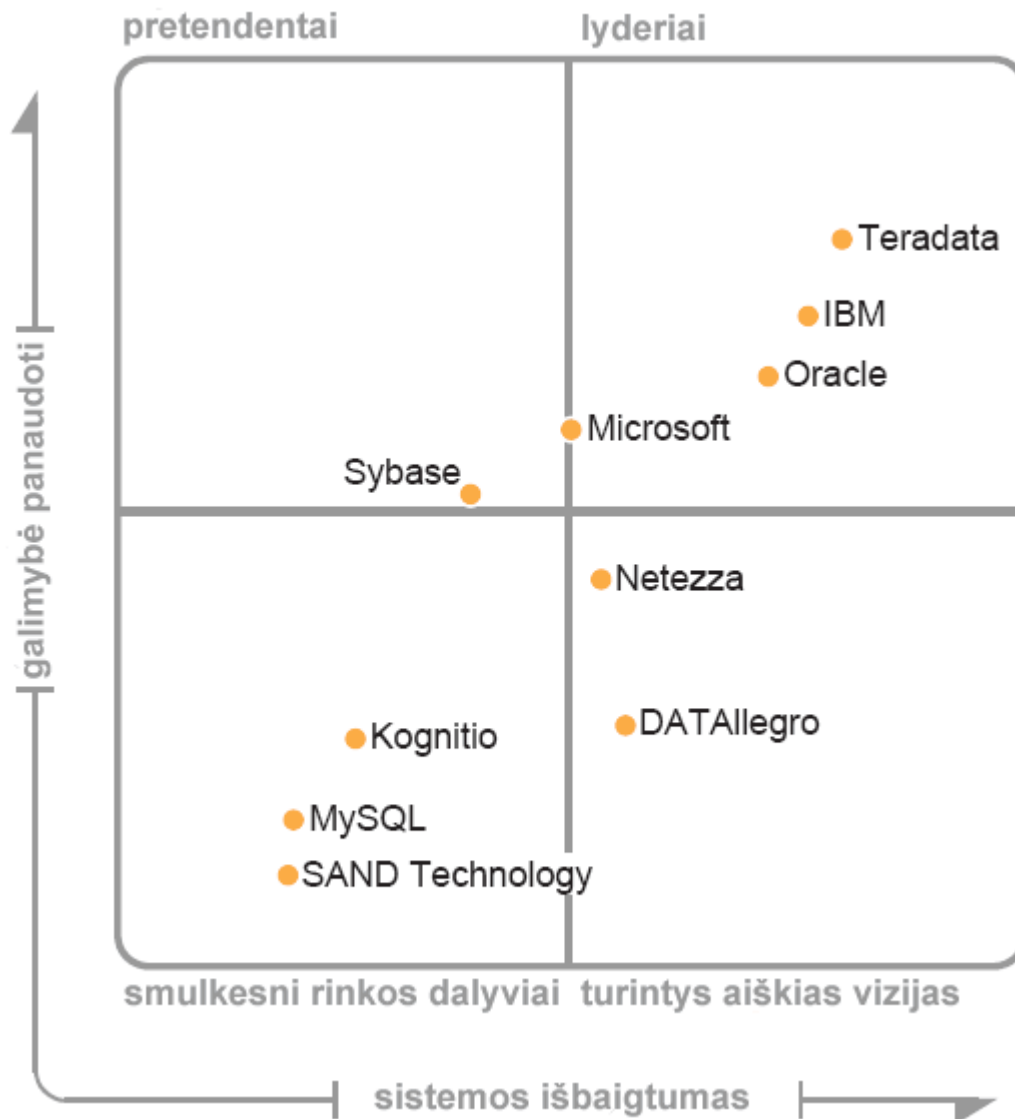
Duomenų saugyklų projektuotojai – viena svarbiausių šio tyrimo vartotojų aibių. Projektuotojų pagrindinis tikslas – pilnai suprojektuoti teisingai ir patikimai veikiančią, greitą ir saugią sistemą. Kiekviename sistemos kūrimo etape ši vartotojų grupė susiduria su atitinkamomis problemomis bei sunkumais.

2.4. Saugyklų kūrimo metodikos analizė

Tobulėjant šiuolaikinėms technologijoms, didėjant duomenų kiekiui darosi vis sudėtingiau juos išanalizuoti ir daryti greitus, efektyvius ir teisingus sprendimus. Duomenų bazės jau peržengė terabaitines ribas ir žmogus jau tampa nepajėgus išanalizuoti visą duomenų gausą. Tokiame milžiniškame kiekyje informacijos gali slėptis ir strategiškai svarbi ir niekinė informacija. Tokios problemos paskatino atsirasti aukštos kokybės taikomiesiems paketams, programavimo įrankiams, duomenų analizės priemonėms, kurios padeda nepasimesti informacijos gausoje. Savo ruožtu tai kartu padidino ir vartotojų prieinamumą prie pažangiausių technologijų, atvėrė elektroninės komercijos, on-line analizės ir kitas galimybes.

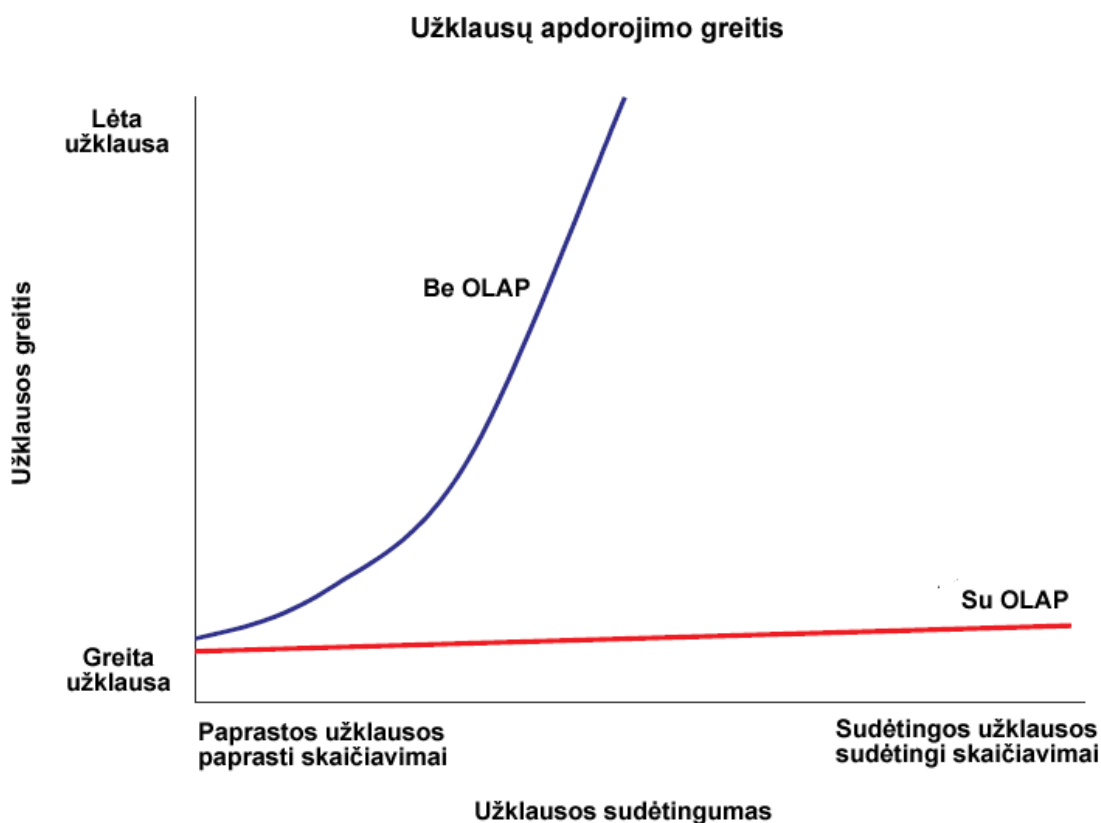
Mes daugiau kalbėsime apie OLAP duomenų analizės įrankius. Šioje srityje yra nemažai nuveikta. Per pastaruosius 18 mėnesių, kai kurie duomenų saugyklų kūrėjų rinkos naujokai, gana sparčiai iškilo nuo mažų įmonių iki vidutinių ir dar didesnių. Šiuo metu duomenų saugyklų rinka orientuojasi į verslo informaciją (Business Intelligence). Gana gausiai į šią rinką ateina nauji programinės įrangos gamintojai, taip pat nesnaudžia ir senieji, siūlydami naujus sprendimus. Skirtingai nei anksčiau, dabar duomenų bazės dydžiai, kuriais gali operuoti duomenų sandėliai, nėra tokie aktualūs. Šiandien, duomenų saugyklos, kurių apimtis siekia nuo 5 iki 20 terabaitų, yra dominuojančios [8].

Pasižiūrėjus į duomenų saugyklų programinės įrangos rinką, galima ją pavaizduoti plokštumoje (3 paveikslėlis).



3 pav. Duomenų saugyklų programinės įrangos gamintojai

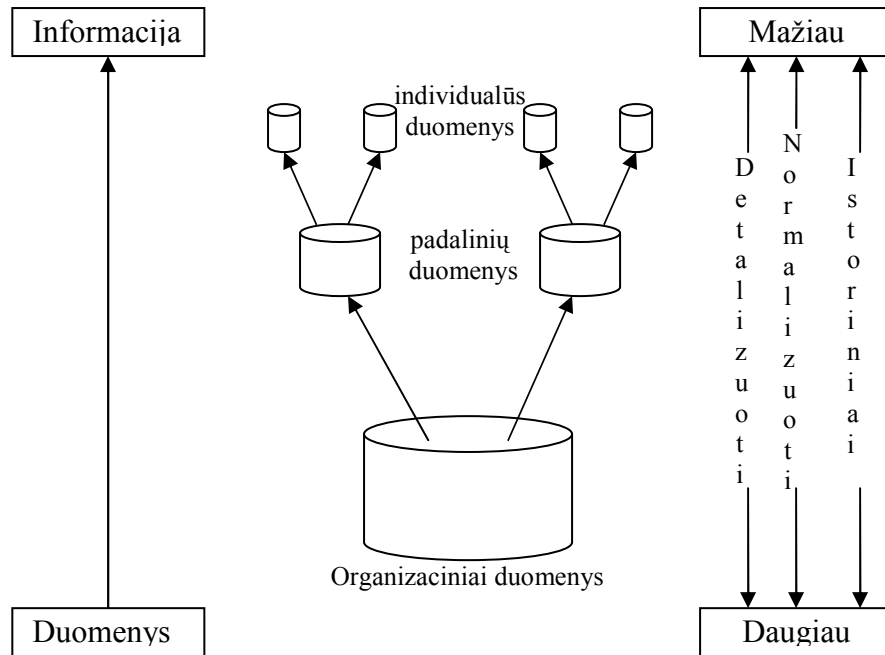
OLAP duomenų analizės privalumai yra akivaizdūs – tai spartesnis duomenų apdorojimas, jų analizė, palyginus su įprastomis užklausomis. OLAP sistemų teikiamus privalumus iliustruoja 4 paveikslėlyje pateiktas grafikas.



4 pav. Užklausų apdorojimas, pasitelkiant OLAP analizės įrankius

Informacijos apdorojimo tikslas yra paversti duomenis naudinga informacija. Online analytical processing (OLAP) yra svarbus metodas, padedantis pasiekti šį tikslą, dirbant su duomenų saugyklomis (angl. Data Warehouse). Nuolat didėjant duomenų bazėse saugomų duomenų kiekiui, vis didėjant jų apdorojimo laikui, duomenų saugyklų poreikis ir panaudojimas vis didėja. OLAP (angl. Online Analytical Processing) yra interaktyvi verslo informacijos analizė [2], [3]. Sistemos vartotojai gali peržiūrėti informaciją įvairiomis dimensijomis (tokiomis kaip pelningumas, pardavimai, išlaidos ir kt.). OLAP sistemos pagalba, vartotojas gali pereiti nuo vienos verslo duomenų dimensijos (pvz.: „metiniai visų padalinių pardavimai“) prie kitos (pvz.: „pelningiausi padaliniai per paskutiniuosius tris mėnesius“) ir pereidinėti tarp skirtingų duomenų detalumo lygių (pvz.: dienos, mėnesio ar ketvirčio pardavimai“). Ši interaktyvi duomenų analizė yra vadinama multidimensine analize. Pagrindinis faktorius, apibrėžiantis OLAP įrankius – yra daugybė skirtingų pagrindinio tikslo įgyvendinimų – tai duomenų analizės variklis, kuris organizacijos duomenis perkelia į multidimensinę erdvę interaktyviai analizei.

5 paveikslėlis parodo duomenų saugyklos vaidmenį, verčiant duomenis informacija ir keletą pagrindinių skirtumų tarp skirtingų duomenų saugyklų architektūrų tipų.



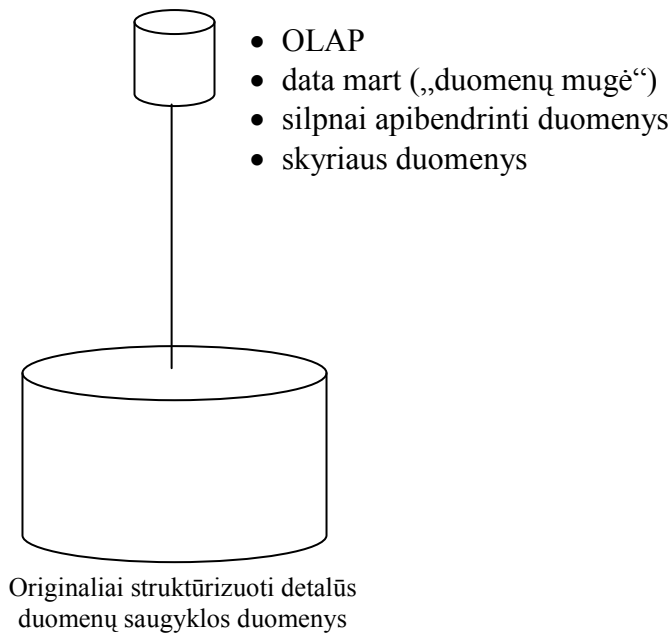
5 pav. Duomenų saugyklos architektūra

Skirtingi duomenų saugyklos tipai gali būti apibrėžti taip:

- **Organizacinė struktūra** – dar kitaip vadinama kaip „atominė“, „korporacinė“. Duomenų saugyklos paskirtis – patenkinti visus organizacijos informacinius poreikius.
- **Suskirstymas pagal padalinius** – šio tipo duomenų saugyklos duomenys yra struktūrizuoti taip, jog atitiktų keliamus informacinius reikalavimus atskiro padalinio, žiūrint pagal organizacijos veiklos funkcijas (pvz.: pardavimų skyrius, plėtros skyrius ir t.t.). Tokio tipo duomenų saugykloje saugomi duomenys, reikalingi konkrečiam organizacijos skyriui ar padaliniui.
- **Individuali struktūra** – duomenys šiame lygyje yra dar labiau struktūrizuoti, kad atitiktų konkrečios veiklos funkcijos informacinius reikalavimus.

2.4.1. Skirtingi OLAP duomenų lygių pavadinimai duomenų apdorojime

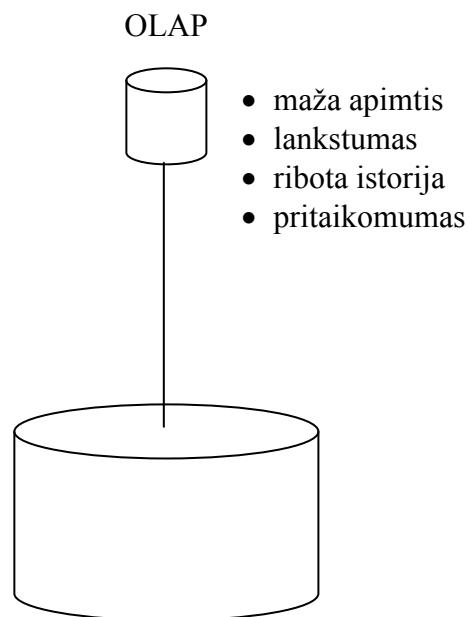
Didėjantis vartotojų poreikis informacijos apdorojimui ar sprendimų priėmimui (angl. decision support), skatina OLAP sistemų panaudojimą. OLAP – tai tam tikras duomenų saugyklos priedas. Iš tikrųjų, duomenų saugyklose, kur duomenys yra suskirstyti pagal atskirų organizacijos padalinių informacinius poreikius, yra idealiausia panaudoti OLAP sistemų teikiamus privalumus. Kitaip, šio lygio duomenų apdorojimas dar yra vadinamas „data mart“ (liet. „duomenų mugė“). Aštuntame paveikslėlyje yra pavaizduota, OLAP sistemų pavadinimai pagal apdorojamų duomenų lygius.



6. pav. Skirtingi OLAP lygių pavadinimai duomenų apdorojime

2.4.2. OLAP duomenų šaltiniai

OLAP sistemose yra naudojami duomenys, paimti iš duomenų saugyklų. Šiose duomenų saugyklose esanti detali, istorinė informacija, yra duomenų saugyklų pagrindas ir sudaro puikų pagrindą OLAP duomenų analizės sistemoms. Žemiau esantis paveikslėlis parodo ryšį tarp organizacijos poreikius atitinkančių duomenų lygio ir OLAP duomenų lygio.



7. pav. OLAP duomenų lygis

Keletas OLAP sistemose naudojamų duomenų lygio charakteristikų yra tokios:

- **Maža apimtis** – palyginus su visos organizacijos duomenimis, OLAP sistemose naudojama labai maža dalis šių duomenų.
- **Lankstumas** – OLAP sistemų veikimas ir duomenų apdorojimas yra kur kas lankstesnis, nei visos organizacijos duomenų sandėlio duomenų apdorojimas. Šis lankstumas pasiekiamas dėl to, jog OLAP naudojamas žymiai mažesnis kiekis duomenų, kuriuos reikia apdoroti, taip pat dėl to, jog OLAP sistemose naudojama programinė įranga yra orientuota į lankstumą, priešingai, nei programinė įranga, naudojama visos organizacijos duomenų sandėliuose, kuri yra sukurti apdoroti didžiulius kiekius duomenų.
- **Ribota istorija** – OLAP sistemos retai kada operuoja senesne, nei šešių mėnesių senumo informacija, o visos organizacijos duomenų saugomi duomenys paprastai siekia nuo penkių iki dešimties metų.

Pritaikomumas – OLAP priemonės galima pritaikyti konkrečioms organizacijos ar padalinio poreikiams, tuo tarpu visos organizacijos duomenys, esantys duomenų sandėliuose, yra kolektyviniai.

2.4.3. Bendrasis duomenų saugyklos proceso modelis

Paprastai, duomenų saugyklos sukūrimo uždavinys gali būti įgyvendintas skirtingais būdais arba pasitelkiant skirtingas technologijas. Skirtingos metodologijos palaiko skirtingus sprendimus ir technologijas. Dauguma duomenų saugyklų specialistų įvardija 10 duomenų saugyklos proceso etapų [9], [14]:

1) *Verslo reikalavimų surinkimas ir analizė*

Verslo reikalavimų surinkimo ir analizės etapas įtakoja praktiškai visus likusius duomenų saugyklos proceso etapus. Šiame etape, saugyklos projektuotojai turi išsiaiškinti ne tik kokius duomenis būsimi sistemos vartotojai nori matyti, bet ir kokio pobūdžio veiklas vartotojai atlieka, kokio tipo sprendimus jie priima. Egzistuoja keletas reikalavimų analizės technologijų, tokių kaip interviu, standartinių šablonų naudojimas, reikalavimo prioritetų keitimas, egzistuojančių dokumentų peržiūra ir kitos.

2) *Duomenų modeliavimas*

Yra skiriami trys duomenų modeliavimo lygiai: abstraktus, loginis ir fizinis modeliavimas. Abstraktus modeliavimas apima aistra reikalavimų duomenims modeliavimą aukštame lygyje, neatsižvelgiant į konkrečias DBVS. Loginis

modeliavimo etape, konceptuali schema yra paverčiama logine schema, kurią galima įgyvendinti pasirinktoje DBVS. Patys populiariausia loginiai duomenų modeliai yra reliaciniai ir dimensiniai duomenų modeliai. Galiausiai, fizinio duomenų modeliavimo etape, loginė schema yra pakeičiama fizine schema, kuri aprašo, kaip duomenų bazė yra realizuojama atminties lygmenyje.

3) *Duomenų atvaizdavimas*

Šis procesas apima skirtingų duomenų aprašymų atvaizdavimą duomenų saugykloje. Pavyzdžiui, duomenų šaltinių formatai paprastai yra paverčiami kuriamų lentelių atributų tipais.

4) *ETL projektavimas*

Rininy s funkcijų, skirtų duomenims išgauti, perkelti ir įkelti į duomenų sanėlį, yra vadinama ETL (angl Extract, Transform and Load).

5) *Vartotojo pusės programų kūrimas*

Duomenų saugyklos yra skirtos teikti informaciją vadovams ir darbuotojams. Jiems reikalinga informacija yra pateikiama per skirtingas vartotojo sąsajas, pasitelkiant įvairias užklausas, ataskaitas, OLAP funkcijas ir duomenų išgavimo priemones.

6) *Projekto planavimas*

Šis etapas apima organizacijos duomenų saugyklos įdiegimui pasiruošimo įvertinimą, projekto plano sukūrimą, resursų, reikalingų projektą įgyvendinti, apskaičiavimą, biudžeto išlaikymą ir kt.

7) *Realizavimas*

Yra keletas galimų sprendimų, realizuojant duomenų saugyklą atitinkamoje aplinkoje. Realizavimo etapas atitinka kurį nors iš programinės įrangos gyvavimo ciklų (krioklio, spiralės ar kt.).

8) *Duomenų kokybės valdymas*

Duomenų kokybės valdymas yra gana svarbus duomenų saugyklos proceso atributų. Tai duomenų, esančių duomenų saugykloje, kokybės užtikrinimas.

9) *Verslo nuoseklumo valdymas*

Dauguma duomenų saugyklų yra skirtos palengvinti strateginių klausimų priėmimą. Taigi, yra pakankamai svarbu, jog būtų galimybė kurti archyvus ir atstatyti duomenis jų sugadinimo atveju.

10) *Valdymo keitimas*

Dauguma duomenų saugyklų yra gana stipriai pakeičiamos, lyginant su tuo, kokios jos būna įdiegtos. Šiuos pakeitimus bet kokioje duomenų saugykloje gali būti gana sudėtinga kontroliuoti. Duomenys, susiję su duomenų saugyklos valdymu, neretai yra išsklaidomi ir saugomi įvairiomis formomis. Tačiau, bet kuriuo atveju, yra pakankamai svarbu, kad duomenų saugykla turėtų gerai apgalvotus valdymo keitimo planus.

2.5. Oracle ir MS SQL duomenų analizės kūrimo priemonių lyginamoji analizė

Esminis Microsoft ir Oracle teikiamų duomenų analizės įgyvendinimo priemonių skirtumas yra tai, jog kompanija Microsoft praktiškai visiškai atskyrė duomenų analizės serverius nuo duomenų bazės serverio. Microsoft Analysis Services [16], [17] dažniausiai veikia atskirame, analizės serveryje nei pati duomenų bazė.

Oracle [10], [11] kompanija pasuko kiek kitu keliu – nuo Oracle 9iR2 versijos, kompanija integravo analizės servisus į pačią DBVS. Šis sprendimas yra geras tuo, jog suaprastina visos sistemos priežiūrą, mažina lokalaus tinklo apkrovimą.

Visgi, OLAP serverio [13] buvimas kartu su DBVS serveriu kiek labiau įtakoja bendrą sistemos našumą, nei analizės ir DBVS serverių buvimas atskirai. Pastaruoju atveju, analizės duomenis galima naudoti praktiškai visiškai neįtakojant pagrindinės sistemos našumo. Toks sprendimas yra pakankamai aktualus, kuomet duomenų bazės serverio resursus tenka taupyti ir negalima dėl to lėtinti kitų įmonės veiklos procesų darbo.

Pagrindinės MS SQL ir Oracle duomenų analizės kūrimo priemonių savybės pateiktos žemiau esančioje lentelėje.

2 Lentelė. MS SQL ir Oracle lyginamoji savybių analizė

Savybė	MS SQL	Oracle
Palaikomos operacinės sistemos	Microsoft Windows šeios operacinės sistemos	Visos plačiau paplitusios operacinės sistemos
Daugiamačių duomenų saugyklos vieta	Atskiras servisas (Analysis Services)	Realiacinė Oracle duomenų bazė
Užklausų vykdymo vieta	Vykdomos atskiruose analizės servisuose	Vykdomos pagrindinės duomenų bazės branduolyje
Naudojamos programavimo kalbos	VB script	PL/SQL
OLAP užklausų kalba	MDX užklausos	DML kalba (daugiau naudojama darbui su analitinėmis aplinkomis. SQL kalba – paprastoms užklausoms.
Kubų peržiūros galimybė,	“Pivot Tables” komponentas	Neturi

naudojant MS Excel priemonės		
------------------------------	--	--

Didelis Microsoft Analysis Services [12] privalumas – tai Pivot Tables komponentas. Tai yra COM+ komponentas, kurį galima panaudoti Microsoft Office taikomosiose programose, pavyzdžiui MS Excel. Šis komponentas leidžia prisijungti prie analizės serveryje saugomų duomenų kubų, vykdyti čia daugiamačių duomenų (MDX) užklausas ir pateikia duomenis lentelėmis arba grafikų pagalba. Šio komponento privalumas yra tai, jog jis pilnai integruojasi į tokias vartotojams įprastas programas kaip Microsoft Excel. Juo paprasta naudotis tiems, kas turi bent bazines Microsoft Excel žinias. Tai yra gana svarbi savybė, nes daugeliu atvejų, analitikai yra įpratę naudoti Microsoft produktus duomenų analizei.

Dar vienas svarbus aspektas – tai licenzijų kaina. Neseniai Oracle kompanija sumažino savo produktų licenzijų kainas, siekdama prisitraukti dalį klientų iš Microsoft kompanijos. Tačiau Microsoft kompanijos produktų dominavimas rinkoje, vartotojų įpratimas (pvz.: dirbti su MS Excel programomis), kol kas leidžia būti Microsoft kompanijai šios rinkos lydere. Bazinės Oracle ir Microsoft siūlomų produktų kainos pagal serverių procesorių skaičių, pateiktos žemiau esančioje lentelėje.

3 Lentelė. MS SQL ir Oracle lyginamoji savybių analizė

Procesorių skaičius	Oracle 11g Enterprise Edition	Oracle 11g Enterprise Edition su OLAP arba Data Mining įrankiais	Oracle 11g Enterprise Edition su OLAP ir Data Mining įrankiais	MS SQL 2008 Enterprise Edition
1	40.000 USD	60.000 USD	80.000 USD	24.999 USD
2	80.000 USD	120.000 USD	160.000 USD	49.998 USD
4	160.000 USD	240.000 USD	320.000 USD	99.996 USD
8	320.000 USD	480.000 USD	640.000 USD	199.992 USD
16	640.000 USD	960.000 USD	1.280.000 USD	399.984 USD
32	1.280.000 USD	1.920.000 USD	2.560.000 USD	799.968 USD

Skirtingai nei Oracle siūlomos variacijos, MS SQL Enterprise Editon versijoje nėra atskirų licenzijų OLAP ar duomenų gavybos (angl. Data Mining) priemonėms – visos jos jau ir taip įeina į standartinį paketą.

Taip pat skiriasi ir Microsoft ir Oracle kompanijų politika, pateikiant vartotojams bandomąsias naujų sistemų versijas. Microsoft SQL serverio bandomųjų versijų neįmanoma laisvai parsisiūsti iš kompanijos tinklapio. Jas gali gauti tiksliai registruoti programinės įrangos testuotojai, programinės įrangos platintojai ir pan. Šiems klientams bandomosios sistemų versijos atsiunčiamos kompaktiniuose diskuose ir neretai jos turi veikimo laiko limitą.

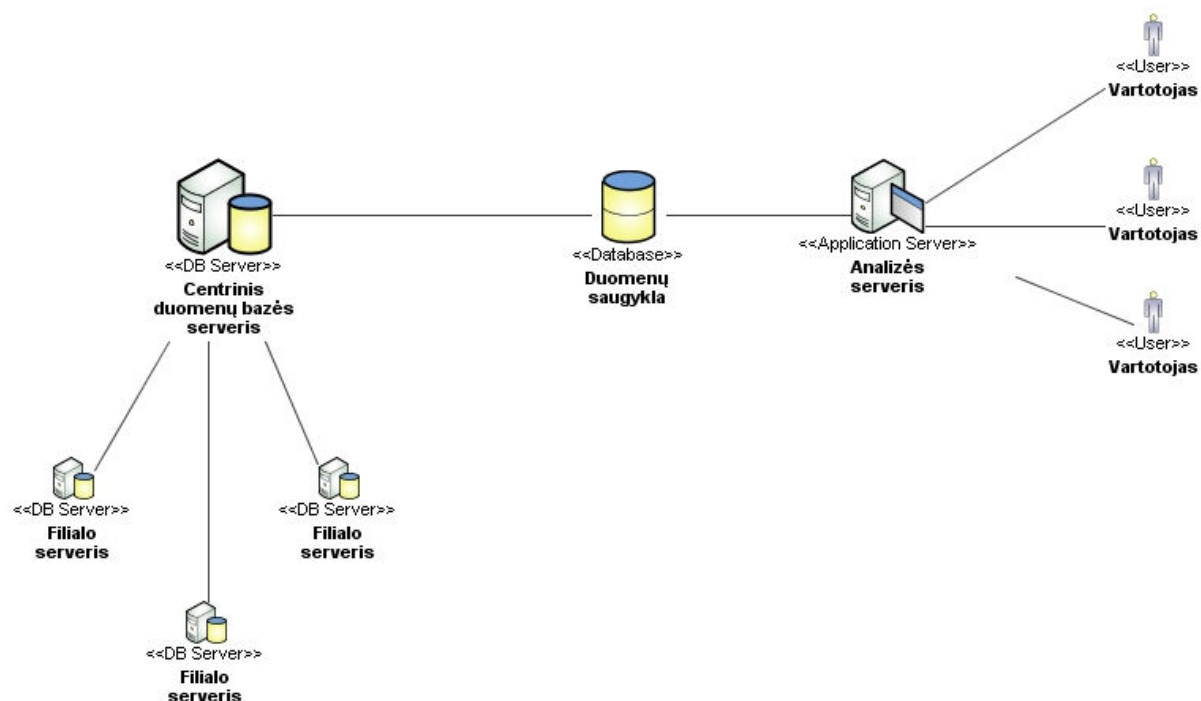
Priešingą politiką bandomųjų sistemų versijų atžvilgiu vykdo Oracle kompanija. Šios firmos produktus galima parsisiųsti tiesiai iš kompanijos tinklapio, juos įdiegti ir naudoti vieno sistemos prototipo kūrimui. Kompanijos svetainėje pateikiami veikiantys programinės įrangos paketai, be jokių laiko ar funkcionalumo apribojimų.

Dar vienas aspektas, kuris apibūdina MS SQL ir Oracle sistemas – tai jų diegimas. Bendrai yra sutinkama, kad MS SQL serveris ir jo komponentai yra žymiai lengviau įdiegiami, nei Oracle siūlomi produktai. Tai lemia keletas priežasčių:

- MS SQL veikia tik MS Windows operacinių sistemų šeimos operacinėse sistemose
- Operacinės bei duomenų bazės sistemų gamintojas yra ta pati kompanija
- Paprastesnis diegimo procesas (vartotojui pateikiama mažiau konfigūravimo alternatyvų)

2.6. Architektūros ir galimų įgyvendinimo priemonių variantų analizė

Verslo valdymo ir apskaitos sistema TSM yra sukurta, panaudojant Oracle duomenų bazių valdymo sistemą. Ši sistema yra įdiegta tiek centriniam įmonės padalinyje, tiek filialuose. Bendruoju atveju, sistemos architektūra atrodytų taip, kaip pavaizduota 8 paveikslėlyje.

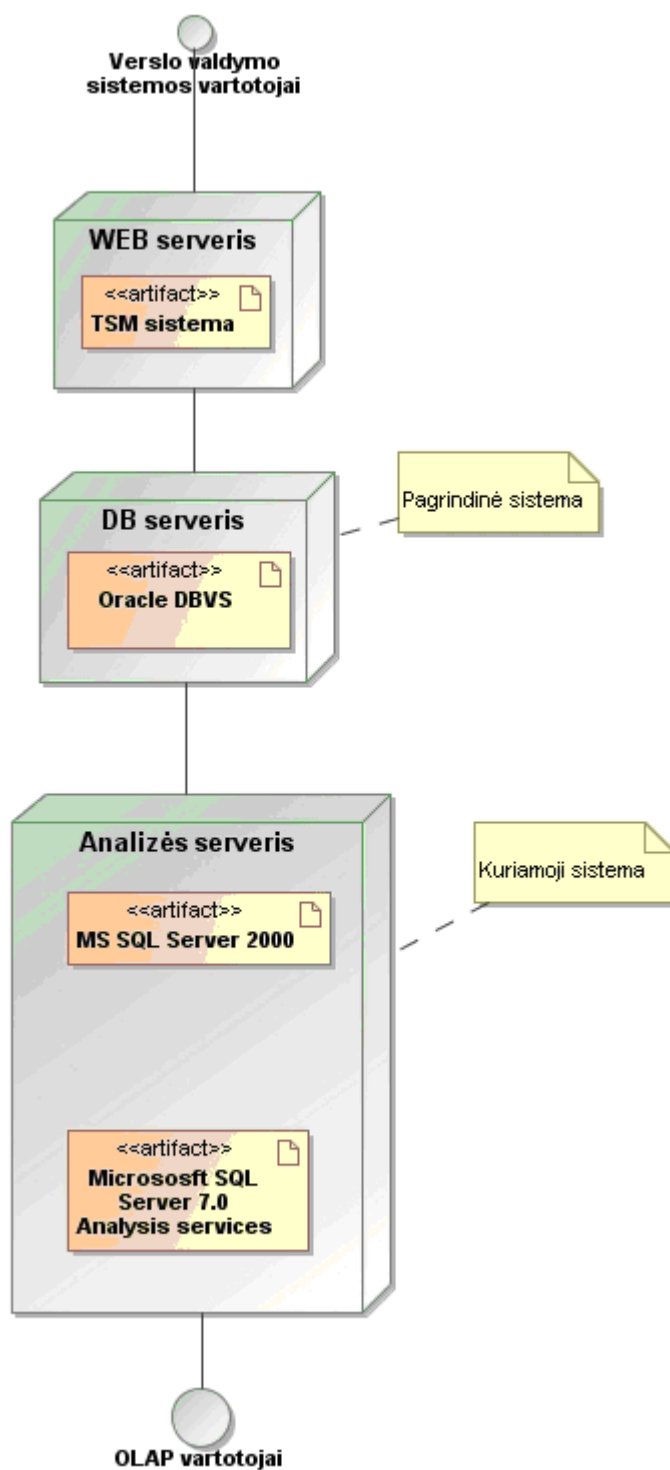


8 pav. Bendra sistemos architektūra

Kiekviename įmonės padalinyje veikia po atskirą duomenų bazės serverį. Centrinėje duomenų bazėje yra kaupiami visų padalinių duomenys. Jie yra sinchronizuojami su filialų duomenų bazėmis replikacijų pagalba.

Centrinėje duomenų bazėje esantys duomenys yra naudojami įvairioms analizėms, ataskaitoms. Šie duomenys įvairių vaizdų (angl. views), sąryšių bei duomenų transformacijų pagalba yra transformuojami į reikiamas struktūras ir patalpinami OLAP duomenų saugykloje, kur vėliau jie naudojami analizei.

Prijungus kuriamąją sistemą prie esančios įmonėje, bendrą sistemos diegimo diagramą galia pavaizduoti taip, kaip 9 paveikslėlyje.



9 pav. Sistemos diegimo diagrama

TSM verslo valdymo sistema yra sukurta Oracle duomenų bazės pagrindu. Vartotojo sąsajai yra pasirinktas interneto naršyklės sprendimas. Tad yra atskiras WEB serveris, kuriame veikia pati TSM valdymo sistema, WEB serveris, skirtas vartotojo sąsajai palaikyti. Duomenų

bazei skirtas atskiras DB serveris. Jame veikia Oracle duomenų bazė, kurioje talpinama visa verslo veiklos informacija.

2.7. Darbo tikslas ir siekiami privalumai

Diegiamai sistemai yra nustatyta keletas tikslų:

1. Sumažinti pagrindinio serverio apkrovimą, atminties naudojimą, tuo padidinant pagrindinės sistemos efektyvumą.
2. Sukurti lengvesnes duomenų analizės priemones, skirtas analizuoti verslo valdymo sistemos duomenis, tuo palengvinant sistemos vartotojų darbą.
3. Efektyviau išnaudoti sistemos darbą nakties metu.
4. Suteikti atitinkamiems vartotojams papildomos informacijos apie įmonės veiklą.
5. Tarnauti kaip atitinkamų ataskaitų rengimo priemonė.

Įmonėje įdiegus naująją analizės sistemą, ji turėtų žymiai palengvinti įmonėje vykstančius pardavimų planavimo bei sprendimų priėmimo procesus. Įdiegus šią sistemą ir pasiteisinus jai, ji būtų pritaikoma ir kitiems įmonėje vykstantiems procesams.

Norint pasiekti sistemai iškeltus tikslus, būtina atlikti šias užduotis:

1. Sukurti atskiros duomenų bazės schemą, atitinkančią duomenų analizei keliamus reikalavimus.
2. Sukurti duomenų integravimo priemones, skirtas duomenų paėmimui iš centrinio duomenų bazės serverio bei jų transformavimui į atskirą duomenų bazės serverį.
3. Sukurti įmonės klientų analizės poreikius tenkinantį duomenų kubą.
4. Išbandyti ir ištestuoti sistemos veikimą.
5. Sudaryti duomenų saugyklų kūrimo bei tobulinimo metodiką naujų įmonės veiklos procesų analizei.

2.8. Reikalavimai duomenims

Kuriant naująją sistemą, laikoma, kad už duomenų pilnumą, tikslumą ir korektiškumą atsako verslo valdymo sistema TSM. Taip yra todėl, kad šias sąlygas būtina išpildyti, siekiant užtikrinti bendrą sėkmingą kasdieninį visos sistemos veikimą. Tad atskirų reikalavimų naujosios sistemos duomenims nėra keliami.

Duomenis į analizės duomenų saugyklą numatyta perkelti kiekvieną naktį. Dėl šios priežasties, OLAP priemonėmis matomi duomenys nebus patys naujausi. Tai verčia daryti TSM verslo valdymo sistemos apkrovimas – tik naktį jis yra mažesnis. Duomenų perkėlimas dieną į

analizės duomenų saugyklą reikalautų nemažai serverio resursų, o tai mažintų bendrą sistemos darbo našumą, juolab, kad analizės duomenų naujumas nėra toks aktualus – pakanka juos atnaujinti vieną kartą per parą. Todėl pasirinktas laikas yra naktis, kuomet serverių apkrovimas yra mažiausias.

Dar vienas reikalavimas duomenims – tai, jog OLAP priemonėmis išgauti duomenys būtų identiški TSM sistemos priemonėmis (ataskaitomis, formomis) išgaunamiems duomenims:

1. Duomenys atitinkamai datai sutaptų su TSM valdymo sistemos duomenimis, išgautais įvairių ataskaitų ir kitų priemonių pagalba.
2. Išliktų lietuviškos raidės,.
3. OLAP priemonėmis išgautuose duomenyse būtų matomi pavadinimai, o ne vien kodai (pvz.: prkių grupių pavadinimai, o ne jų kodai ir pan.).

2.9. Nefunkciniai reikalavimai ir apribojimai

2.9.1. Reikalavimai veikimui

Vienas svarbiausių naujai sistemai keliamų reikalavimų jos veikimui - tai sistemos našumas. Svarbu, kad naujoji sistema pernelyg neapkrautų TSM verslo valdymo sistemos serverio tiek duomenų atnaujinimo, tiek analizės kubų generavimo metu. Tai vienas svarbiausių reikalavimų, keliamų naujai sistemai.

Pagrindinės sistemos apkrovimo problemą galima spręsti keliais būdais:

1. Mažinant analizuojamų duomenų kiekį, naudojant tik tai tam tikro intervalo (pvz.: laiko intervalo) duomenis. Tokio sprendimo trūkumas – tai suprastėjanti duomenų saugykla (joje saugomi nepilni duomenys).
2. Perkelti duomenis į atskirą, vien tik analizei skirtą serverį. Tokio sprendimo privalumas tas, jog replikavimo pagalba, duomenys būtų perkeltami į atskirą OLAP duomenų bazės serverį (duomenų replikavimas reikalauja žymiai mažiau pagrindinės sistemos resursų, nei jų analizė pagrindiniame serveryje)

Naujai sistemai kurti pasirinktas antrasis variantas. Replikacijų pagalba, pagrindinės sistemos duomenys bus perkeltami į atskirą OLAP duomenų bazės serverį, kur vėliau jie galės būti apdorojami, skaičiuojamos reikalingos reikšmės ir pan. Tokį pasirinkimą lėmė pakankamai opus centrinio duomenų bazės serverio resursų klausimas. Tad pasirinktas sprendimas leis minimaliai apkrauti centrinę sistemą.

Taip pat svarbus reikalavimas, keliamas naujai sistemai – tai kubų pjūvių ir reikšmių generavimo greitis. Kadangi papildomas, analizei skirtas duomenų bazės serveris yra gerokai

silpnesnis nei pagrindinis duomenų bazės serveris, svarbu tinkamai išnaudoti jo resursus. Kubų pjūvių generavimo greitį galima padidinti, pasitelkiant įvairius sprendimus:

1. Mažinant analizuojamų duomenų kiekį, naudojant tikrai tam tikro intervalo (pvz.: laiko intervalo) duomenis. Tokio sprendimo trūkumas – tai suprastėjanti duomenų saugykla (joje saugomi nepilni duomenys).
2. Iš anksto apskaičiuojant reikalingas skaičiuojamasias reikšmes ir padaryti pakankamai pradinių apskaičiavimų (angl. aggregations).
3. Naudoti optimizacijas pagal panaudojimą (ang. *usage based optimizations*).

Naujosios sistemos skūrimui, pasirinktas duomenų bazės perkėlimas į atskirą analizės serverį. Tai labiausiai lėmė „labai brangūs“ centrinio duomenų bazės serverio resursai. Kadangi papildomas analizės duomenų bazės serveris yra gerokai silpnesnis, siekiant užtikrinti kuo greitesnį kubų pjūvių generavimą, šiam procesui paspartinti bus pasitelktos visos anksčiau išvardintos priemonės. Vėliau, pasitvirtinus naujosios sistemos naudingumui ir jos teikiamiems analizės privalumams, bus galima pasitelkti našesnę ir geresnę techninę įrangą, skirtą duomenų analizei.

2.9.2. Reikalavimai sąveikai/suderinamumui su kitomis sistemomis

Nemažiau svarbūs yra naujosios sistemos integralumo klausimai. Ypač svarbu, kad naujai sukurta analizės sistema kuo labiau integruotųsi į esamą įmonės informacinę bei verslo valdymo sistemas. Naujoji sistema turėtų dirbti be didelių sistemų administratorių pastangų, tai yra praktiškai automatiškai ir be įsikišimų. Tad naujosios analizės sistemos darbui užtikrinti, yra svarbu, kad:

1. Būtų nuolatinis ryšys tarp OLAP ir analizės duomenų bazės serverio, o šio serverio ryšys trėtų būti nuolat palaikomas su centriniu įmonės duomenų bazės serveriu.
2. Svarbu užtikrinti OLAP analizės kubų atnaujinimą kiekvieną naktį (išskyrus savaitgalius, kuomet tai nėra reikalinga).
3. OLA analizės serveris būtų integruotas su įmonėje esančia Microsoft Windows Active Directory vartotojų autentifikavimo sistema.

Visų sistemų integracija yra svarbi, siekiant užtikrinti efektyvų ir nenutrūkstantį įmonės sistemų darbą. Tad naujosios analizės sistemos integracijai su kitomis įmonėje esančiomis sistemomis reikia skirti nemažai dėmesio ir stengtis užtikrinti kiek įmanoma geresnę visų įmonės sistemų integraciją.

2.9.3. Kiti reikalavimai (patikimumui, saugumui, plėtimui)

Be efektyvumo, integracijos su kitomis sistemomis, nemažiau svarbus yra duomenų saugumo klausimas, naudojant informacines sistemas. Tai yra vienas kritinių apskėtų, kuriuos turi užtikrinti įmonėje veikiančios sistemos. Yra ypač svarbu užtikrinti, kad vartotojai, kuriems skirta informacija ir duomenys, galėtų ją pasiekti, o tie vartotojai, kuriems ji nėra reikalinga, prieš šių išteklių priėjimo neturėtų. Kalbant apie kuriamąją analizės sistemą, tai galima užtikrinti Microsoft SQL Server Analysis Services esančiomis vartotojų autorizavimo priemonėmis. Tai galima pasiekti, skirtingoms Microsoft Active Directory vartotojų grupėms suteikiant atitinkamas priėjimo prie sistemos galimybes.

Dar vienas svarbus klausimas – tai sistemos patikimumas. Tai liečia tiek sistemoje naudojamų duomenų, tiek pačios sistemos veikimo patikimumą. Būtina pasiekti, jog sistema dirbtų be sutrikimų ir kad vartotojas, kuriam reikalinga atitinkama informacija, ją gautų greitai ir kad ji būtų patikima. Sistemos patikimumas – tai svarbus rodiklis, kuris leidžia lengviau „adaptuoti“ naujas sistemas į įmonės veiklos procesus, įgyti jos vartotojų pripažinimą. Nepatikimos sistemos dažniausiai nustoja naudotis ir galiausiai visiškai pašalinamos iš įmonės veiklos procesų.

Taip pat svarbu, jog kuriama analizės sistema galėtų būti plečiama. Konkrečiu atveju, kuriama sistema apims įmonės klientų analizę. Atsiradus poreikiui, ją bus galima modifikuoti, išplėsti, pridėdant papildomo funkcionalumo, suprojektuojant naujus analizės kubus bei kitais būdais.

2.10. Rizikos faktorių analizė

Siekiant užtikrinti efektyvų sistemos darbą, užtikrinti visalaikį jos pasiekiamumą, būtina įvertinti galimas rizikas ir numatyti, kaip šias rizikas suvaldyti bei kiek įmanoma sumažinti. Tik numačius galimas rizikas ir jų valdymo būdus, galima pasiekti efektyvų ir patikimą sistemos darbą.

Vienas patogiausių būdų nustatyti galimas rizikas ir jas maksimaliai sumažinti – tai sudaryti rizikos faktorių lentelę.

4. Lentelė. Sistemos rizikos faktorių analizė

Rizikos faktorius	Rizikos laipsnis	Rizikos faktoriaus mažinimo priemonės
Analizės serverio darbo sutrikimas	Kritinis	Sutrikus analizės serverio (OLAP) darbui, sistema tampa nepasiekiamą. Todėl svarbu užtikrinti nenutrūkstamą

		energijos tiekimą. Taip pat svarbu, kad analizės serveris būtų nesunkiai pasiekiamas ir fiziškai – kad sugedus atitinkamiems serverio techninės įrangos komponentams, juos būtų galima pakeisti ir atstatyti serverio funkcionalumą.
Duomenų nekorektiškumas	Aukštas	Būtina užtikrinti, kad analizei skirti duomenys būtų kiek įmanoma tikslesni. Todėl prieš atiduodant sistemą naudojimui, būtina ją visapusiškai ištestuoti. Taip pat numatyti galimas problemas, kurios iškiltų esant nekorektiškiems duomenims ir turėti planus, kaip tas problemas pašalinti.
Nepavykęs automatinis duomenų atnaujinimas	Vidutinis	Būtina užtikrinti, kad duomenų atnaujinimas vyktų skaldžiai ir duomenys nuolat būtų naujausi. Tam tikslui galima įdiegti pranešimų sistemą, kuri informuotų apie pavykusį/nepavykusį duomenų atnaujinimą. Visgi, vienos dienos duomenų atnaujinimas nebūtų toks kritinis, kuomet analizuojami mėnesio ar metų pardavimų duomenys.
Norimų analizės funkcijų nebuvimas	Žemas	Visus analizės pūvius bei funkcijas su būsimais sistemos vartotojais būtina suderinti iš anksto, numatyti jų įgyvendinimo galimybes ir būdus.
Sistemos lūžis	Žemas	Nuolat (kasdien) darant atsargines duomenų kopijas, sistemos lūžio riziką galima sumažinti iki minimumo.

2.11. Rezultato kokybės kriterijai

Pagrindinis projekto kokybės kriterijus – tai sukurtos sistemos funkcionalumas. Tad įvertinus visus būsimai sistemai keliamus funkcinius ir nefunkcinius reikalavimus, kurią sistemą galima įvertinti kokybės kriterijais, išvardintais 5 lentelėje.

5. Lentelė. Sistemos kokybės kriterijai

Kokybės kriterijus	Komentarai
Integracija	Ar pilnai naujoji sistema integruojasi į įmonėje veikiančias sistemas ir procesus?
Stabilumas	Kaip stabiliai veikia sistema? Ar korektiškai veikia sistemos duomenų atnaujinimas? Kiek buvo sistemos darbo sutrikimų ?
Našumas	Kaip greitai dirba sistema? Kiek trunka duomenų atnaujinimas, jų apdorojimas?
Patikimumas	Ar patikimi ir korektiški sistemos pateikiami duomenys? Ar numatyti visi kritiniai atvejai? Kaip dažnai pasitaiko sistemos klaidų, kurias tenka taisyti?
Efektyvumas	Ar sistema dirba efektyviai? Ar pagerėjo ir pagreitėjo darbo kokybė? Ar naujos sistemos įdiegimas netrukdo ir nestabdo esamų įmonės sistemų darbo?
Patogumas	Ar sistema yra patogi vartotojams? Ar tolerantiška klaidoms? Ar lengvai įsissavinama nepatyrusiems vartotojams? Kiek vidutiniškai vartotojui trunka išmokti naudotis sistema?
Išsamumas	Ar įgyvendinti visi funkciniai ir nefunkciniai sistemai keliami reikalavimai? Ar numatyti visi atvejai? Ar išpildyti visi vartotojų pagedavimai? Ar vartotojai patenkinti sistemos darbu?

Atsižvelgus į visus šiuos kriterijus, galima bus įvertinti sistemos kokybę bei naudą įmonei, įvertinti, ar projektas sėkmingas. Pasiteisinus, jį galima plėsti, didinti sistemos funkcionalumą, sukurti naujas analizės galimybes, naujus duomenų analizės kubus.

2.12. Analizės kubų projektavimo metodikų analizė

Duomenų saugyklos projektavimas yra gana sudėtinga užduotis. Norint pasiekti gerų rezultatų, būtina naudoti sisteminius metodus. Yra keli pagrindiniai duomenų saugyklų kūrimo ir projektavimo žingsniai, kurie yra atliekami, nepriklausomai nuo duomenų saugyklos kūrimo metodikos.

2.12.1. Verslo poreikių surinkimas ir analizė

Šis pirminis etapas įtakoja beveik visus tolimesnius etapus duomenų saugyklų kūrimo procese. Šiame žingsnyje, duomenų saugyklos projektuotojai turi ne tikrai išgauti iš būsimų sistemos vartotojų, kokius duomenis jie nori matyti, kokiose darbo vietose sistema bus naudojama, kokie sprendimais bus priimami duomenų sandėlio pagalba. Yra išvystyta keletas reikalavimų analizės technikų, kuriose naudojami interviu su būsimais sistemos vartotojais, Jungtinis projektavimas (angl. JAD) standartinių šablonų naudojimą, reikalavimų grupavimą pagal jų svarbą, egzistuojančių dokumentų apžvalgą. Interviu su būsimos sistemos vartotojais ir jungtinis projektavimas (JAD) – tai dvi pagrindinės technikos išgauti reikalavimus iš vartotojų.

2.12.2. Duomenų modeliavimas

Yra trys duomenų modeliavimo lygmeys: konceptualus projektavimas, loginis projektavimas, ir fizinis projektavimas. Konceptualus projektavimas pateikia duomenų sandėlio reikalavimus duomenims aukštame lygyje, „neprisirišant“ prie jokios konkrečios duomenų bazių valdymo sistemos (DBVS). Šiame etape tam paprastai asitelkiamos ER arba duomenų klasių diagramos. Loginio projektavimo metu, konceptuali schema yra paverčiama į loginę schemą, kur panaudojama konkreti duomenų bazių valdymo sistema. Populiariausi duomenų sandėlio loginiai modeliai yra santykinis ir matuojamasis modelis. Trečiajame, fizinio projektavimo etape, loginė schema yra konvertuojama į fizinę, kuri jau nurodo, kaip duomenų bazė yra organizuojama fiziniame atmintyje.

Skiriami keturi pagrindiniai duomenų saugyklų aplinkos elementai:

1. Duomenų šaltinių sistemos (nutolusios duomenų bazės, struktūrizuoti failai, taikomosios programos ir pan.).
2. Duomenų gavimo, transformavimo ir pakrovimo (angl Extract, Transform and Load – ETL) sistemos.
3. Pačios duomenų saugyklos.
4. Ataskaitų pateikimo sistemos ir portalai.

2.12.3. Duomenų šaltiniai

Kasdien verslo valdymo sistemos sukaupia didžiulius kiekius duomenų. Verslo valdymo ir kitų įmonėse naudojamų sistemų yra įvairiausių ir sukurtų įvairiomis technologijomis. Vienos jų duomenis saugo reliacinėse duomenų bazėse, kitos failuose ir pan.

Tad projektuojant duomenų saugyklą, būtina išanalizuoti esamus duomenų šaltinius, iš kurių informaciją reikalinga pasiimti, kad ji vėliau būtų apdorojama ir analizuojama duomenų analizės sistemose (pavyzdžiui, OLAP).

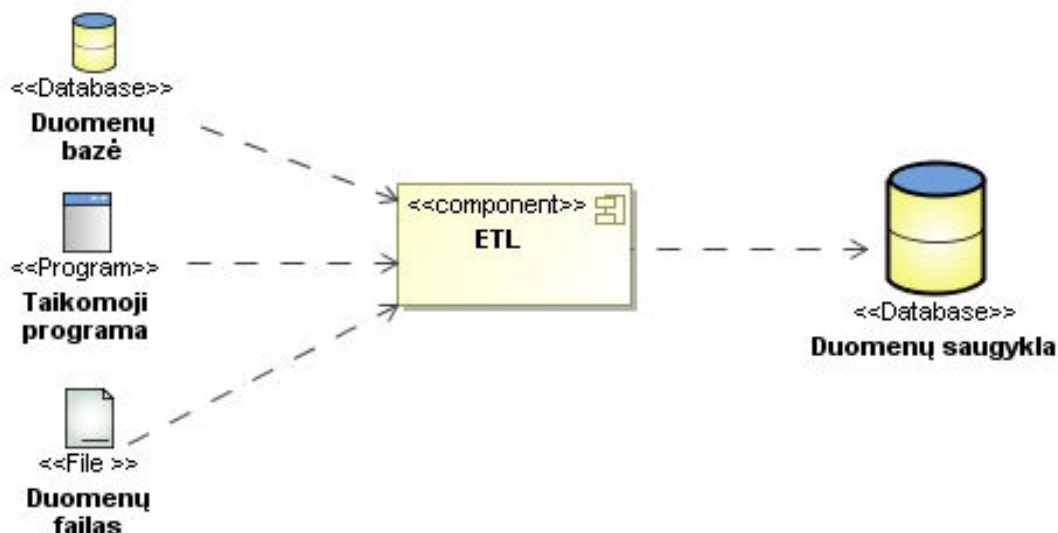
2.12.4. ETL projektavimas

Informacija, saugoma duomenų saugykloje, surenkama iš operacinių taikomųjų sistemų, kurios negali būti tiesiogiai naudojamos kaip strateginės informacijos šaltiniai. Rinkinys priemonių, skirtų išgauti, transformuoti ir įkelti duomenis iš operacinių sistemų į duomenų saugyklas, yra vadinamas ETL [4], [5] (angl. Extract, Transform and Load).

ETL projektavimas yra vienas svarbiausių etapų, siekiant užtikrinti efektyvų ir patikimą duomenų sandėlio naudojimą. Tik gerai suprojektavus ETL modelius, duomenų sandėlyje saugoma informacija bus naujausia ir patikima.

Paprastai, ETL procesas yra sudarytas iš trijų etapų:

1. Reikalingų duomenų atrinkimo (angl. Extract). Šie duomenys gali būti skirtingose duomenų bazėse esančios lentelės, duomenų failai ar kita struktūrizuota informacija. Šio proceso metu duomenys yra išgaunami iš kitų duomenų šaltinių (tai gali būti OLTP, liktinės (angl. legacy) sistemos, centrinės duomenų bazės, duomenų failai, ODBC ir pan.).
2. Antrajame etape (angl. Transform), reikalingi duomenys yra nuskaitomi, sutvarkomi (agreguojami, apskaičiuojamos reikalingos tarpinės reikšmės ar pan.). Šio proceso metu išgauti duomenys yra paverčiami į nuoseklią duomenų saugyklos schemą, taikant verslo taisykles. Taip pat šis procesas yra atsakingas už išgautų duomenų patvirtinimą, tų duomenų tikslumą, tipo pakeitimą. Tai vienas sudėtingiausių ETL etapų.
3. Trečiasis etapas – tai duomenų įkėlimas į duomenų saugyklą (angl. Load). Šio proceso metu, atrinkti ir sutvarkyti duomenys išsaugomi naujoje duomenų bazėje.



10 pav. ETL procesas

Projektuojant ETL, galimi du keliai: tai kaskart atnaujinant duomenis, nuskaityti visus duomenis, juos sutvarkyti ir įkelti į duomenų saugyklą. Antrasis būdas – tai CDC (angl Change Data Capture) metodų taikymas [16], [18].

Pirmasis būdas (visus duomenis nuskaityti, apdoroti ir įkelti į duomenų saugyklą reikalauja daug resursų. Kuomet duomenų yra nedaug, šis metodas pasiteisina (nes yra pats paprasčiausias). Tačiau kuomet duomenų yra daug, šis būdas yra netinkamas – stipriai apkrauna sistemas. O jeigu norima nuolat matyti naujausią informaciją, tai naudojant šį metodą, to padaryti praktiškai neįmanoma (nes reikia nemažai laiko, kol duomenys nuskaityti, sutvarkomi ir įkeliami į duomenų saugyklą). Antrasis (CDC metodų taikymas) būdas – tai daug racionalesnis, tačiau kartu ir gerokai sudėtingesnis.

2.12.5. Pasikeitusių duomenų nustatymo (CDC) metodas

Yra skiriami dviejų tipų pasikeitusių duomenų nustatymo (angl. CDC) metodai – tai sinchroniniai ir asinchroniniai. Asinchroniniai savo ruožtu skirstomi į keturis metodus:

- Hotlog
- Distributed hotlog
- Autolog archived
- Autolog online

Efektyvus pasikeitusių duomenų surinkimas vis dar yra nemažas iššūkis projektuojant duomenų saugyklas. Čia išgaunant duomenis iš duomenų šaltinių, yra reikalingi tik tie

duomenys, kurie buvo pakeisti ar pridėti. Žiūrint iš koncepcinės pusės, yra du būdai pastebėti duomenų pakitimus: taikomųjų programų lygyje arba duomenų lygyje.

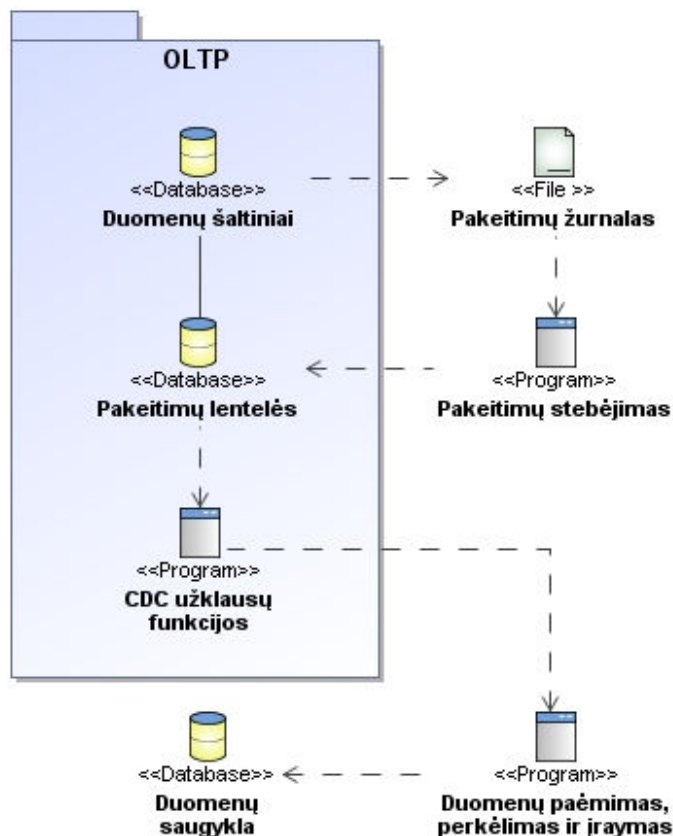
Pasikeitusių duomenų atrinkimo mechanizmai lengvina duomenų integraciją, ją daro efektyvesnę ir greičiau atliekamą. CDC leidžia:

- Atnaujinti duomenis, sistemoms dirbant įprastu režimu. Tai leidžia atnaujinti duomenis, tam visiškai netrukdam kitų sistemų darbo ir nereikalaujant sustabdyti kitų vartotojų darbą su sistema. Tai taupo laiką ir palengvina duomenų integraciją.
- Turėti pačius naujausius duomenis. CDC pagalba galima gauti naujausius duomenis dažniau, netgi realiu laiku – taip savalaikie informacija aprūpinant vartotojus ir kitus darbuotojus, priimančius sprendimus.
- Taupyti kaštus. Atrenkant tiksliai pakitusius duomenis, CDC reikia žymiai mažiau resursų, nei įprastais ETL metodais. Tokiu būdu mažinami kaštai techninės, programinės įrangos, taip pat žmoniškųjų resursų sąskaita.

CDC sprendimai yra suprojektuoti taip kad maksimizuotų ETL efektyvumą, mažinant išteklių vartojimą tiksliai iki pasikeitusių duomenų kopijavimo/transformavimo (t.y. iki delta pokyčio). Pasikeitusių duomenų nustatymo (CDC) metodai yra pagrįsti šiais komponentais:

- **Duomenų pasikeitimo stebėjimo posistemės.** Tai programinės įrangos komponentai, atsakingi už pasikeitusių duomenų identifikavimą duomenų šaltiniuose. Šios posistemės gali būti optimizuotos arba specialiai sukurtos duomenų šaltiniams (t.y. naudoti duomenų bazės žurnalus, triggeriu, arba naudoti bendrus metodus, tokius kaip duomenų pakeitimo žurnalų palyginimus.
- **Pasikeitusių duomenų paslaugos (servisai).** Šie servisai pateikia funkcijas, leidžiančias duomenis filtruoti (pavyzdžiui, atrinkti tiksliai į duomenų bazę įrašytus duomenis), rūšiuoti ir pan.
- **Pasikeitimų įrašymas.** Šie mechanizmai yra atsakingi už patikimą pakeistų duomenų išsaugojimą.

Bendruoju atveju, CDC metodą galima pavaizduoti grafiškai (žemiau esančiame paveikslėlyje).

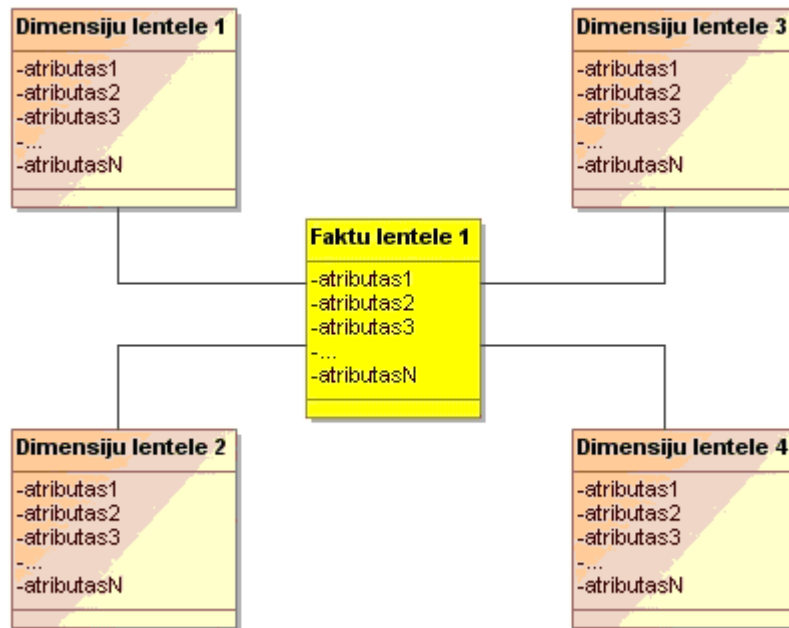


11 pav. CDC metodas

2.12.6. Duomenų saugyklų modeliai

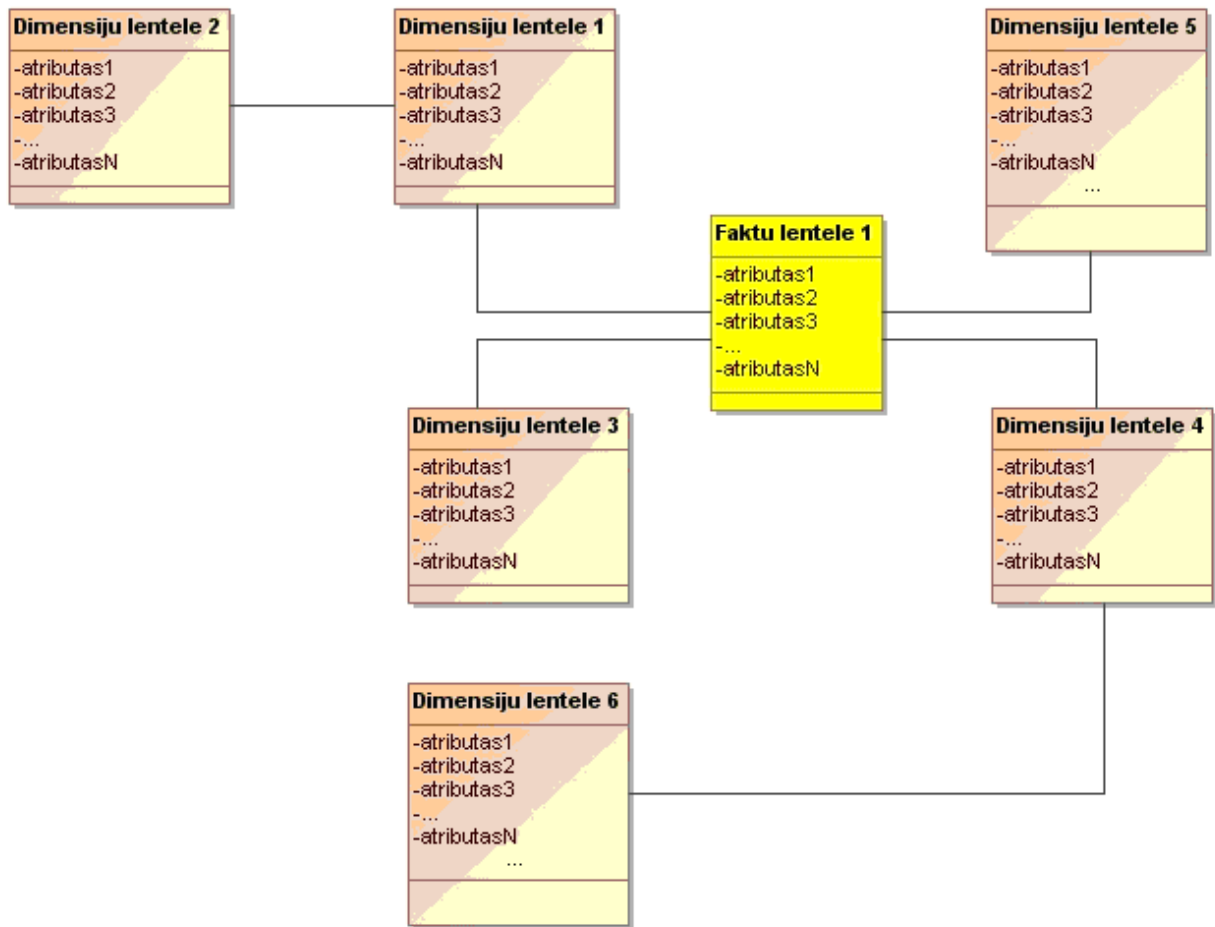
Norint užtikrinti efektyvų ir patikimą darbą su duomenų saugykla, būtina ją tinkamai suprojektuoti [15], [16]. Yra skiriamos trys pagrindinės duomenų saugyklų schemas:

1. **Žvaigždės schema (angl. Star schema).** Žvaigždės tipo duomenų saugyklos modelį sudaro faktų ir dimensijų lentelės. Faktų lentelėse saugoma kiekybiniai ir faktiniai verslo duomenys. Ši informacija dažnai yra sudaryta iš skaitmenų, gali susidėti iš daugelio skilčių ir milijonų ar net milijardų eilių. Dimensijų lentelės paprastai yra mažesnės ir jose saugomi tekstiniai duomenys, apibūdinantys dimensijas. SQL užklausos, kad atrinktų reikalingą informaciją, naudoja sujungimus tarp faktų ir dimensijų lentelių, duomenų apribojimus. Vykdomo sparta yra svarbus bet kokios schemas parametras, ypač sprendimų paramos sistemose, nes čia dažniausiai operuojama su didžiuliais duomenų kiekiais.



12 pav. Žvaigždės schema

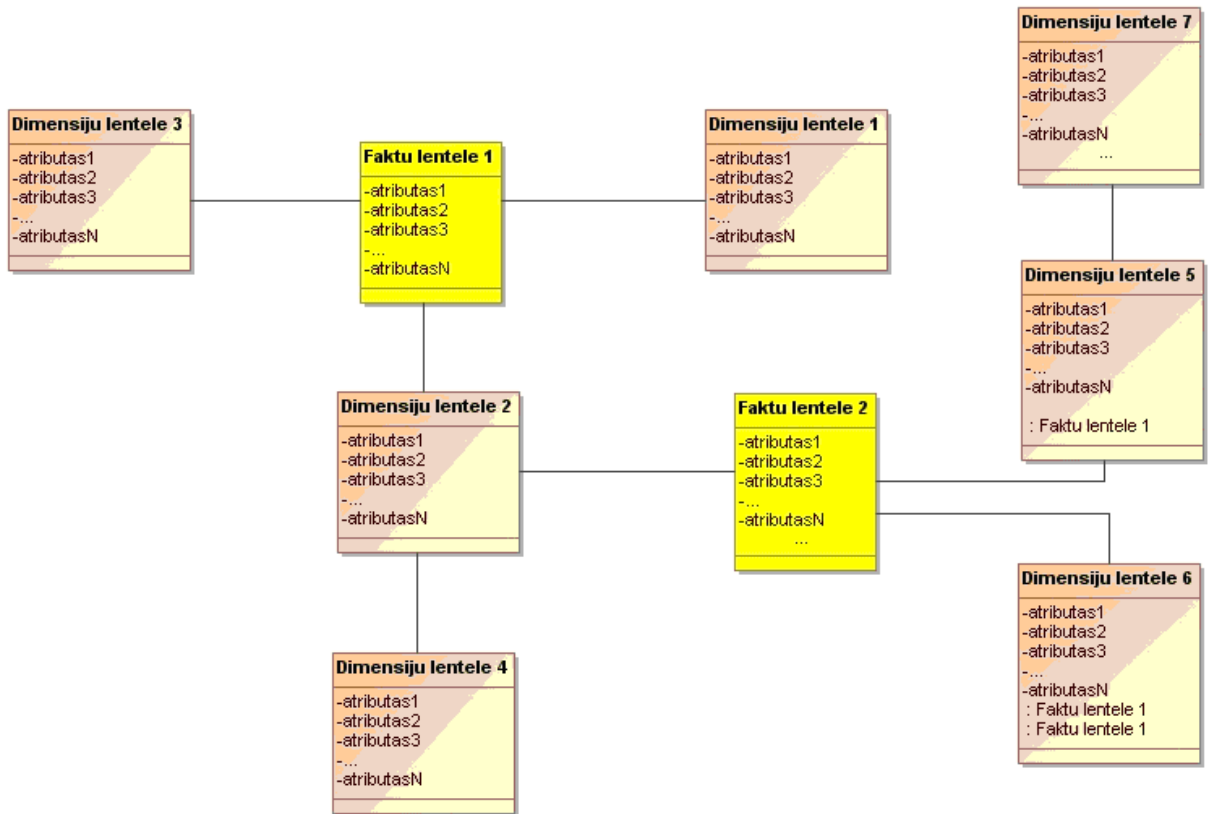
2. **Snaigės schema.** Ši duomenų saugyklos architektūra, iš esmės, yra sudėtingesnis, žvaigždės schemas variantas. Svarbiausias skirtumas snaigės schemoje yra tai, jog dimensijų lentelės šioje schemoje yra normalizuotos, todėl jos turi tipišką reliacinės duomenų bazės struktūrą. Snaigės schemas paprastai yra naudojamos tada, kai dimensijų lentelė tampa labai didelė ir kai žvaigždės schema negali būti naudojama dėl duomenų struktūrų sudėtingumo.



13 pav. Snaigės schema

3. **Žvaigždžių klasteris (galaktikos schema).** Kiekvienai žvaigždės ar snaigės schemai galima suprojektuoti žvaigždžių klasterio schemą. Ši schema yra gerokai sudėtingesnė, nei žvaigždė ar snaigė, kadangi ją sudaro keletas faktų lentelių. Tai leidžia dimensijų lentelėms būti bendroms tarp keleto faktų lentelių. Toks duomenų saugyklos projekto sprendimas yra labai lankstus, tačiau toką duomenų saugyklą gali būti sunku valdyti ir palaikyti.

Didžiausias žvaigždžių klasterio schemas trūkumas yra sudėtingas projektas, nes turi būti apsvarstyta daug duomenų grupavimo variantų.



14 pav. Žvaigždžių klasterio schema

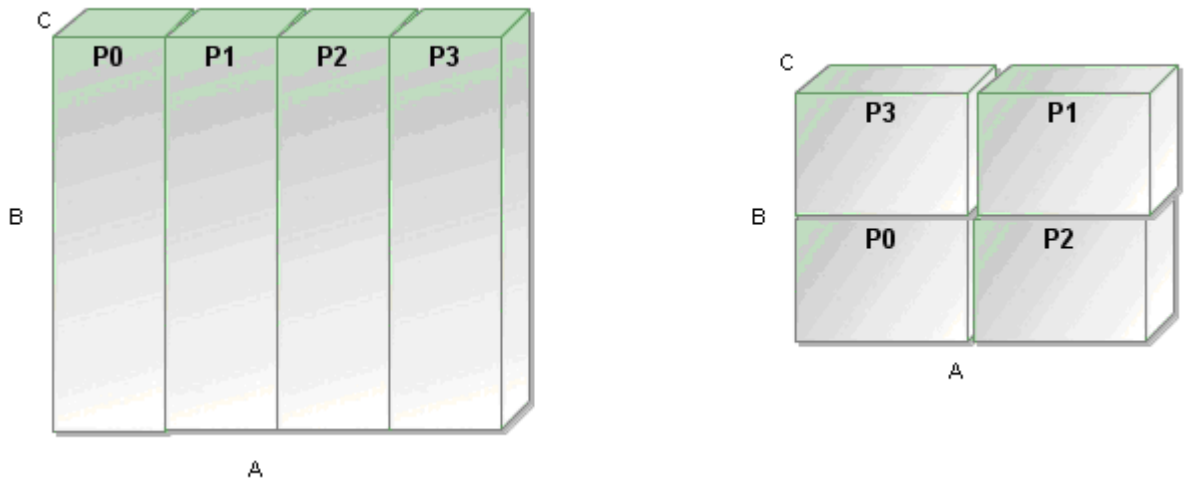
Kiekviena duomenų saugyklos schema turi savų privalumų ir savų trūkumų. Tad svarbu parinkti tinkama duomenų saugyklos struktūrą, kad būtų pasiekti optimaliausi našumo, patikimumo, paprastumo rodikliai.

2.13. Analizės kubų veikimo ir atnaujinimo optimizavimas

Projektuojant duomenų saugyklą, yra labai svarbu atsižvelgti į kuriamų duomenų saugyklų veikimo spartą ir galimus optimizavimo būdus. Tik pakankamai optimizavus duomenų saugyklą, jos veiks efektyviai, mažiau reikalaus pagrindinės sistemos resursų. Galima išskirti šiuos pagrindinius analizės kubų veikimo ir atnaujinimo optimizavimo būdus:

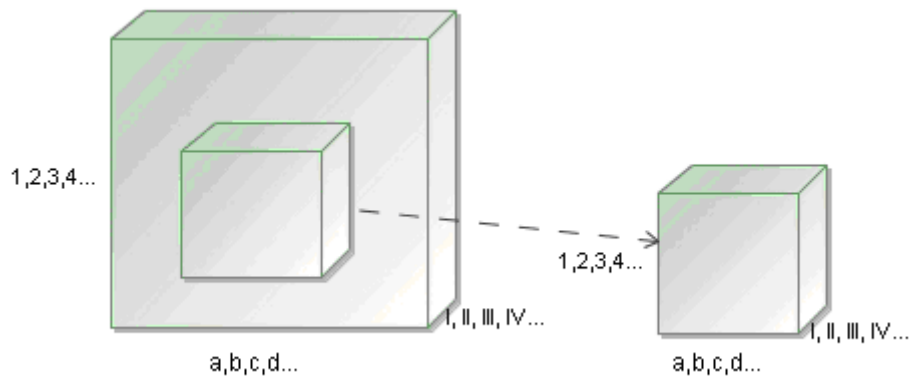
- 1. Duomenų padalijimas (angl partitioning).** Tai leidžia dideles duomenų lenteles ir indeksus padalinti į mažesnes dalis. Padalijus duomenis į mažesnes dalis, yra lengviau ir greičiau daromos atsarginės duomenų kopijos, duomenų įkėlimas ir jų vykdymas. Duomenų padalijimas padidina sistemos našumą bei duomenų pasiekiamumą. Lentelė gali būti padalinta pagal bet kuri tos lentelės stulpelį. Tad suskaldžius duomenų kubą į nepriklausomus skirsnius (ang.

partitions), galima atlikti keletą skaičiavimų vienu metu – tai ypač padidina našumą multiprocessorinėse sistemose.



15 pav. Duomenų padalijimas

- 2. Dalinių kubų (angl. partial cubes) panaudojimas.** Šis kubų veikimo optimizavimo būdas gali būti panaudotas tik to paties analizės serverio kubų apdorojimo algoritmuose. Dalinių kubų panaudojimas leidžia atnaujinti informaciją, apdorojant tik reikalingus kubus, o tų kubų kurių nereikia, nekeičiant. Taip yra apdorojama tik reikalinga informacija – taigi, sumažinamos reikalingų apdoroti duomenų apimtys. – tad padidėja bendras sistemos našumas.



16 pav. Dalinių kubų panaudojimas

- 3. Agregacijų skaičiaus mažinimas.** Yra svarbu mažinti iki minimumo bendravimą tarp procesų – tai leidžia efektyviau išnaudoti daugiaprocesorines sistemas. Pilname duomenų kube, sąryšių skaičius yra eksponentinis dydis. Projektuojant duomenų kubą, yra svarbu kaip įmanoma sumažinti agregacijų

skaičių. Kadangi atnaujinant kubą, turi būti atnaujinamos ir agregacijos, tai kuomet jų yra pakankamai daug, kubo atnaujinimo laikas gali smarkiai padidėti.

Microsoft siūlo keletą patarimų, kaip optimizuoti analizės kubus, juos kuriant Microsoft Analysis Services programinės įrangos pagalba:

- **Optimalus agregacijų suprojektavimas.** Yra svarbu pabrėžti, jog didėjant kubui, agregacijų skaičius didėja eksponentiškai. Agregacijos – tai iš anksto paskaičiuotos tarpinės reikšmės kiekvienam kubo hierarchijos lygiui. Agregacijų kūrimui yra siūloma nemažai įvairių priemonių. Vienos jų tolygiai paskirsto užklausas pagal dimensijas, kitos leidžia kurti agregacijas, atsižvelgiant į kubų naudojimo istoriją ir pan.
- **Duomenų saugyklos tipo parinkimas.** Microsoft Analysis Services siūlo vieną iš trijų duomenų saugyklų tipų:

- **MOLAP (angl. multidimensional OLAP)** – tai tradicinis duomenų saugyklos tipas. MOLAP atveju, duomenys yra saugomi multidimensiniuose kubuose, o ne reliacinėje duomenų bazėje. Pagrindiniai MOLAP pranašumai – tai efektyvus užklausų atlikimas ir greitas duomenų atrinkimas. Visi skaičiavimai yra atliekami kubo sukūrimo metu ir panaudoti užklausų vykdymo metu. Šiuo atveju, duomenys dažniausiai saugomi tame pačiame serveryje, todėl yra labai greitai pasiekiami.

Pagrindiniai MOLAP trūkumai – tai ribotas duomenų kiekis, kuris gali būti saugomas. Taip pat MOLAP tipo saugyklos užima pakankamai daug vietos.

- **ROLAP (angl. relational OLAP)** – šiuo atveju visi duomenys yra sukaupiami reliacinėse duomenų bazėse. ROLAP tipo duomenų saugyklose galima saugoti gerokai didesnius duomenų kiekius nei MOLAP tipo. Tačiau ROLAP tipo saugyklose, užklausų atlikimo laikas yra gerokai ilgesnis, nei MOLAP saugyklose. Taigi, ROLAP tipo saugyklų privalumas yra tai, jog čia gali būti saugoma praktiškai neribotas duomenų kiekis. Pagrindinis trūkumas – tai gerokai lėtesnis užklausų apdorojimas, nei MOLAP tipo saugyklose (netgi apdorojant tą patį duomenų kiekį, ROLAP nusileidžia užklausų vykdymo sparta MOLAP tipo saugykloms).

- **HOLAP (angl. hybrid OLAP)** technologijos apjungia tiek MOLAP, tiek ROLAP privalumus. HOLAP saugykloje dalį duomenų galima saugoti MOLAP, kitą dalį – ROLAP tipo saugyklose. Paprastai, HOLAP saugykla šaltinio duomenis saugo duomenų šaltinio saugykloje, o savo saugykloje saugo tik agregacijas. Tačiau tai ilgina užklausų vykdymo trukmę. Kad išvengtų šios problemos, kuriami papildomi indeksai šaltinio duomenų saugykloje arba papildomos agregacijos HOLAP saugykloje. Tačiau pastaruoju atveju, prarandama disko vieta.

Kokio tipo duomenų saugyklą pasirinkti, priklauso nuo projektuojamo analizės modelio konkrečiu atveju - tai priklauso nuo apdorojamų duomenų kiekio, reikalingos duomenų apdorojimo spartos ir pan.

- **Šaltinio duomenų saugyklos atskyrimas nuo analizės saugyklos.** Jeigu įmanoma, šias saugyklas reiktų atskirti į atskirus serverius. Tai padidina darbo efektyvumą, jei sistemomis nuolat naudojasi daug vartotojų, kadangi tiek darbinės, tiek analizės sistemų darbo sparta yra labai svarbi.
- **Naudoti MOLAP ir HOLAP kubus vietoje ROLAP,** jeigu turima pakankamai atminties didesniems duomenų kiekiams saugoti. Tai žymiai sutrumpina užklausų vykdymo laiką, tačiau sunaudojama daugiau atminties.
- **Duomenų saugyklos duomenų bazę naudoti tik skaitymo režimu.** Tai sumažina kubo atnaujinimo laiką ir sutrumpina užklausų vykdymo laiką.
- **Kubus atnaujinti, kai sistema naudojasi mažiausiai vartotojų.** Kai sistema naudojasi mažiausia vartotojų, kubų atnaujinimas vyksta sparčiau.
- **Padidinti agregacijų skaičių.** Kuo daugiau agregacijų sukuriama kubo kūrimo metu, tuo greičiau yra vykdomos užklausos. Jeigu turima pakankamai daug vietos duomenų saugojimui, reiktų sukurti kuo daugiau agregacijų.
- **Skaidyti kubus dalimis (particijomis)** – tai leidžia vienu metu apdoroti keletą kubų ir efektyviau išnaudoti multiprocesorines sistemas.
- **Naudoti indeksus šaltinio duomenų bazėje** – tai paspartina duomenų paiešką bei apdorojimą. Tačiau reikalauja daugiau disko vietos.
- **Naudoti tik reikalingus, naujus duomenis.** Paprastai, analizei nėra reikalinga istorinė informacija, o tik paskutinių kelių mėnesių. Tad atnaujinant kubą galima nurodyti, kokius duomenis naudoti. Naudojant tik naujausius

duomenis, galima sutaupyti daug kubo atnaujinimo laiko ir reikalingos atminties resursų.

2.14. Analizės išvados

Atlikus analizę, galima daryti šias išvadas:

1. Tiriant įmonės, naudojančios TSM verslo valdymo sistemą, veiklos procesus nustatyta, jog standartinių sistemos ataskaitų nepakanka įmonės klientų analizei. Taip pat standartinės priemonės nėra tokios lanksčios. Dar vienas trūkumas – tai sistemos resursų stoka. Sudėtingų ataskaitų generavimas naudoja pernelyg daug sistemos resursų ir mažina bendrą sistemos našumą.
2. Iki šiol nėra priemonių, kurios leistų analizuoti įmonės klientų poreikius – yra kaupiama tik informacija, tačiau jai apdoroti priemonių jokių nėra. Todėl įmonei yra jų poreikis.
3. Analizės metu nustatyta galima sistemos architektūra – dėl darbinės sistemos resursų stokos nuspręsta naujai sistemai paskirti atskirus duomenų bei analizės serverius. Tai leis minimaliai apkrauti centrinį duomenų bazės serverį ir minimaliai įtakoti pagrindinės sistemos apkrovimą.
4. Taip pat nustatyti būsimos sistemos funkciniai (reikalingi duomenys, jų formatai ir pan.) ir nefunkciniai (reikalavimai sistemos našumui, patikimumui, saugumui ir pan.) reikalavimai, reikalavimai duomenims (už jų korektiškumą atsako TSM verslo valdymo sistema). Sistemos saugumo reikalavimai bus išpildyti, panaudojant esamus Active Directory vartotojus, jiems suteikiant skirtingas prieigas prie analizės priemonių galimybes.
5. Atlikus esamų sprendimų literatūros šaltiniuose analizę, pasirinktas analizės sistemos kūrimo sprendimo variantas – nuspręsta naudoti MOLAP duomenų saugyklą, panaudoti atskirą duomenų bazės bei analizės serverius. MOLAP duomenų saugyklos modelis pasirinktas, siekiant didinti sistemos veikimo spartą.
6. Pasiruošta sistemos diegimui – išanalizuota esama situacija, infrastruktūra, atliktas pirminis naujos sistemos integravimo bandymas (išgaunant duomenis iš esamos (TSM) verslo valdymo sistemos ir juos patalpinant atskiroje duomenų saugykloje).

3. Vaistinių tinklo duomenų analizės sistemos reikalavimų specifikacija

Analizės etapo metu surinktus vartotojų reikalavimus reikia patikslinti bei detaliam specifikuoti. Tam atliekama kuriamos sistemos reikalavimų specifikacija ir analizė. Tikslus būsimų sistemos vartotojų reikalavimų specifikuojimas padeda išvengti nesusipratimų programinės įrangos kūrimo metu ir garantuoja, jog sukurtas produktas maksimaliai atitiks vartotojų reikalavimus.

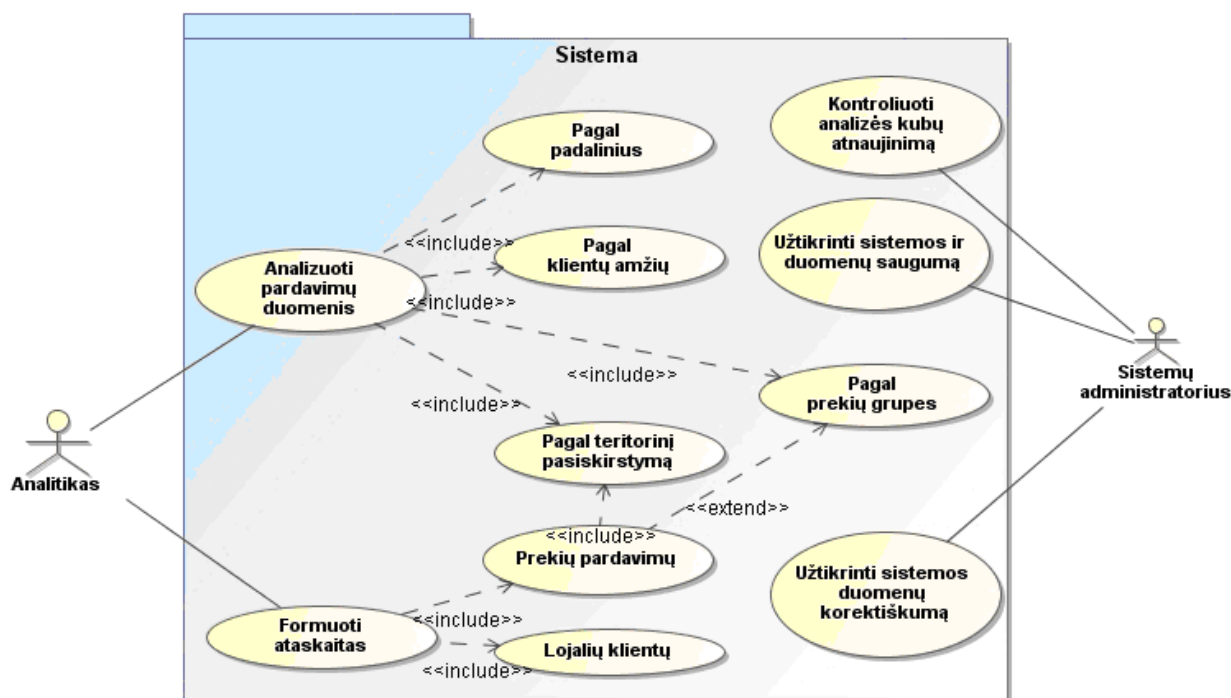
Šio projekto tikslas – sukurti UAB „Nemuno vaistinė“ klientų analizės sistemą, paremtą duomenų saugyklos pagrindu, kuria galėtų naudotis marketingo darbuotojai. Kuriamą sistemą padės efektyviau analizuoti pardavimų, klientų lojalumo ir jų poreikių duomenis, padės priimti strateginius marketingo sprendimus. Taip pat kuriamą sistemą leis efektyviau panaudoti sukauptą istorinę informaciją, pasiūlyti klientams reikalingiausias prekes bei paslaugas.

Reikalavimų specifikuojimo metu būtina tiksliai apibrėžti vartotojų reikalavimus būsimai sistemai, juos suderinti su užsakovu. Tik gerai atlikus reikalavimų specifikaciją ir analizę bei ją suderinus su užsakovais, galima tikėtis sėkmingo projekto.

3.1. Sistemos panaudojimo atvejų diagramos ir specifikacijos

Sistemos panaudojimo atvejų diagramos ir jų specifikacijos leidžia tiksliai identifikuoti būsimus sistemos vartotojus, jų poreikius bei atliekamus veiksmus. Visiems sistemos vartotojams sudaromos panaudojimo atvejų diagramos, jų specifikacijos bei tų vartotojų panaudojimo atvejų sekų, veiklos diagramos.

Sistemos vartotojai – tai analitikai, kurie analizuoja įmonės klientus, prekių pardavimus. Analitiko panaudojimo atvejų diagrama pavaizduota žemiau esančiame paveikslėlyje:



17 pav. Sistemos panaudojimo atvejų diagrama

Sistemų administratoriaus funkcijos – tai prižiūrėti bendrą sistemos darbą, užtikrinti saugumą.

Marketingo analitikas analizuoja sistemos pateikiamus duomenis ir priima strateginius sprendimus. Analitikas sistemos pagalba gali analizuoti pardavimų duomenis bei formuoti įvairias ataskaitas. Toliau pateikiamos kiekvieno panaudojimo atvejo specifikacijos.

- **Panaudojimo atvejis „Analizuoti pardavimų duomenis“**

Panaudojimo atvejis „Analizuoti pardavimų duomenis“ apima duomenų analizę įvairiais pjūviais (padalinių, klientų amžiaus grupių, teritorinio pasiskirstymo, prekių kategorijų). Panaudojimo atvejo specifikacija pateikta žemiau esančioje lentelėje.

5 lentelė. Panaudojimo atvejo „Analizuoti pardavimų duomenis“ specifikacija

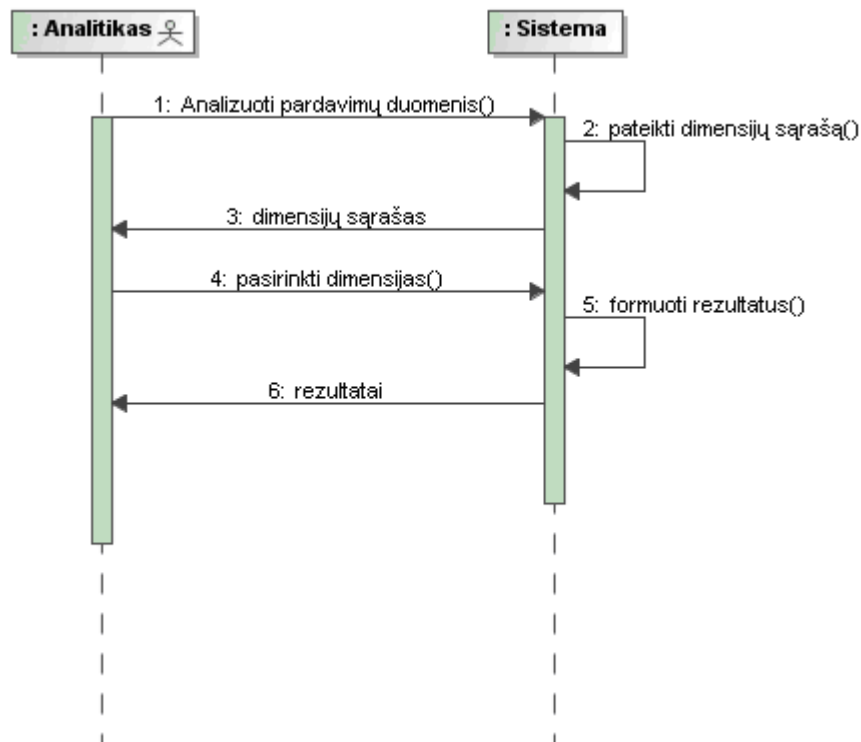
Panaudojimo atvejis „Peržiūrėti naujienas“	
Aktorius	Analitikas
Prieš sąlyga	
Sužadinimo sąlyga	Vartotojas nori peržiūrėti sukauptus pardavimų duomenis
Susiję panaudojimo atvejai	Išplečia PA
	Apima PA
	Specializuoja PA
Pagrindinis įvykių srautas	Sistemos reakcija ir sprendimai
1. Vartotojas pasirenka duomenų analizę	1. Sistema pateikia dimensijų sąrašą
Po sąlyga	Vartotojas peržiūri duomenis norimais pjūviais

Alternatyvūs scenarijai

1. Nepavyko prisijungti prie duomenų saugyklos

Sistema suformuoja klaidos pranešimą.

Panaudojimo atvejo „Analizuoti pardavimų duomenis“ sekų diagrama pavaizduota žemiau esančiame paveikslėlyje.



18 pav. Panaudojimo atvejo „Analizuoti pardavimų duomenis“ sekų diagrama

- **Panaudojimo atvejis „Pagal padalinius“**

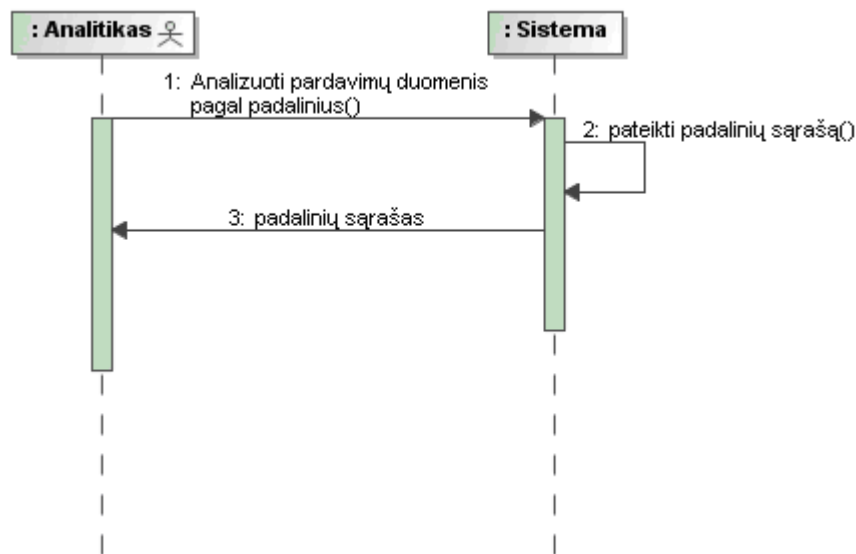
Panaudojimo atvejis „Pagal padalinius“ išplečia panaudojimo atvejį „Analizuoti pardavimų duomenis“. Sistemos vartotojui pasirinkus panaudojimo atvejį „Analizuoti pardavimų duomenis“, vartotojas gali pasirinkti, pagal kokius kriterijus tuos duomenis analizuoti. Pasirinkus „Pagal padalinius“, sistema suformuoja padalinių sąrašą.

6 lentelė. Panaudojimo atvejo „Pagal padalinius“ specifikacija

Panaudojimo atvejis „Peržiūrėti naujienas“		
Aktorius		Analitikas
Prieš sąlyga		
Sužadinimo sąlyga		Vartotojas nori peržiūrėti sukauptus pardavimų duomenis pagal padalinius
Susiję panaudojimo atvejai	Išplečia PA	Analizuoti pardavimų duomenis
	Apima PA	
	Specializuoja PA	
Pagrindinis įvykių srautas		Sistemos reakcija ir sprendimai

1. Vartotojas pasirenka duomenų analizę 2. Pasirenkama duomenų analizė pagal padalinius	1. Sistema pateikia dimensijų sąrašą 2. Sistema suformuoja padalinių sąrašą
Po sąlyga	Vartotojas peržiūri duomenis norimais pjūviais
Alternatyvūs scenarijai	
1. Nepavyko prisijungti prie duomenų saugyklos	Sistema suformuoja klaidos pranešimą.

Panaudojimo atvejo „Pagal padalinius“ sekų diagrama pavaizduota žemiau esančiame paveikslėlyje.



19 pav. Panaudojimo atvejo „Pagal padalinius“ sekų diagrama

- **Panaudojimo atvejis „Pagal klientų amžių“**

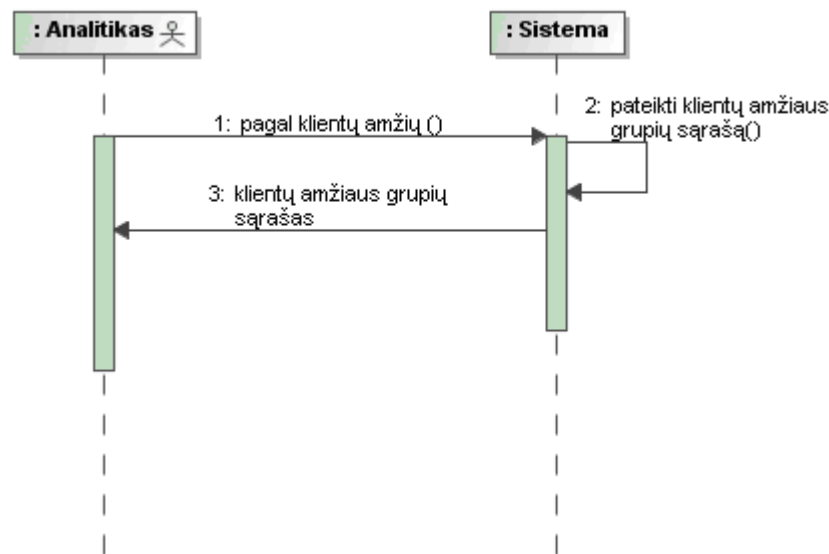
Panaudojimo atvejis „Pagal klientų amžių“ išplečia panaudojimo atvejį „Analizuoti pardavimų duomenis“. Sistemos vartotojui pasirinkus panaudojimo atvejį „Analizuoti pardavimų duomenis“, vartotojas gali pasirinkti, pagal kokius kriterijus tuos duomenis analizuoti. Pasirinkus „Pagal klientų amžių“, sistema suformuoja klientų amžiaus grupių sąrašą.

7 lentelė. Panaudojimo atvejo „Pagal klientų amžių“ specifikacija

Panaudojimo atvejis „Peržiūrėti naujienas“		
Aktorius	Analitikas	
Prieš sąlyga		
Sužadinimo sąlyga	Vartotojas nori peržiūrėti sukauptus pardavimų duomenis pagal klientų amžių	
Susiję panaudojimo atvejai	Išplečia PA	Analizuoti pardavimų duomenis
	Apima PA	
	Specializuoja PA	
Pagrindinis įvykių srautas	Sistemos reakcija ir sprendimai	

1. Vartotojas pasirenka duomenų analizę 2. Pasirenkama duomenų analizė pagal padalinius	1. Sistema pateikia dimensijų sąrašą 2. Sistema suformuoja klientų amžiaus grupių sąrašą
Po sąlyga	Vartotojas peržiūri duomenis norimais pjūviais
Alternatyvūs scenarijai	
1. Nepavyko prisijungti prie duomenų saugyklos	Sistema suformuoja klaidos pranešimą.

Panaudojimo atvejo „Pagal klientų amžių“ sekų diagrama pavaizduota žemiau esančiame paveikslėlyje.



20 pav. Panaudojimo atvejo „Pagal klientų amžių“ sekų diagrama

- **Panaudojimo atvejis „Pagal prekių grupes“**

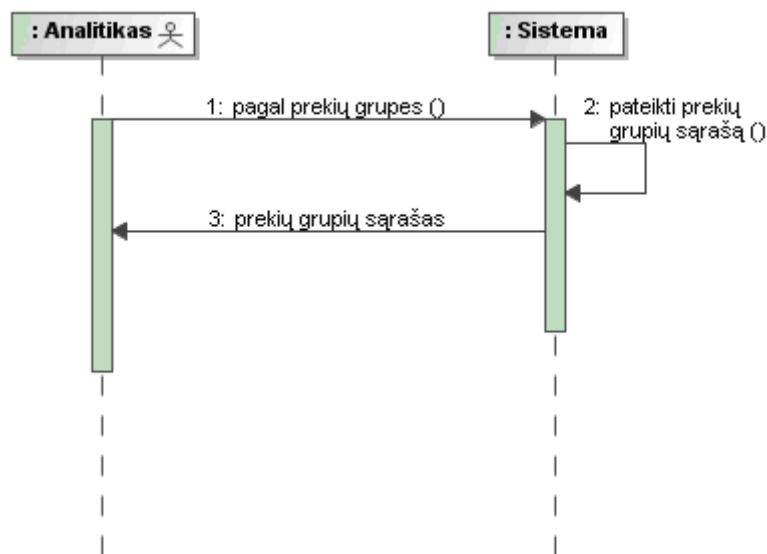
Panaudojimo atvejis „Pagal prekių grupes“ išplečia panaudojimo atvejį „Analizuoti pardavimų duomenis“. Sistemos vartotojui pasirinkus panaudojimo atvejį „Analizuoti pardavimų duomenis“, vartotojas gali pasirinkti, pagal kokius kriterijus tuos duomenis analizuoti. Pasirinkus „Pagal prekių grupes“, sistema suformuoja prekių grupių sąrašą.

8 lentelė. Panaudojimo atvejo „Pagal klientų amžių“ specifikacija

Panaudojimo atvejis „Peržiūrėti naujienas“		
Aktorius	Analitikas	
Prieš sąlyga		
Sužadinimo sąlyga	Vartotojas nori peržiūrėti sukauptus pardavimų duomenis pagal prekių grupes	
Susiję panaudojimo atvejai	Išplečia PA	Analizuoti pardavimų duomenis
	Apima PA	
	Specializuoja PA	
Pagrindinis įvykių srautas	Sistemos reakcija ir sprendimai	

1. Vartotojas pasirenka duomenų analizę 2. Pasirenkama duomenų analizė pagal padalinius	1. Sistema pateikia dimensijų sąrašą 2. Sistema suformuoja prekių grupių sąrašą
Po sąlyga	Vartotojas peržiūri duomenis norimais pjūviais
Alternatyvūs scenarijai	
1. Nepavyko prisijungti prie duomenų saugyklos	Sistema suformuoja klaidos pranešimą.

Panaudojimo atvejo „Pagal prekių grupes“ sekų diagrama pavaizduota 20 paveiksle.



20 pav. Panaudojimo atvejo „Pagal prekių grupes“ sekų diagrama

- **Panaudojimo atvejis „Pagal teritorinį pasiskirstymą“**

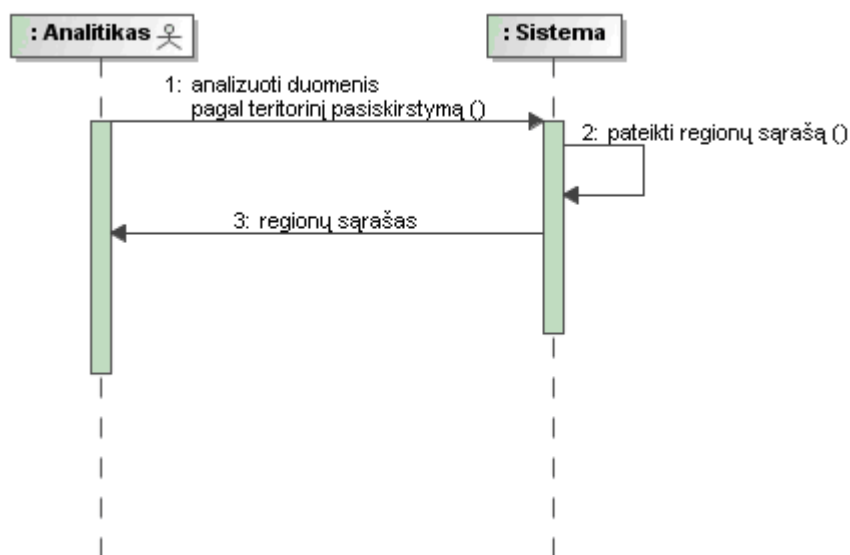
Panaudojimo atvejis „Pagal teritorinį pasiskirstymą“ išplečia panaudojimo atvejį „Analizuoti pardavimų duomenis“. Sistemos vartotojui pasirinkus panaudojimo atvejį „Analizuoti pardavimų duomenis“, vartotojas gali pasirinkti, pagal kokius kriterijus tuos duomenis analizuoti. Pasirinkus „Pagal teritorinį pasiskirstymą“, sistema suformuoja regionų sąrašą. Suformuojama s sąrašas pagal gyventojų skaičių tam tikram regione. Pagal tai galima analizuoti pardavimų informaciją kaimo, miesto vietovėse ir pan.

8 lentelė. Panaudojimo atvejo „Pagal teritorinį pasiskirstymą“ specifikacija

Panaudojimo atvejis „Peržiūrėti naujienas“		
Aktorius	Analitikas	
Prieš sąlyga		
Sužadinimo sąlyga	Vartotojas nori peržiūrėti sukauptus pardavimų duomenis pagal teritorinį pasiskirstymą	
Susiję panaudojimo	Išplečia PA	Analizuoti pardavimų duomenis
	Apima PA	

atvejai	Specializuoja PA	
Pagrindinis įvykių srautas		Sistemos reakcija ir sprendimai
1. Vartotojas pasirenka duomenų analizę 2. Pasirenkama duomenų analizė pagal padalinius		1. Sistema pateikia dimensijų sąrašą 2. Sistema suformuoja prekių grupių sąrašą
Po sąlyga		Vartotojas peržiūri duomenis norimais pjūviais
Alternatyvūs scenarijai		
1. Nepavyko prisijungti prie duomenų saugyklos		Sistema suformuoja klaidos pranešimą.

Panaudojimo atvejo „Pagal teritorinį pasiskirstymą“ sekų diagrama pavaizduota žemiau esančiame paveikslėlyje.



21 pav. Panaudojimo atvejo „Pagal teritorinį pasiskirstymą“ sekų diagrama

- **Panaudojimo atvejis „Formuoti ataskaitas“**

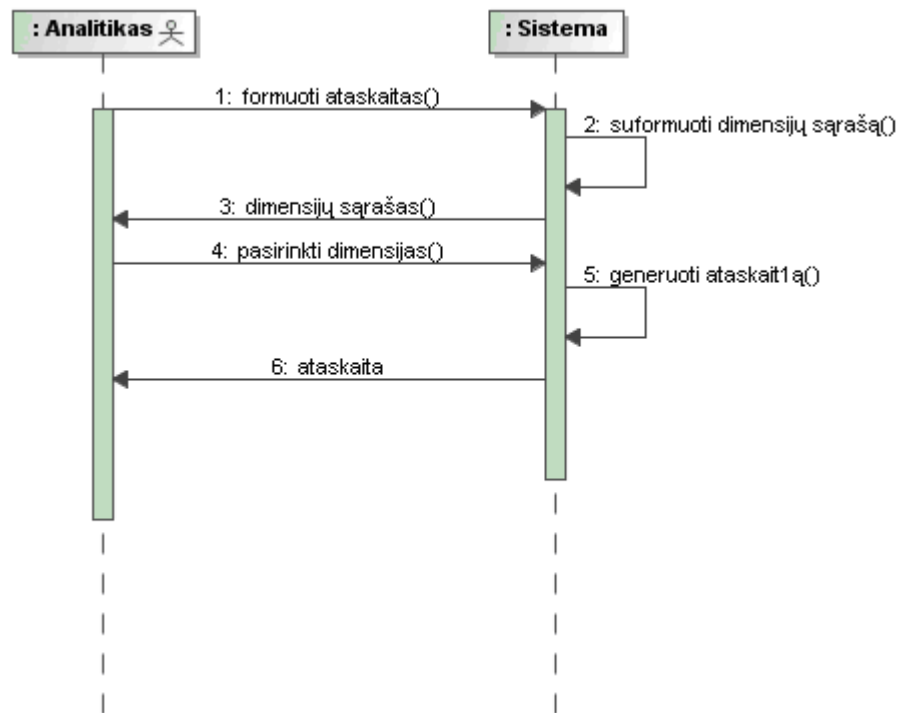
Panaudojimo atvejis „Formuoti ataskaitas“ apima įvairių analitinių ataskaitų formavimą. Analitikui, analizuojančiam sukauptus pardavimų duomenis, yra svarbu sukurti įvairias ataskaitas, kuriose būtų galima matyti duomenis norimais pjūviais, diagramomis.

9 lentelė. Panaudojimo atvejo „Formuoti ataskaitas“ specifikacija

Panaudojimo atvejis „Peržiūrėti naujienas“		
Aktorius	Analitikas	
Prieš sąlyga		
Sužadinimo sąlyga	Vartotojas nori sukurti analizuojamų duomenų ataskaitas	
Susiję panaudojimo atvejai	Išplečia PA	
	Apima PA	
	Specializuoja PA	
Pagrindinis įvykių srautas		Sistemos reakcija ir sprendimai

1. Vartotojas pasirenka ataskaitų formavimą	1. Sistema pateikia dimensijų sąrašą
Po sąlyga	Vartotojas suformuoja norimas ataskaitas
Alternatyvūs scenarijai	
1. Nepavyko prisijungti prie duomenų saugyklos	Sistema suformuoja klaidos pranešimą.

Panaudojimo atvejo „Formuoti ataskaitas“ sekų diagrama pavaizduota žemiau esančiame paveikslėlyje.



22 pav. Panaudojimo atvejo „Formuoti ataskaitas“ sekų diagrama

3.2. Analizės kubo specifikacija

Sukauptų duomenų analizei naudojamas duomenų kubas KLIENTAI. Šis kubas skirtas analizuoti informaciją apie įmonės klientus, jų perkamas prekes, analizuoti klientų lojalumą, amžiaus grupes ir pan.

Būsimo klientų analizės kubo struktūra:

10 lentelė. Duomenų kubo struktūra

Dimensijos	
Data	Operacijos data
Metai	Operacijos atlikimo metai
Ketvirtis	Operacijos atlikimo ketvirtis
Mėnesis	Operacijos atlikimo mėnuo
Diena	Operacijos atlikimo diena

Lytis	Kliento lytis (V-Vyras, M-Moteris)
Padalinys	Įmonės padalinys
Rajonas	Rajonas, kuriame yra padalinys
Miestas	Miestas, kuia,e yra padalinys
Gyventojų skaičius	Gyventojų skaičius mieste/gyvenvietėje, kuriame yra padalinys
Prekės pavadinimas	Prekės pavadinimas
Prekių grupė	Prekių grupės pavadinimas
Kompensacijos pavadinimas	Kompensacijos pavadinimas
Matai	
Nuolaidų suma	Suteiktų nuolaidų suma
Nuolaidų procentas	Nuolaidų procentas
Bendra suma	Bendra suma
Draudimo nuolaida	Draudimo nuolaida
Kompensacijos suma	Kompensacijos suma
Skaičiuojamieji laukai	
Nuolaidų suma (AVG)	Vidutinė nuolaidų suma
Nuolaidų procentas (AVG)	Vidutinis nuolaidų procentas
Bendra suma (AVG)	Vidutinė bendra suma
Draudimo nuolaida (AVG)	Vidutinė draudimo nuolaida
Kompensacijos suma (AVG)	Vidutinė kompensacijos suma
Apsilankymų intervalas	Klientų apsilankymo intervalas (savaitės, mėnesiai)
Amžiaus grupė	Klientų amžiaus grupė

3.3. Detali reikalavimų specifikacija

Išgavus reikalavimus iš būsimų sistemos vartotojų, specifikavus būsimo kubo struktūrą, toliau reikia tiksliai nustatyti, iš kur kiekviena dimensija ir matas įgaus reikšmes. Nustačius reikalingus laukus, lenteles, iš kurių reikšmės bus reikalingos, projekto realizacijos metu bus kuriami duomenų perkėlimo servisi (angl. DTS – Data Transfer Services), kurių pagalba bus perkeliama reikalingi laukai ir reikšmės iš darbinės į analizės duomenų bazes. Žemiau esančioje lentelėje surašyti visi projektuojamo analizės kubo laukai ir lentelės bei jų laukai, iš kurių paimama informacija į analizės kubą.

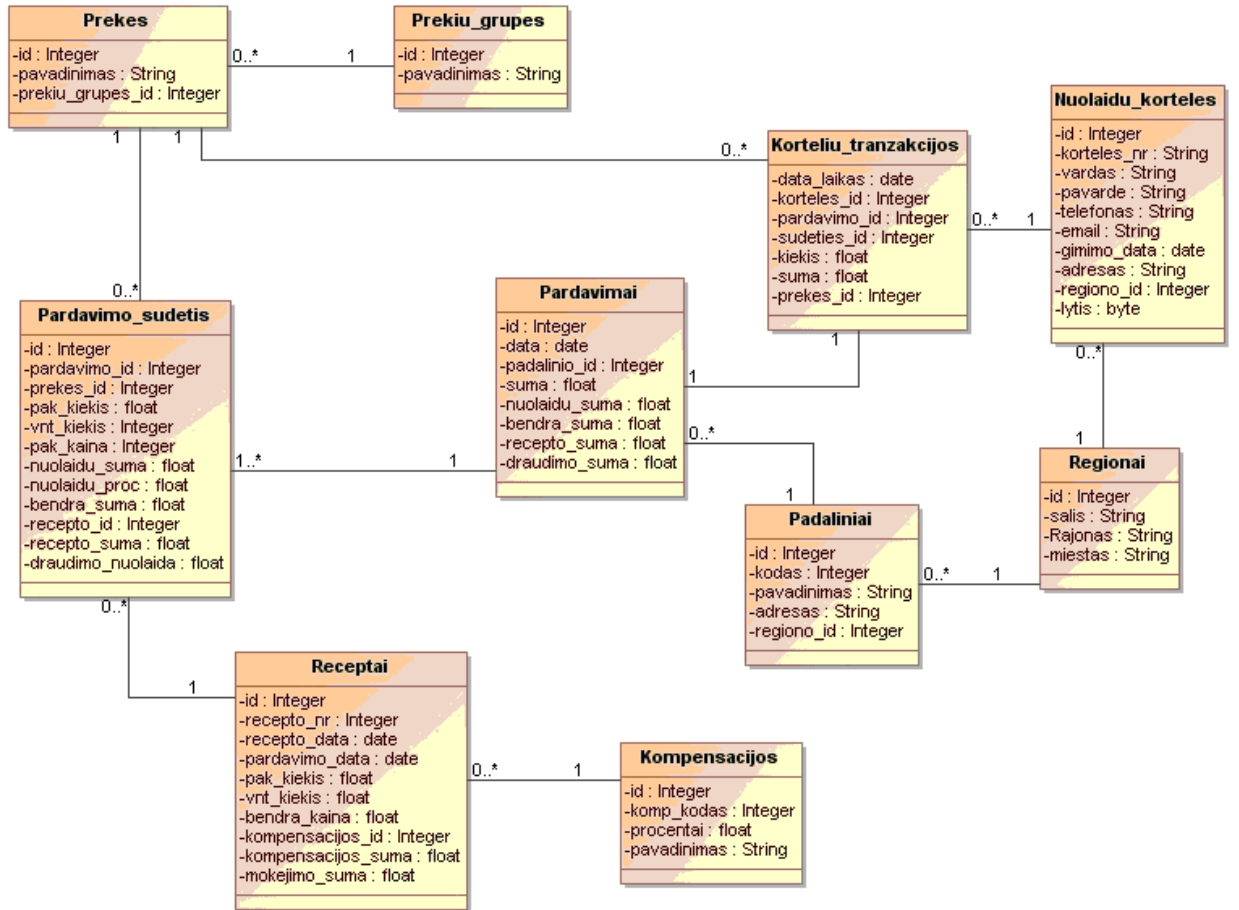
11 lentelė. Analizės kubo sąryšiai su duomenų bazės lentelėmis

Dimensija	Duomenų bazės lentelė	Laukas
Data	pardavimai@TSM2	Data
Metai	pardavimai@TSM2	Data
Ketvirtis	pardavimai@TSM2	Data

Mėnesis	pardavimai@TSM2	Data
Diena	pardavimai@TSM2	Data
Lytis	nuolaidu_korteles@TSM2	Lytis
Padalinys	padaliniai@TSM2	Pavadinimas
Rajonas	regionai@TSM2	Rajonas
Miestas	regionai@TSM2	Miestas
Gyventoju skaičius	gyventoju_sk@TSM2	Gyventoju skaičius
Prekės pavadinimas	prekes@TSM2	Pavadinimas
Prekių grupė	prekiu_grupes@TSM2	Pavadinimas
Kompensacijos pavadinimas	kompensacijos@TSM2	Pavadinimas
Matas	Duomenų bazės lentelė	Laukas
Nuolaidų suma	pardavimai@TSM2	Nuolaidu_suma
Nuolaidų procentas	pardavimai@TSM2	Nuolaidu_proc
Bendra suma	pardavimai@TSM2	Bendra suma
Draudimo nuolaida	pardavimai@TSM2	Draudimo_nuolaida
Kompensacijos suma	receptai@TSM2	Kompensacijos_suma
Skaičiuojamieji laukai	Formulė	
Nuolaidų suma (AVG)	AVG(Nuolaidų suma)	
Nuolaidų procentas (AVG)	AVG(Nuolaidų procentas)	
Bendra suma (AVG)	AVG (Bendra suma)	
Draudimo nuolaida (AVG)	AVG(Draudimo nuolaida)	
Kompensacijos suma (AVG)	AVG(Kompensacijos suma)	
Apsilankymų intervalas	Function <i>lp_gauti_apsilankymu_intervala (p_kortelesid IN NUMBER)</i>	
Amžiaus grupė	case when gimimo_data >= sysdate-365*18 then '0-18' when gimimo_data < sysdate-365*18 and gimimo_data >= sysdate-365*36 then '18-36' when gimimo_data < sysdate-365*37 and gimimo_data >= sysdate-365*44 then '37-44' when gimimo_data < sysdate-365*45 and gimimo_data >= sysdate-365*62 then '45-62' when gimimo_data < sysdate-365*63 then '63+' else '---' end	

3.4. Dalykinės srities modelis

Visi analizės kubo duomenys yra paimami iš analizės duomenų bazės lentelių. TSM2 – tai antrinė duomenų bazė, atskira nuo pagrindinės (TSM) duomenų bazės. TSM2 (analizės) duomenų bazėje yra tik tos lentelės, kurių duomenys naudojami projektuojamame analizės kube. TSM2 duomenų bazės struktūra ir lentelių sąryšiai pavaizduoti žemiau esančiame paveikslėlyje.



22 pav. Duomenų bazės TSM2 schema

4. Klientų analizės sistemos projektas

Vienas svarbiausių projekto etapų – tai pačios sistemos projektas. Šį etapą įtakoja visi anksčiau buvę etapai (vartotojų reikalavimų išgavimas ir analizė, reikalavimų specifikacijos). Tik gerai parengus sistemos projektą ir atsižvelgus į vartotojų reikalavimus bei poreikius, galima sukurti naudingą ir efektyvų produktą, kuris teiks naudą vartotojams.

Projekto dekompozicija [1] gali būti atliekama dviem kryptimis: iš viršaus į apačią ir iš apačios į viršų. Dažniausiai projekto struktūra sudaroma abiem kryptimis. Valdymo lygmens dekompozicijos lygiai nustatomi iš viršaus į apačią, o techninio lygmens dekompozicijos lygiai – iš apačios į viršų. Projekto pagrindinis tikslas išskaidomas iki uždavinių, o po to, numatomi darbai, kurie gali būti lengvai įvertinti kaštų ar laiko rodikliais ir būtini sprendžiant prieš tai išskirtus uždavinius; jie grupuojami į kompleksus, o po to į sistemas.

Rengiant šį projektą, naudojama „iš viršaus žemyn“ metodika – t.y. pradžioje specifikuojamas galutinis rezultatas (norimas duomenų kubas), o toliau detalizuojama, kaip ir iš ko jis bus sudarytas.

4.1. Sistemos pagrindimas ir esmės išdėstymas

UAB „Nemuno vaistinė“ užsiima didmenine ir mažmenine prekyba vaistais. Įmonės tinklas nuolat sparčiai plečiasi. Norint pasiūlyti klientams tai, ko jiems reikia, yra svarbu tirti rinką, analizuoti klientų poreikius.

Nuo 2003 metų pradėtos taikyti nuolaidų kortelės. Klientams, perkantiems prekes su nuolaidų kortele, visoms prekėms taikomos nuolaidos. Tai skatina klientus naudotis šiomis kortelėmis. Klientui kortelė – tai galimybė įsigyti reikalingas prekes pigiau. O prekybininkui – tai būdas stebėti klientų elgseną, įpročius, analizuoti jų poreikius. Taip galima pateikti klientams tai, ko jiems reikia. Nuo pat 2003 metų, kai buvo pradėtos platinti pirmosios nuolaidų kortelės, iki šiol įmonėje nebuvo praktiškai jokios klientų analizės sistemos. Yra tikrai pardavimų duomenys, tačiau jie realiai niekur nepanaudojami. Nuolaidų kortelės veikė tikrai kaip klientų pritraukimo priemonė – tiesiog, klientai gauna kiekvieną kartą nuolaidą už prekes.

Tačiau sukauptus pardavimų duomenis galima panaudoti, ir gana efektyviai – išanalizavus pardavimus, kuomet panaudojamos kortelės, galima nustatyti, kokias prekes dažniausiai perka klientai, koks yra asortimento pasiskirstymas tarp didmiesčių ir mažesnių miestelių klientų. Tai leistų efektyviai nustatyti klientų poreikius, pasiūlyti jiems reikalingų prekių bei pritraukti daugiau klientų - tuo pačiu ir padidinti apyvartumus.

Taigi, atsirado poreikis analizuoti šiuos duomenis. Tam buvo nuspręsta sukurti įmonės klientų analizės sistemą, kuri leistų atlikti visus aukščiau išvardintus veiksmus. Kadangi klientų kortelių jau yra sukaupta per 600.000, o pardavimų informacijos kasdien sukaupiami milžiniški kiekiai, tad tokio pobūdžio analizei praktiškai nėra kito varianto, kaip tik sukurti analizės kubą, kurio pagalba būtų galima greitai ir efektyviai gauti norimus rezultatus.

Taigi, nuspręsta sukurti atskirą domenų saugyklą, kurioje būtų saugoma informacija apie įmonės padalinių pardavimus, nuolaidų kortelių panaudojimo istorijos ir kita susijusi informacija.

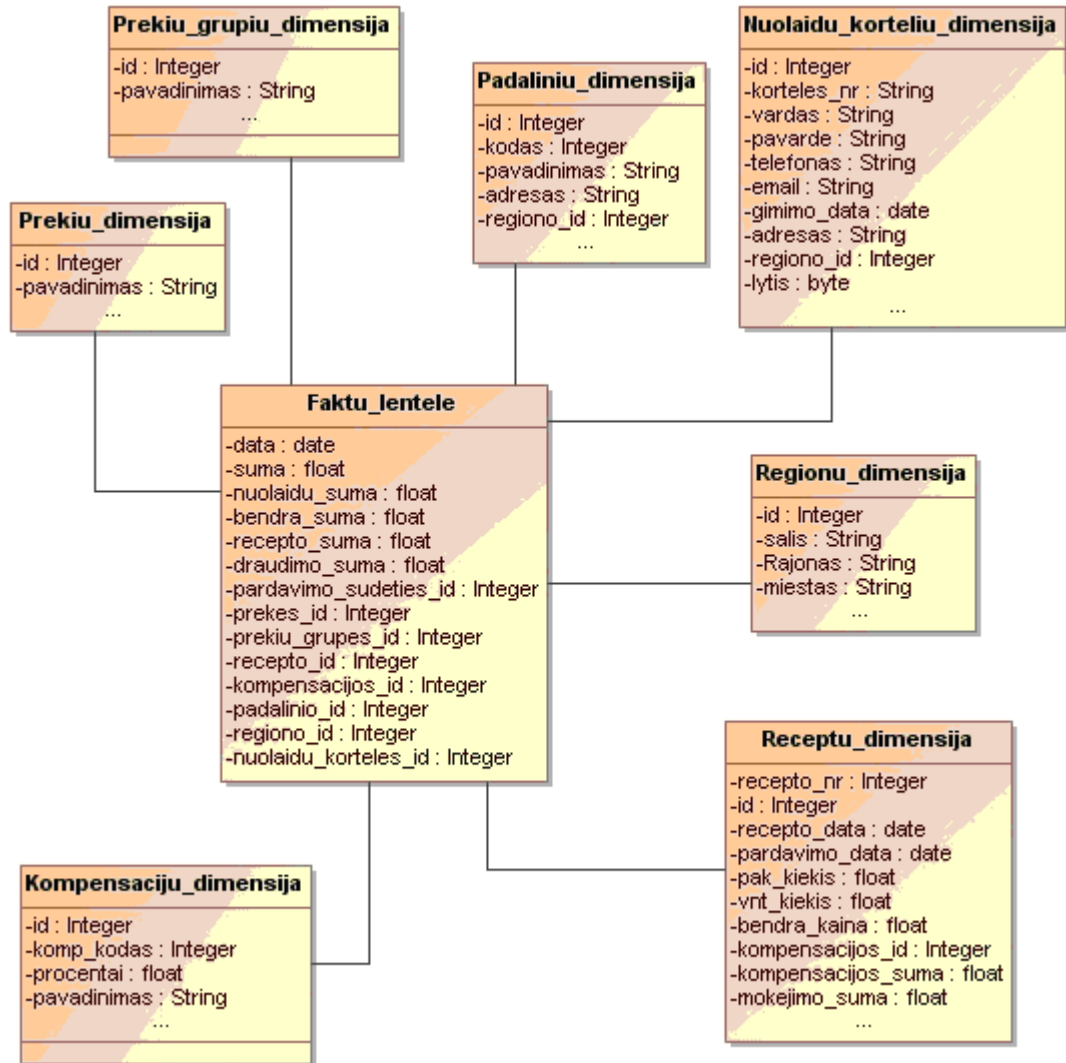
4.2. Sistemos architektūra

Labai svarbus sistemų kūrimo etapas yra projekto parengimas. Išgavus iš būsimų sistemos vartotojų visus reikalavimus, galima pradėti kurti sistemos projektą.

Gerai parengtas ir su būsimos sistemos vartotojais suderintas sistemos projektas leidžia sumažinti projekto įgyvendinimo klaidų tikimybę, leidžia užtikrinti būsimų sistemos vartotojų reikalavimų išpildymą. Tinkamai parengtas sistemos projektas leidžia numatyti reikalingus būsimos sistemos elementus, jų sujungimo būdus, veiksmų sekas.

Rengiant šį projektą, panaudota „iš viršaus žemyn“ projektavimo technologija – t.y. pradžioje specifikuojamas galutinis rezultatas (analizės kubas), o toliau detalizuojama, kaip ir iš ko jis bus gaunamas, kokių elementų reiks ir pan.

Suprojektuoto analizės kubo schema pavaizduota 23 paveikslėlyje.



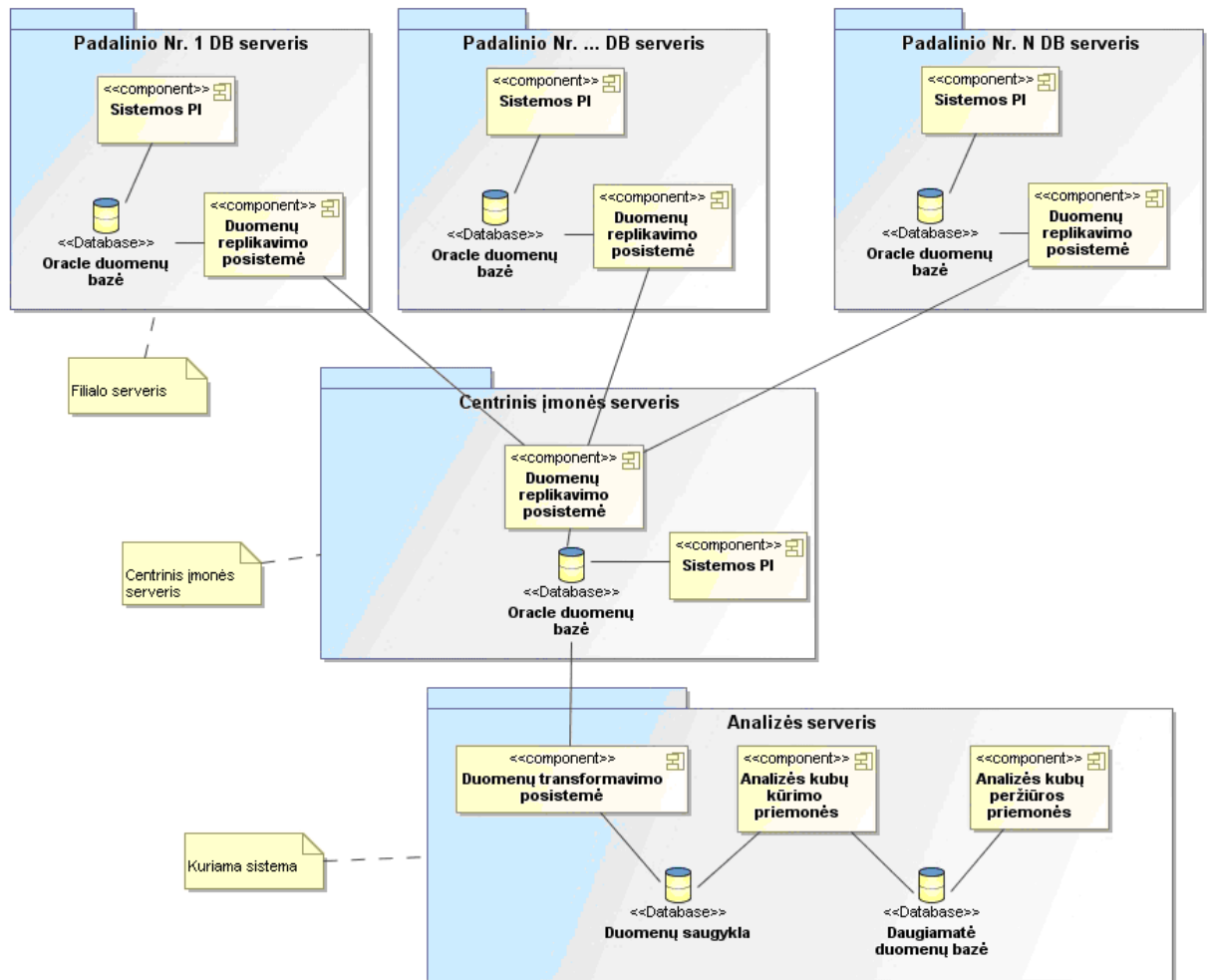
23 pav. Kuriamos sistemos OLAP schema

4.2.1. Bendra sistemos architektūra

UAB „Nemuno vaistinė“ turi per 200 filialų, kurių kiekviename veikia po atskirą filialo serverį. Šiuose serveriuose saugoma atitinkamo filialo informacija. Per duomenų replikavimo posistemės, šie duomenys yra apsieičiami su centriniu įmonės serveriu.

Kuriant įmonės klientų analizės sistemą, vienas svarbiausių reikalavimų buvo kuo mažiau apkrauti darbinę įmonės sistemą. Kadangi duomenų centriname duomenų bazės serveryje yra labai daug, jų išrinkimas, apdorojimas reikalauja nemažai pagrindinės sistemos resursų. Tad naujosios sistemos įgyvendinimui nuspręsta skirti atskirą duomenų bazės bei analizės serverius. Atskirame duomenų bazės serveryje bus saugoma įmonės klientų duomenų analizei reikalinga informacija. Ji nuolat atnaujinama centrinės duomenų bazės duomenimis. Toks sprendimas leidžia minimaliai apkrauti pagrindinę sistemą (nes duomenų nukopijavimas

reikalauja gerokai mažiau pagrindinės sistemos resursų, nei tų duomenų analizė tame pačiame darbiname serveryje. Bendra sistemos architektūra pavaizduota žemiau esančiame paveikslėlyje.



24 pav. Bendra sistemos architektūra

4.2.2. Reikalavimai sistemos funkcionavimo palaikymui

Projektuojant sistemą, labai svarbu yra numatyti reikalingos įrangos techninius parametrus, siekiant užtikrinti stabilų visos sistemos veikimą.

Projektuojama sistema – tai dalis įmonės esamos sistemos. Visi įmonės duomenys saugomi centrinėje Oracle duomenų bazėje. Į šia duomenų bazę replikų pagalba patenka duomenys iš visų nutolusių įmonės filialų serverių. Filialo serveryje taip pat veikia Oracle duomenų bazės serveris, kuriama saugoma lokali filialo informacija. Lokaliai filialo serveriui keliami minimalūs reikalavimai nurodyti žemiau esančioje lentelėje.

12 lentelė. Minimalūs reikalavimai filialo serveriui

Įrenginys	Parametrai
Procesorius	INTEL CELERON 430, 1.8 GHZ/FSB 800/512KB
Pagrindinė plokštė	G31M-S2L, iG31/ICH7/FSB1333/2xDDR2-800 DC max 4GB/PCI-Ex16/PCIEx1/2xPCI/ATA/4x SATA2/VGA/8ch/ALC662/Gb-LAN/mATX
Operatyvioji atmintis	1GB 800MHZ DDR2 NON-ECC CL5 DIMM
Kietasis diskas	WD CAVIAR 80GB SATA2 7200RPM 8MB
Grafikos lustų rinkinys	Integruotas
Garso plokštė	Integruotas
Klaviatūra	---
Pelė	---
Operacinė sistema	Windows XP Pro EN OEM

Kaip aprašyta bendroje sistemos architektūroje, visi duomenys yra saugomi centriniame įmonės serveryje. Šiam įrenginiui keliami rekomendaciniai techniniai reikalavimai pateikti žemiau esančioje lentelėje.

13 lentelė. Reikalavimai pagrindiniam serveriui

Įrenginys	Parametrai
Procesorius	Intel(R) Xeon(R) CPU X5450 @ 3.00GHz, 6144 KB cash, 8 branduolių
Operatyvioji atmintis	8GB
Kietasis diskas	320GB
Grafikos lustų rinkinys	Integruotas
Garso plokštė	Integruotas
Klaviatūra	--
Pelė	--
Operacinė sistema	Linux šeimos OS

Reikalavimų dalyje buvo nustatyta, jog analizei skirtus duomenis reikia saugoti atskirame duomenų bazės serveryje. Kadangi analizės kubų realizacija vykdoma Microsoft Anglysis Services, paprastumo dėlei nuspręsta naudoti Microsoft SQL Server 2000 duomenų bazės serverį (nes jo licenzija ir taip yra nupirkta, tad tai mažina projekto išlaidas). Analizės duomenų bazės serveriui keliami techniniai reikalavimai pateikti žemiau esančioje lentelėje.

14 lentelė. Minimalūs reikalavimai analizės duomenų bazės serveriui

Įrenginys	Parametrai
Procesorius	x86 Family 15 Model 4 Stepping 3 GenuineIntel ~3000 Mhz
Operatyvioji atmintis	2.5GB 800MHZ DDR2 NON-ECC CL5 DIMM
Kietasis diskas	WD CAVIAR 80GB SATA2 7200RPM 8MB
Grafikos lustų rinkinys	Integruotas
Garso plokštė	Integruotas
Klaviatūra	---
Pelė	---

Analizės serveriui, kuriame apdorojami, atnaujinami duomenų analizės kubai, taip pat keliami didesni reikalavimai, kadangi yra svarbu trumpas duomenų kubų atnaujinimo, užklausių apdorojimo laikas. Analizės servisai (angl. Analysis Services) veikia tame pačiame analizės serveryje. Tad atskirai reikalavimai šiam serveriui nekeliami, nes jie veikia kaip papildomi servisai Microsoft SQL Server 2000 serveryje.

Be įprastų techninės įrangos parametrų, nemažiau svarbūs ir ryšio tarp serverių parametrai. Visi filialų serveriai turi turėti tiesioginį ryšį su centriniu duomenų bazės serveriu. Kadangi filialai yra išsidėstę skirtingose geografinėse vietovėse, optimaliausia jų tarpusavio sujungimo būdas – tai pasinaudoti internetiniu ryšiu, sujungiant serverius VPN IPsec tunelio pagalba. Šiais ryšio kanalais duomenys periodiškai sinchronizuojami su centriniu serveriu.

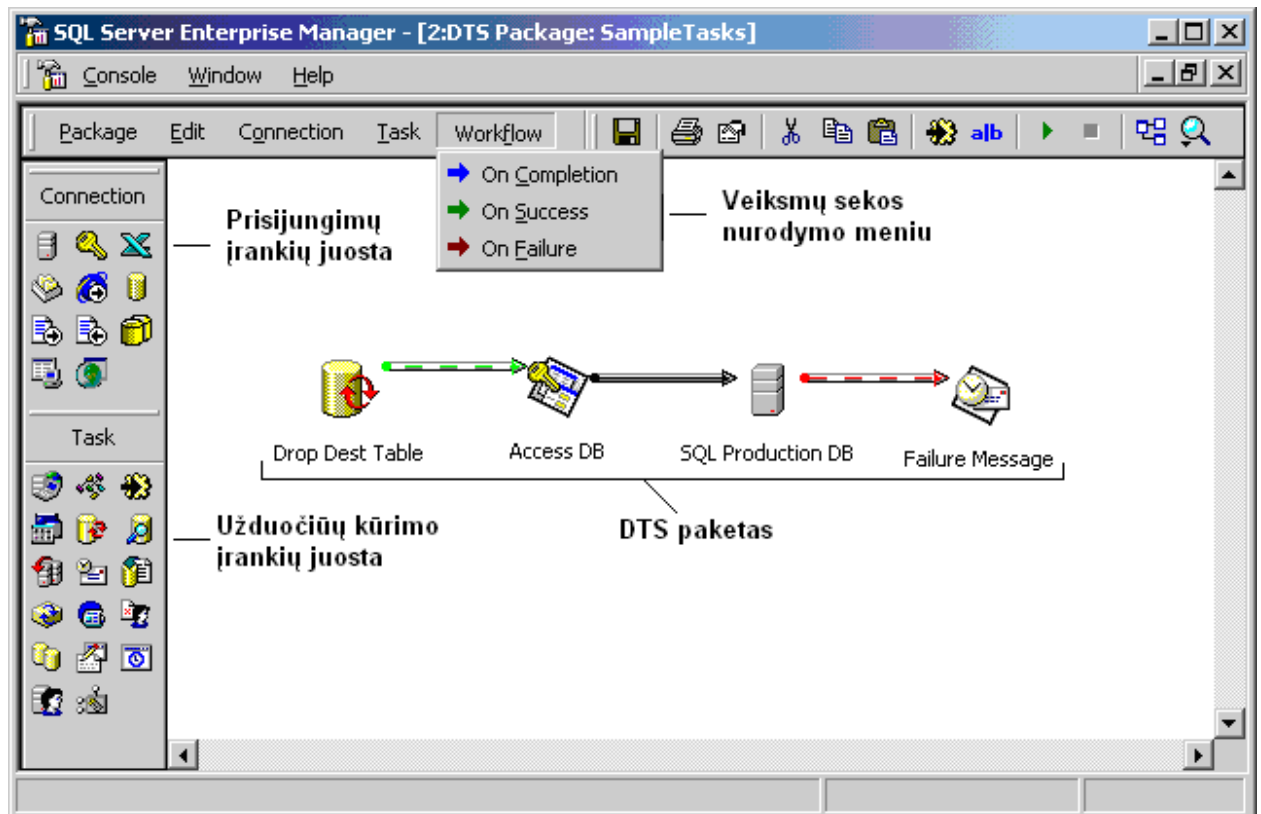
Visi kiti serveriai (centrini, analizės duomenų bazės, analizės serveris) – yra viename lokaliame tinkle, tad juos sujungti pakanka įprasto vietinio LAN ryšio užtikrinimo.

4.2.3. Duomenų transformacijų servisai

Analizės etapo metu buvo nurodyta, jog vienas svarbiausių reikalavimų - tai kuo mažiau apkrauti pagrindinę sistemą. Dėl to nuspręsta visus analizei reikalingus duomenis saugoti atskirame duomenų bazės serveryje. Tad yra svarbu užtikrinti savalaikį ir patikimą duomenų atnaujinimo mechanizmą tarp pagrindinės ir analizės duomenų bazių.

Duomenų atnaujinimui nuspręsta panaudoti DTS (angl. Data Transformation Service). Tai įrankis, kuris yra prie kiekvienos Microsoft SQL serverio licenzijos ir duodamas nemokamai. DTS – tai įrankių rinkinys, kuris naudojamas norint eksportuoti ar importuoti duomenis tarp skirtingų duomenų šaltinių (pvz.: Microsoft SQL server, Microsoft Excel, Oracle DBVS ir pan.). Ryšiai tarp skirtingų duomenų šaltinių yra užtikrinami, pasitelkiant OLE DB – atviro standarto duomenų prieigą.

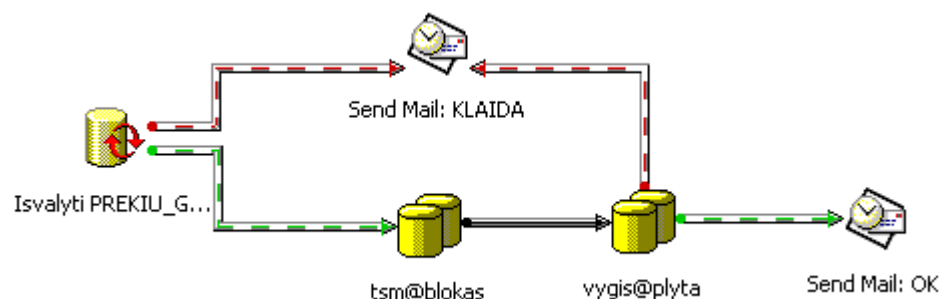
Kartu su Microsoft SQL Server pateikiamas DTS dizaineris leidžia kurti DTS modelį, pasitelkiant grafinius elementus ir juos dėlioiant DTS dizainerio (angl DTS Designer) lange (žr. paveikslėlių apačioje).



25 pav. DTS Designer langas

DTS Designer pagalba, šiame projekte realizuotas reikalingų duomenų perkėlimas iš darbinės duomenų bazės į analizės duomenų bazę.

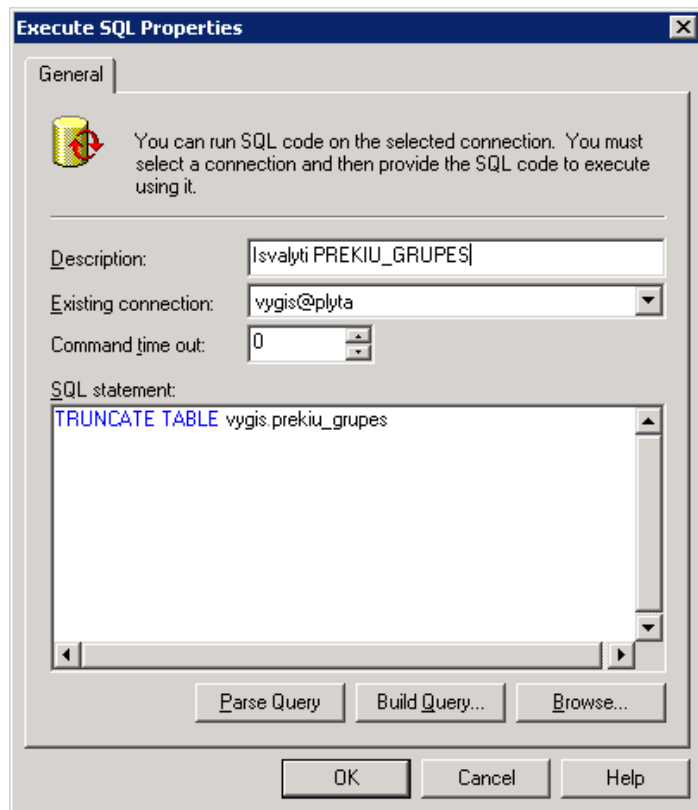
Bendroju atveju, vienos lentelės perkėlimo metu, yra ištrinami seni lentelės duomenys, jie nuskaitomi iš šaltinio duomenų bazės ir išsaugomi analizės duomenų bazės lentelėje. Vienos lentelės perkėlimo DTS paketas pavaizduotas žemiau esančiame paveikslėlyje.



26 pav. Vienos lentelės duomenų atnaujinimo DTS paketas

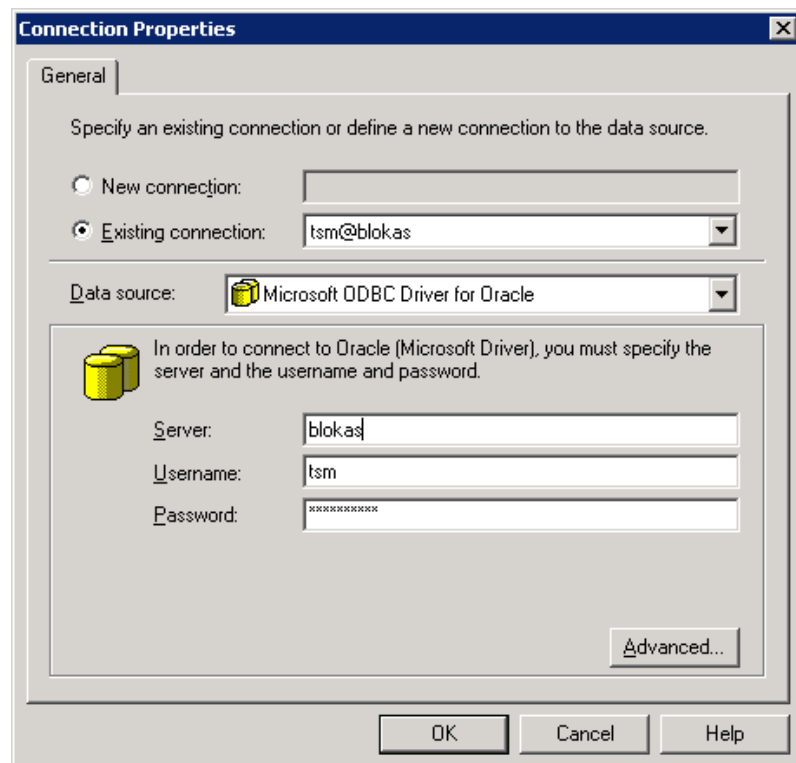
Vienos lentelės duomenų atnaujinimo procesą galima išskaidyti į kelis etapus:

1. Pirmiausia yra pašalinami esamos lentelės duomenys.



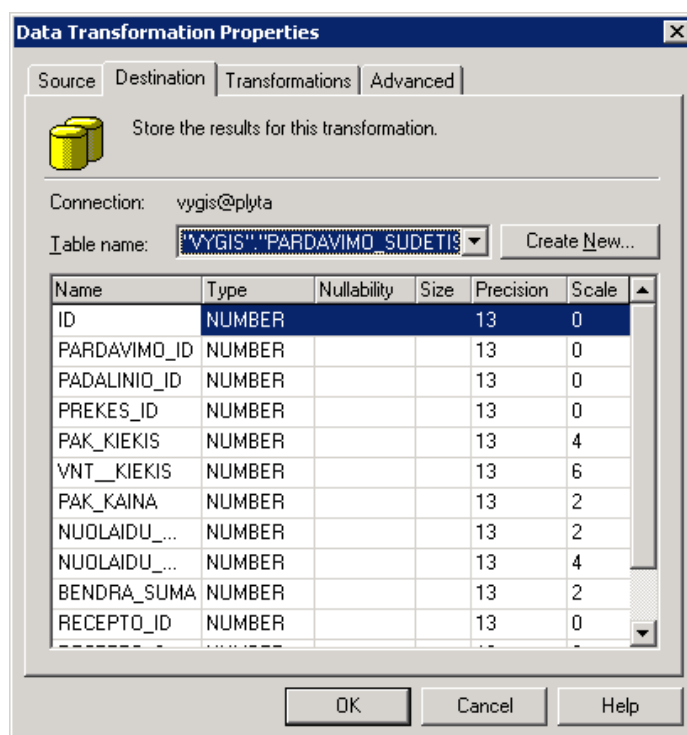
27 pav. Lentelės duomenų išvalymas

2. Nuskaitomi duomenys iš TSM Oracle duomenų bazės serverio.



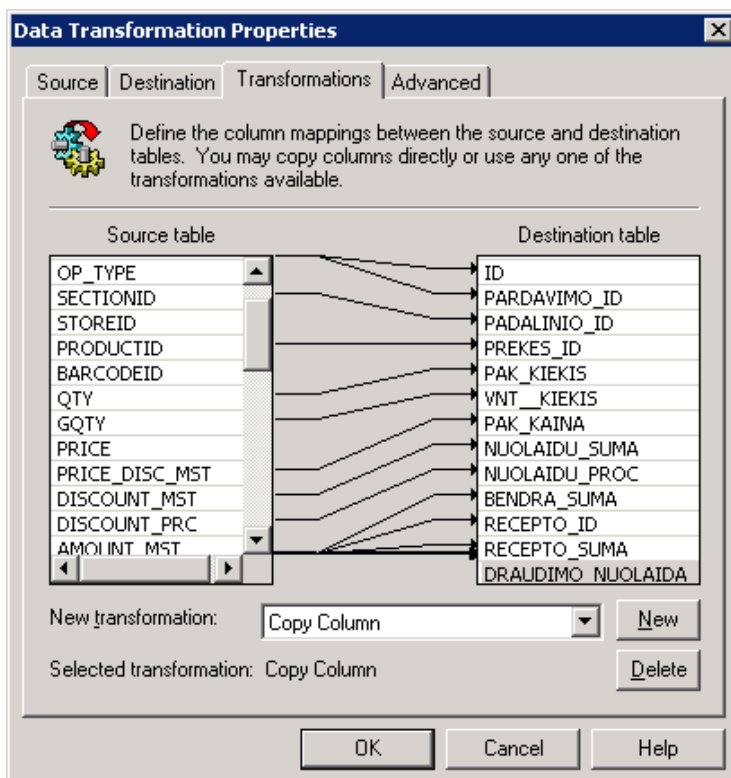
28 pav. Prisijungimas prie pagrindinės DB

3. Prisijungiama prie analizės DB serverio.



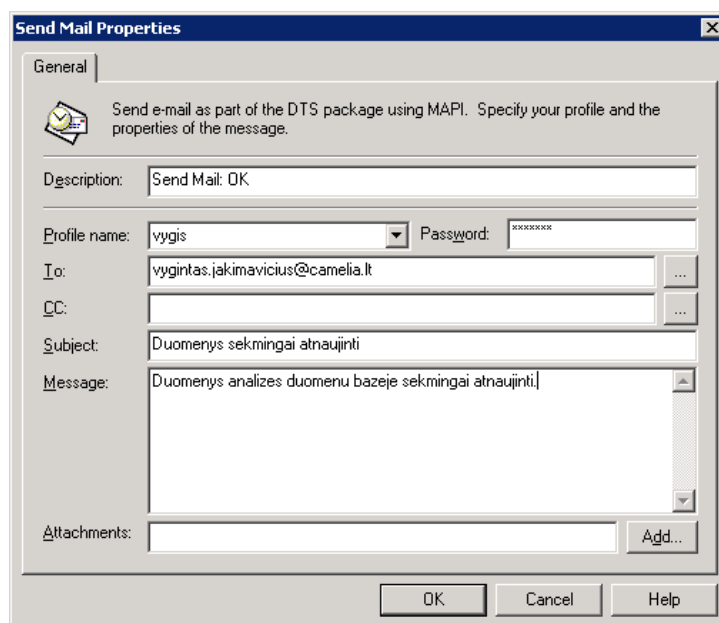
29 pav. Prisijungimas prie analizės DB

4. Nuskaityti duomenys perkeliama į analizės duomenų bazės lentelę.



30 pav. Duomenų perkėlimas iš šaltinio į analizės duomenų bazės lentelę

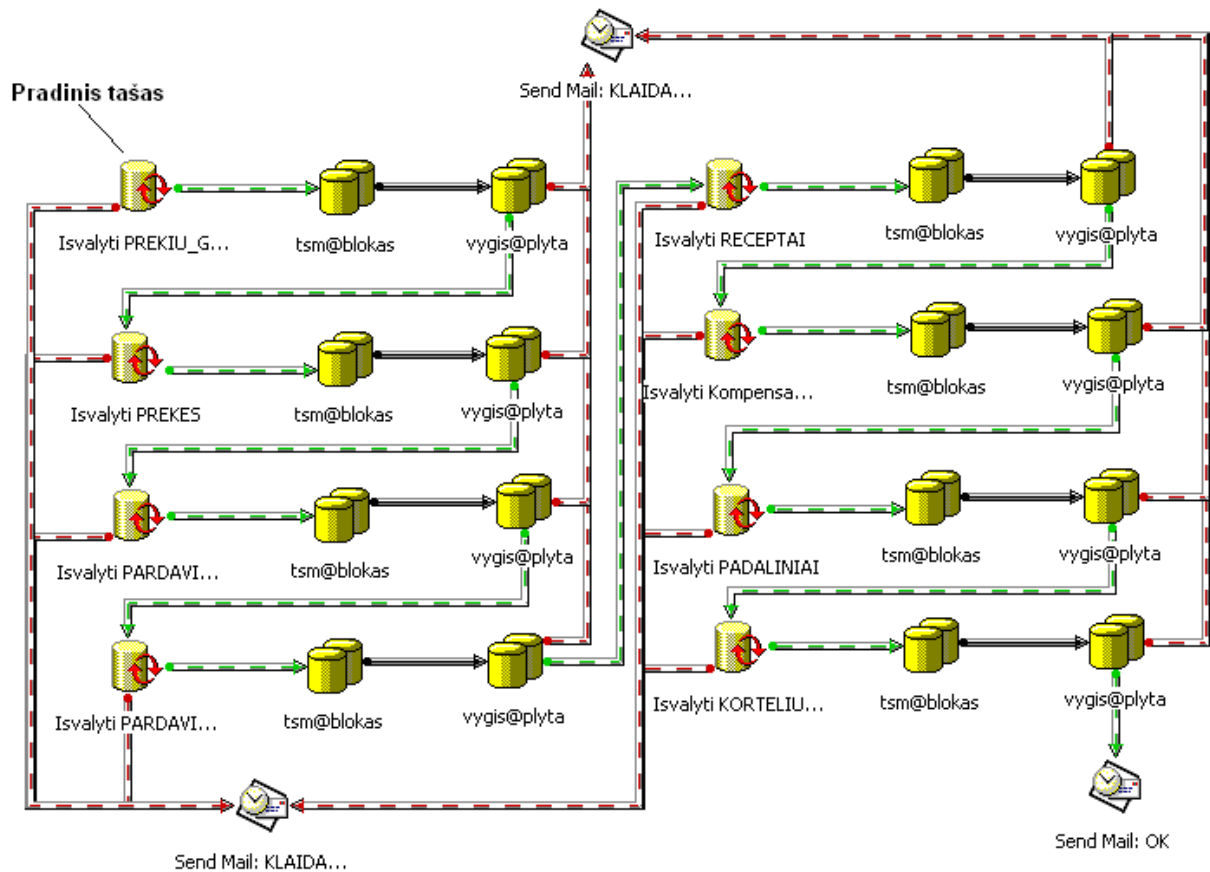
5. Administratoriui išsiunčiamas elektroninis laiškas apie sėkmingą duomenų atnaujinimą.



31 pav. Žinutė administratoriui apie sėkmingą duomenų atnaujinimą.

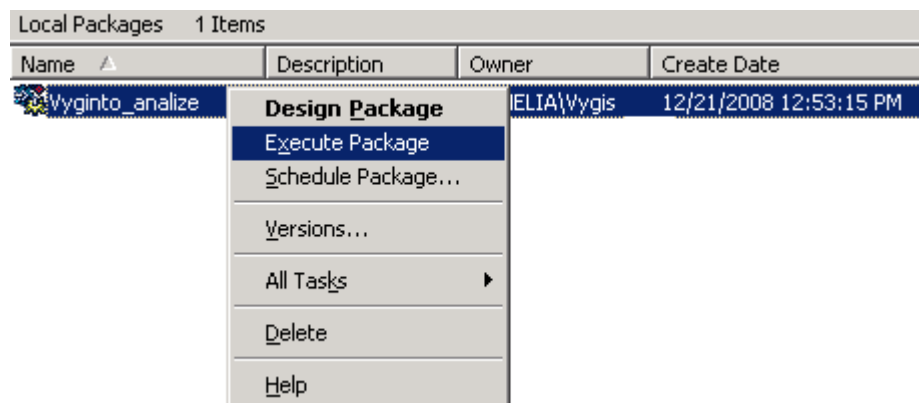
Duomenų atnaujinimo procese numatyta, jog įvykus klaidai, atnaujinimo procesas nutraukiamas ir administratoriui išsiunčiama elektroninė žinutė apie klaidą. Sėkmės atveju, sistemos administratorius apie tai taip pat informuojamas.

Visų analizės duomenų bazės lentelių duomenų atnaujinimo DTS paketas parodytas žemiau esančiame paveikslėlyje.



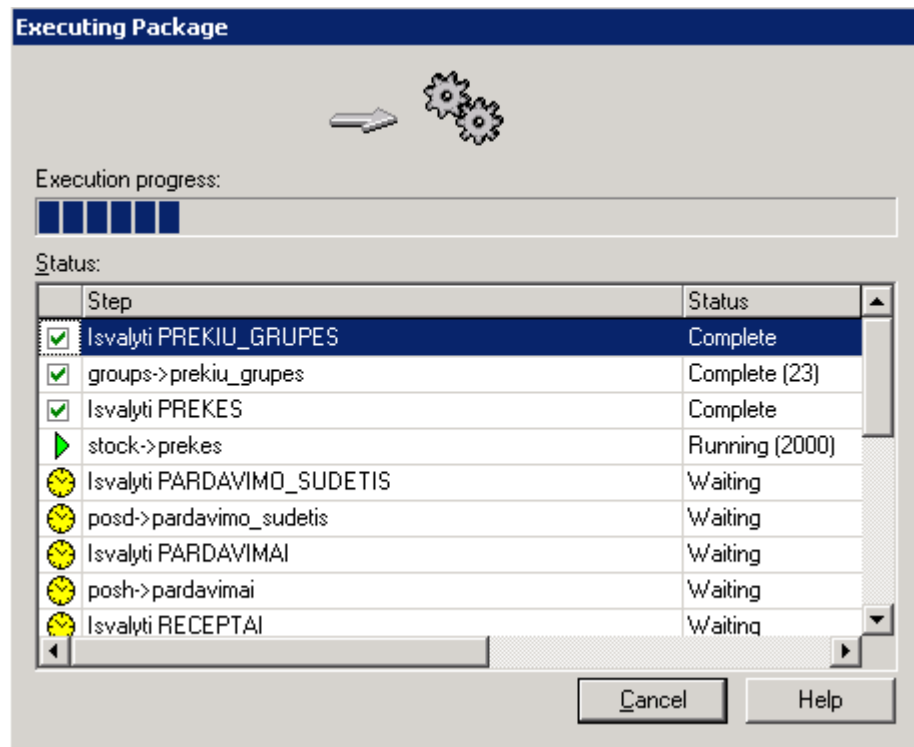
32 pav. Analizės duomenų bazės lentelių atnaujinimo DTS paketas

Paleidus šį DTS paketą, SQL Server Enterprise Manager parodo nurodytų veiksmų eigą ir progresą.



33 pav. DTS paketo paleidimas

Vykiant DTS paketą, SQL Server Enterprise Manager rodo paketo vykdymo eigą.



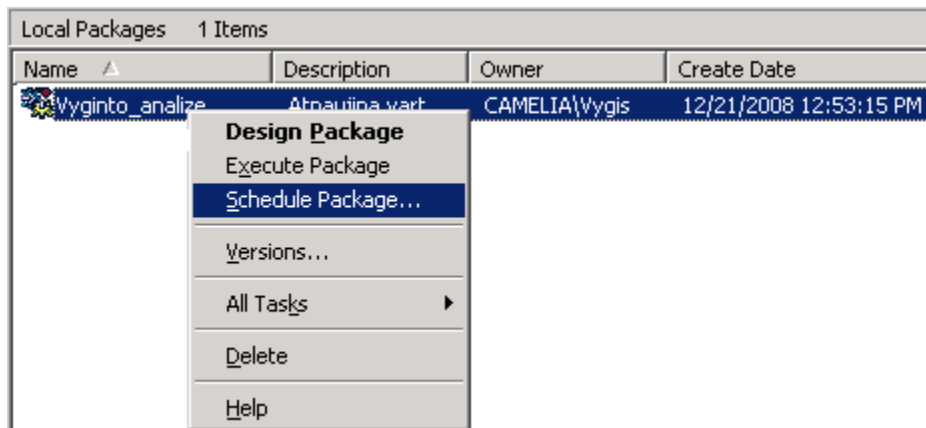
34 pav. DTS paketo vykdymo progresas

Visų analizės duomenų bazės lentelių duomenų atnaujinimo ir visas analizės kubo formavimo procesas detaliau aprašytas sekančiame skyrelyje (Analizės kubo formavimo procesas).

4.2.4. Duomenų atnaujinimo automatizavimas

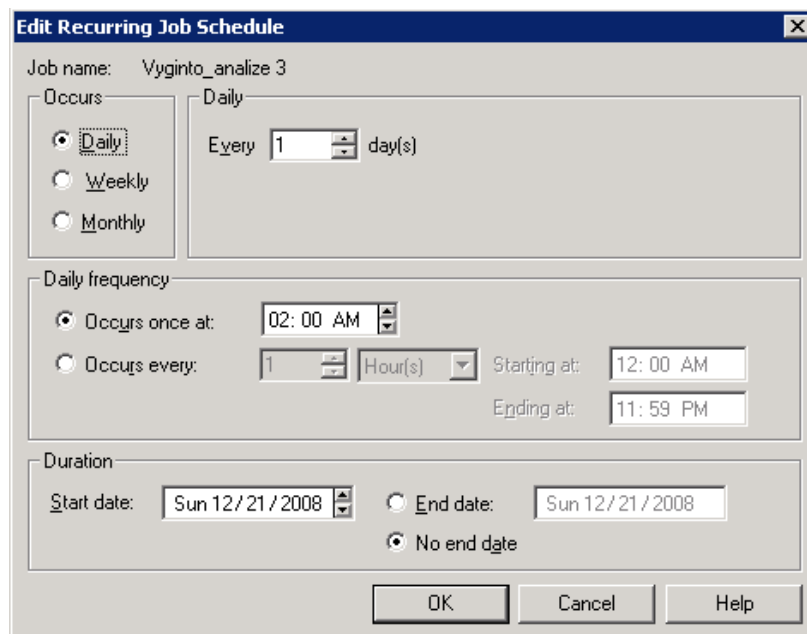
Analizei kirtų duomenų savalaikis atnaujinimas yra svarbus procesas, užtikrinantis, jog vartotojui bus pateikta naujausia informacija. Reikalavimų analizės metu buvo nustatyta, jog visiškai pakankamas duomenų atnaujinimo periodas – kiekvieną naktį.

Tam tikslui bus pasitelktas anksčiau sukurtas DTS paketas, perkeliantis duomenis iš centrinės į analizės duomenų bazę, kuris Microsoft SQL Server 2000 priemonių pagalba bus suplanuotas kasdieniniam paleidimui naktį. Taigi, sukurtam DTS paketui sukuriame paleidimo planą.



35 pav. Duomenų atnaujinimo planavimas – 1 žingsnis

Toliau nustatome periodiškumą, kuriuo bus atnaujinami analizės duomenų bazės duomenys. Mūsų atveju – tai duomenų atnaujinimas kiekvieną naktį, 02:00 val.



36 pav. Duomenų atnaujinimo planavimas – 2 žingsnis

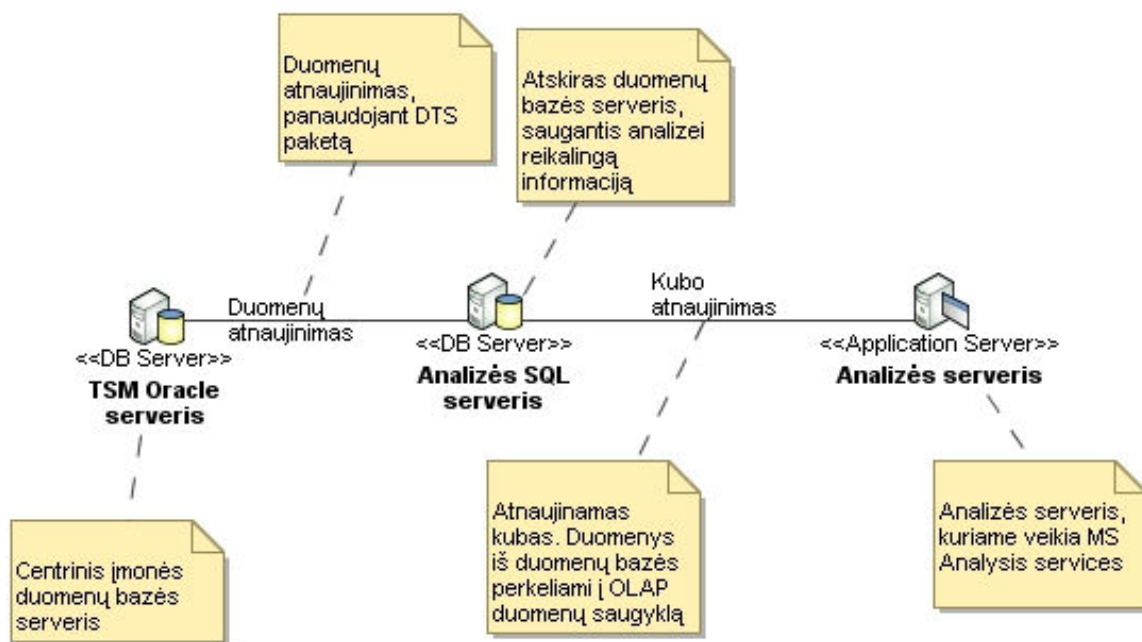
Nustačius duomenų atnaujinimo parametrus, atsiranda įrašas apie nurodytą užduotį SQL Server Agent->Jobs lange.

Name	Category	Enabled	Runnable	Sched...	Status	Last Run St...	Next Run Date
Vyginto_analize	[Uncategorized (Local)]	Yes	Yes	Yes	Not Running	Unknown	(12/22/2008 2:00:00 AM)

37 pav. Duomenų atnaujinimo planavimas – atnaujinimo įrašas

4.2.5. Analizės kubo formavimo procesas

Analizės kubas formuojamas keliais etapais. Tai tam tikra procesų seka. Kiekvieno etapo metu vyksta duomenų išgavimas iš šaltinio sistemas ir tų duomenų įkėlimas į kitą sistemą. Bendras analizės kubo formavimo procesas pavaizduotas žemiau esančiame paveikslėlyje.



38 pav. Kubo formavimo procesas

Analizės kubo formavimo procesas yra kiek sudėtingesnis, kadangi analizei skirtiems duomenims saugoti yra skirtas atskiras SQL duomenų bazės serveris. Pirmojo etapo metu, panaudojant DTS (ang. Data Transfer Services), analizei reikalingi duomenys iš darbinės sistemos yra nukopijuojami į analizės MS SQL duomenų bazės serverį:

- Išvalomos senos lentelės
- Nuskaitomi nauji duomenys iš TSM Oracle duomenų bazės serverio
- Nuskaityti duomenys įrašomi į analizės MS SQL duomenų bazės serverį

Sekantis žingsnis – tai faktų lentelės suformavimas:

- Ištrinama esama faktų lentelė
- Iš naujo suformuojama faktų lentelė
- Apskaičiuojamos tarpinės reikšmės
- Rezultatai išsaugomi analizės duomenų bazės serverio faktų lentelėje

Paskutinis analizės kubo formavimo etapas – tai kubo atnaujinimas:

- Atnaujinamos analizės kubo dimensijos
- Duomenys iš duomenų bazės serverio perkeliama į OLAP duomenų saugyklą
- Suskaičiuojamos agregacijos

Kubo atnaujinimo etapas yra realizuotas Microsoft Analysis Services pakete. Kubo atnaujinimo etapai čia atliekami automatiškai, paleidus kubo atnaujinimo procesą. Priklausomai nuo poreikių, galima atnaujinti kubą trejopai:

- Laipsniškas atnaujinimas (angl. Incremental update) – leidžia atnaujinti kubą, nekeičiant esamų kubo duomenų, o tikrai pridėdamas prie kubo atsiradusius naujus duomenis. Šis kubo atnaujinimo procesas yra vienas paprasčiausių ir greičiausių – neperskaičiuojamos esančios agregacijos, or pridėdamos tikrai naujos – tai sumažina kubo atnaujinimui reikalingą laiką.
- Duomenų atnaujinimas (angl. Refresh data) – šio kubo atnaujinimo metu yra perskaičiuojamos visos kubo reikšmės ir atnaujinami duomenys. Šio kubo atnaujinimo būdo privalumas yra tai, jog vartotojai gali naudotis kubu netgi jo atnaujinimo metu – taigi, esant poreikiui, bet kada galima atnaujinti kubo duomenis.
- Vusiškas atnaujinimas (angl. Full Process) – tai visiškas kubo atnaujinimas. Šis kubo atnaujinimas reikalingas tada, kai pakeičiama pati kubo struktūra ir atsiranda arba panaikinamos tam tikros dimensijos, reikšmės ir pan. Šio atnaujinimo metu kubu naudotis vartotojai negali.

4.2.6. Analizės modelių metaduomenų eksportas naudojant XML standartą

Microsoft SQL Server 2000 naudoja OIM (angl. Open Informatikon Model) metaduomenims apie visus duomenų tipus saugoti – lenteles, laukus, sąryšius, transformacijas. Analysis Services metaduomenys taip pat atitinka OIM standartą.

Tačiau vien turėti analizės modelio metaduomenis neužtenka – juos reikia mokėti gauti, panaudoti. Microsoft Analysis Services turi priemones, leidžiančias sukurtus OLAP analizės kubus eksportuoti bei importuoti į/iš XML standartą atitinkančius dokumentus. Tai yra ypač naudinga – sukurtus OLAP duomenų analizės kubus galima eksportuoti į kitas sistemas, kur juos

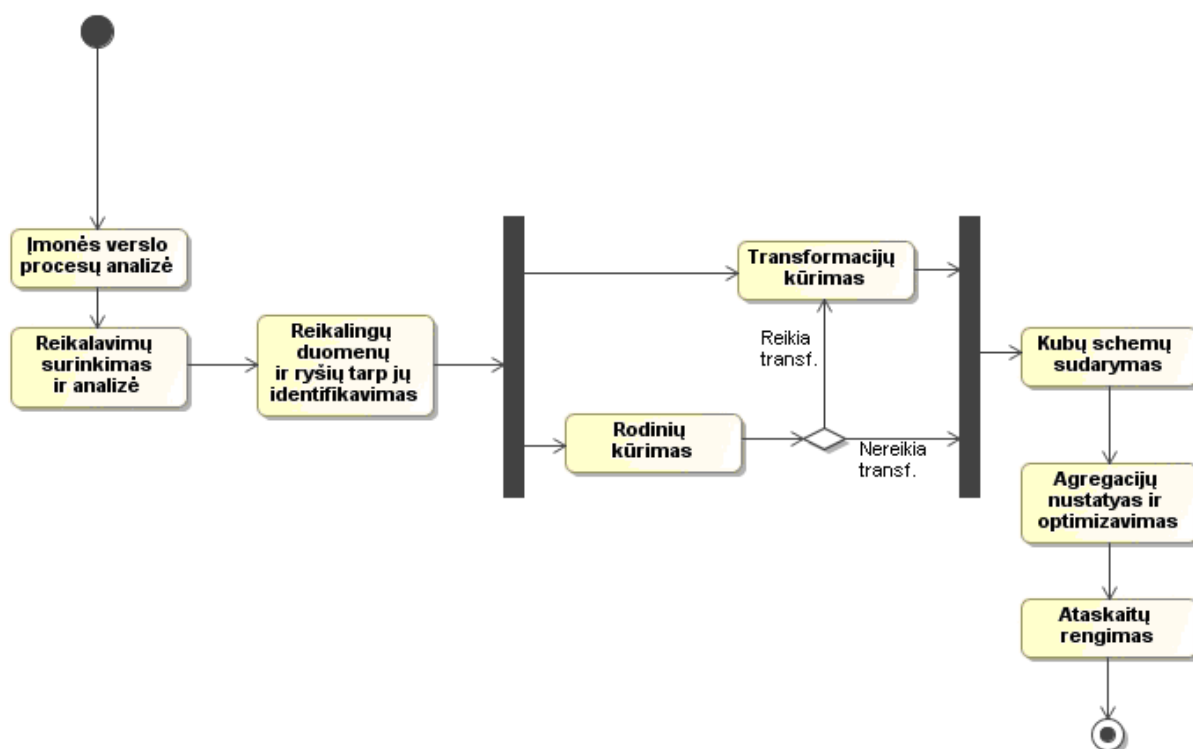
toliau galima apdoroti, redaguoti kitų įrankių pagalba. Taip pačiai galimas ir atvirkštinis variantas – OLAP duomenų kubus galima importuoti į Microsoft Analysis Services.

Sukurtuo įmonės klientų duomenų analizės kubo XML aprašas pateiktas prieduose.

4.2.7. Analizės kubų formavimo procesas

Šio darbo tikslas yra ne tik sukurti konkrečią analizės sistemą ir ją integruoti į įmonės veiklos procesus, bet ir sukurti OLAP analizės kubų formavimo proceso modelį, kuris padėtų panašiuose projektuose, leistų lengviau numatyti galimus sunkumus, sprendimus ir technologijas.

Supaprastintas analizės kubų formavimo proceso modelis, sukurtas remiantis literatūros šaltiniuose pateikiamais metodais, pateiktas žemiau esančiame paveikslėlyje.



39 pav. Analizės kubo formavimo procesas

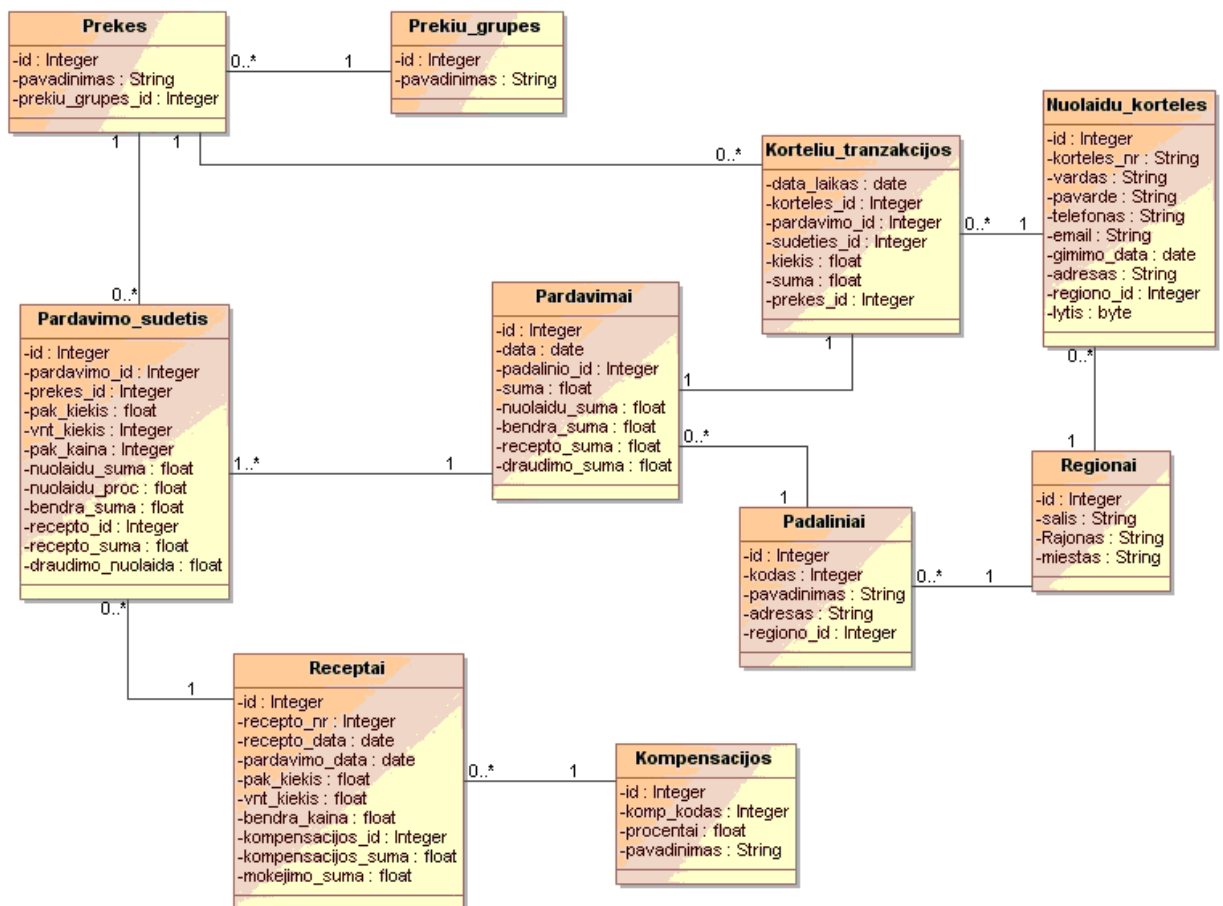
Kiekvienas šio proceso etapas gali būti detalizuotas išsamiai. Dalis šio proceso etapų yra detalizuoti šio projekto skyriuose, kita dalis bus detalizuota sekančių projektų metu.

4.3. Duomenų bazės schema

UAB „Nemuno vaistinė“ įmonės klientų analizės sistema naudoja duomenis, sukauptus verslo valdymo sistemos TSM. Verslo valdymo sistema TSM apima pardavimų

ilgalaikio/trumpalaikio turto apskaitos bei kitas funkcijas. Iš viso, TSM verslo valdymo sistemos duomenų bazę sudaro 600 įvairių lentelių (visos jos išardintos prieduose).

Įmonės klientų analizei naudojami sukaupti pardavimų duomenys. Kadangi projekto analizės etapo metu buvo iškeltas reikalavimas kuo mažiau apkrauti esamą duomenų bazę, buvo pasirinktas sprendimas įmonės klientų analizei paskirti atskirą duomenų bazės serverį, kurio duomenų bazėje būtų saugomi tik tie duomenys, kurie naudojami analizės duomenų saugykloje. Dėl to sumažėjo pati analizei skirtos duomenų bazės dalis. Analizės duomenų bazės loginė schema ir ryšiai tarp lentelių pavaizduoti 40 paveikslėlyje.



40 pav. Analizės duomenų bazės schema

5. Eksperimentinis sistemos tyrimas ir vertinimas

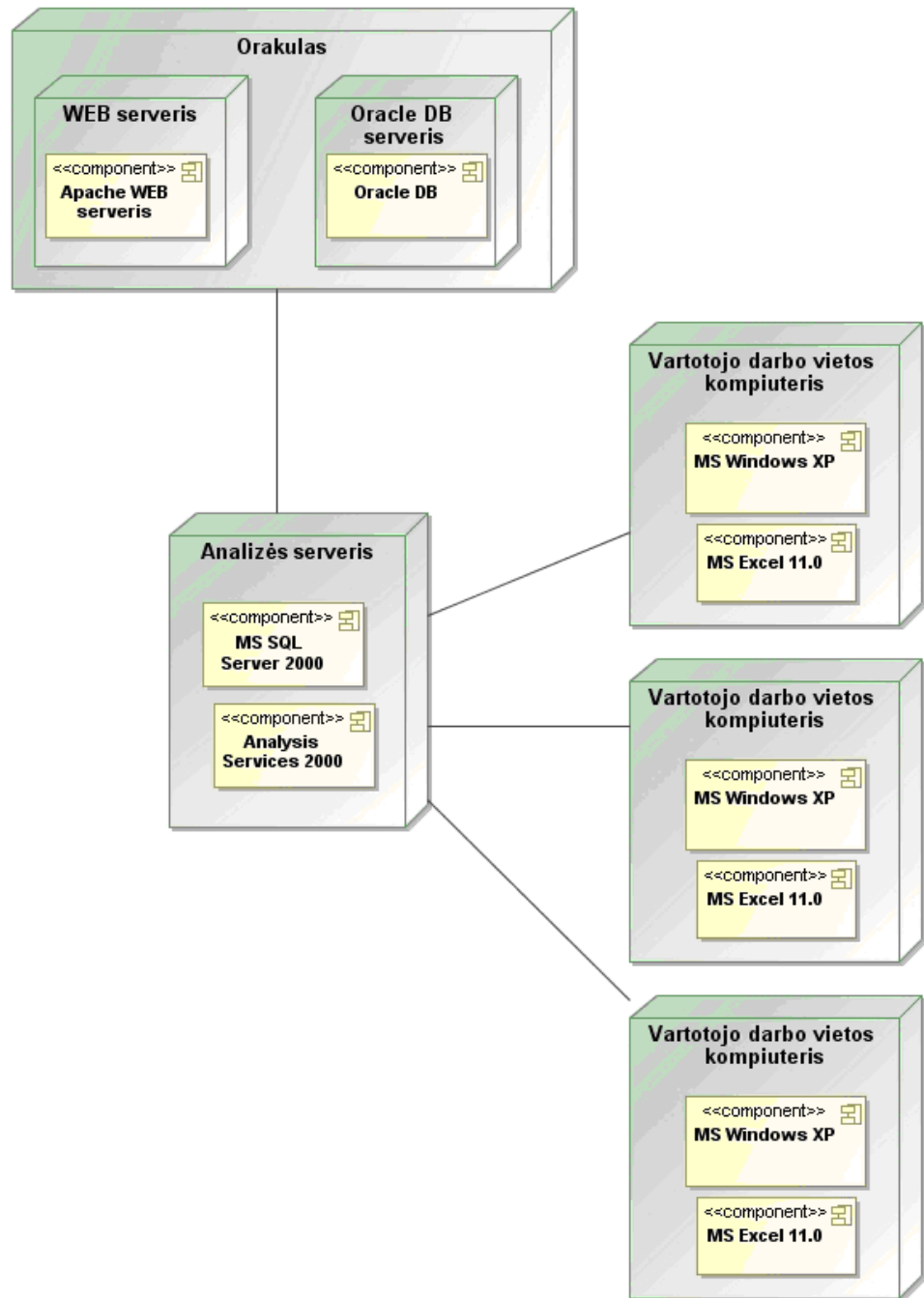
Darbo metu sukurtas įmonės klientų analizės kubo veikimas yra testuojamas ir analizuojami jo veikimo parametrai (atnaujinimo laikas, užklausų generavimo į analizės kubą laikai or pan.).

5.1. Eksperimentinio diegimo aprašymas

UAB „Nemuno vaistinė“ įmonės klientų analizės sistema naudoja duomenis, sukauptus verslo valdymo sistemos TSM. Verslo valdymo sistema TSM apima pardavimų ilgalaikio/trumpalaikio turto apskaitos bei kitas funkcijas. Iš viso, TSM verslo valdymo sistemos duomenų bazę sudaro 600 įvairių lentelių (visos jos išvardytos prieduose).

5.2. Sistemos diegimas

Kuriama įmonės klientų analizės sistemą sudaro du duomenų bazių serveriai, analizės serveris bei vartotojų darbo vietos. Sistemos komponentai ir jų tarpusavio saryšiai pavaizduoti žemiau esančiame paveikslėlyje.



41 pav. Sistemos komponentai ir jų sąryšiai

Sistemą sudaro šie komponentai:

- Oracle DB serveris. Šiame duomenų bazės serveryje veikia Oracle DBVS, kurioje saugomi visi duomenys. Taip pat tame pačiame serveryje veikia ir programos sąsajai skirtas Alache WEB serveris.
- Analizės duomenų bazėje veikianti Oracle DBVS. Šioje duoenų bazėje saugoma informacija, skirta vien įmonės klientų analizei. Ši duomenų bazė veikia atskirai

nuo pagrindinės dėl pagrindinio serverio apkrovimo. Veikiant analizės duomenų bazei atskirame serveryje, mažiau apkraunama pagrindinė duomenų bazė.

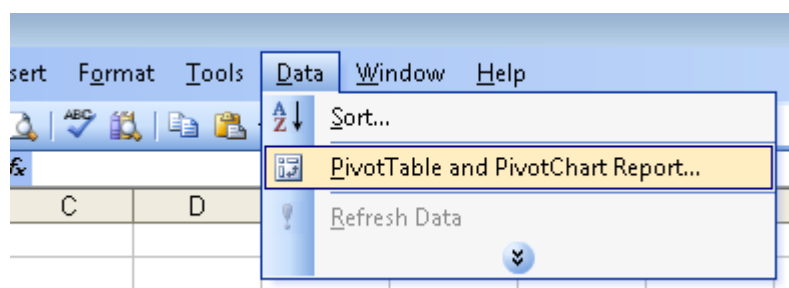
- Analysis Services 2000, esantys analizės serveryje – tai servais, skirti darbui su OLAP kubais. Šie servais naudoja duomenis, esančius OLAP duomenų saugykloje, juos apdoroja, atnaujina, ir leidžia naudotis vartotojams.
- Microsoft Excel 11.0 – programų paketas, kurio pagalba vartotojai gali dirbti su OLAP kubais. Šio programų paketo pagalba, vartotojai gali vykdyti norimas užklausas, generuoti ataskaitas, diagramas, peržiūrėti OLAP serverio pateikiamus duomenis.

5.3. Sistemos naudojimo instrukcija

Įmonės klientų duomenų analizės kubui peržiūrėti naudojamas Microsoft Excel programinės įrangos paketas. Microsoft Excel pagalba yra nesudėtinga peržiūrėti kubo duomenis, rengti ataskaitas, grafikus, vartotojui turint pagrindinius įgūdžius darbo su Microsoft Excel programiniu paketu. Microsoft Excel programų paketas turi specialiai darbui su analizės kubais skirtą modulį - Pivot Table. Šio įrankio pagalba, galima peržiūrėti duomenų kubo duomenis, rengti ataskaitas, kurti grafikus ir pan. Taip pat šis programinės įrangos paketas kubo informacijai peržiūrėti pasirinktas dėl patogumo: visi įmonės vartotojai, kuriems reikia dirbti su analizės kubais, turi šį programinės įrangos paketą – taigi, nereikia papildomų išlaidų vartotojo sąsajos kūrimui, projektavimui.

Toliau trumpai apžvelgiama darbo su OLAP analizės kubu darbo eiga, pasinaudojant Microsoft Excel moduliu Pivot Table.

1. Pasileidus Microsoft excel, pasirenkame meniu punktą *Data->Pivot Table and PivotChart Report*.



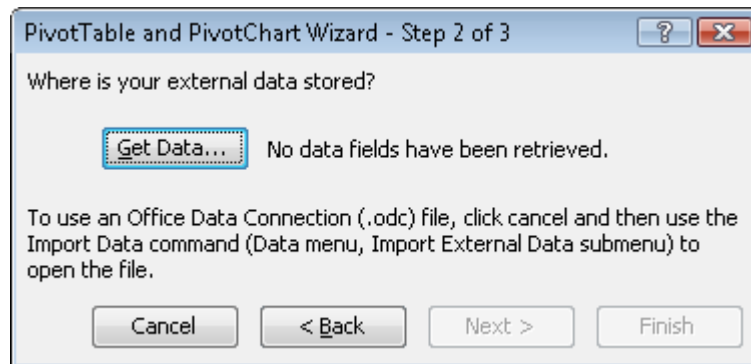
42 pav. PivotTable and PivotChart Report komponentas

2. Atsidariusiame lange, pasirenkamas *External Data Source* pasirinkimas. Toliau spaudžiama *Next*.



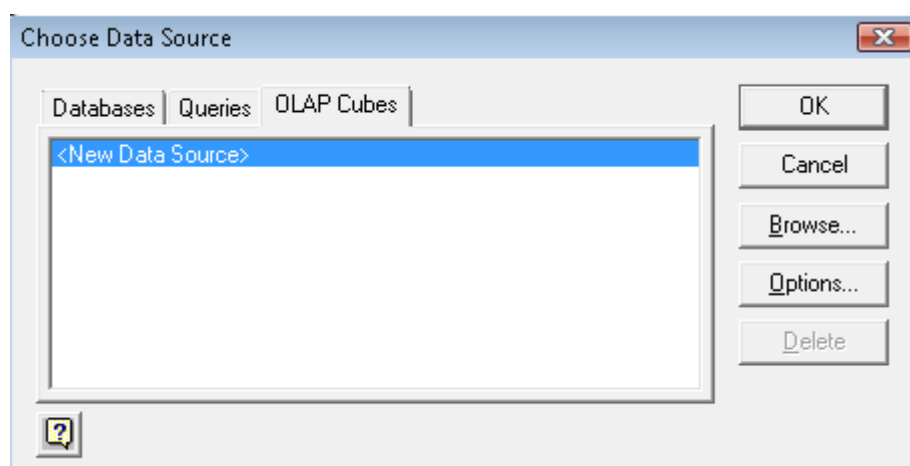
43 pav. External Data Source pasirinkimas

3. Spaudžiamas mygtukas *Get Data*.



44 pav. Išorinio šaltinio pasirinkimas

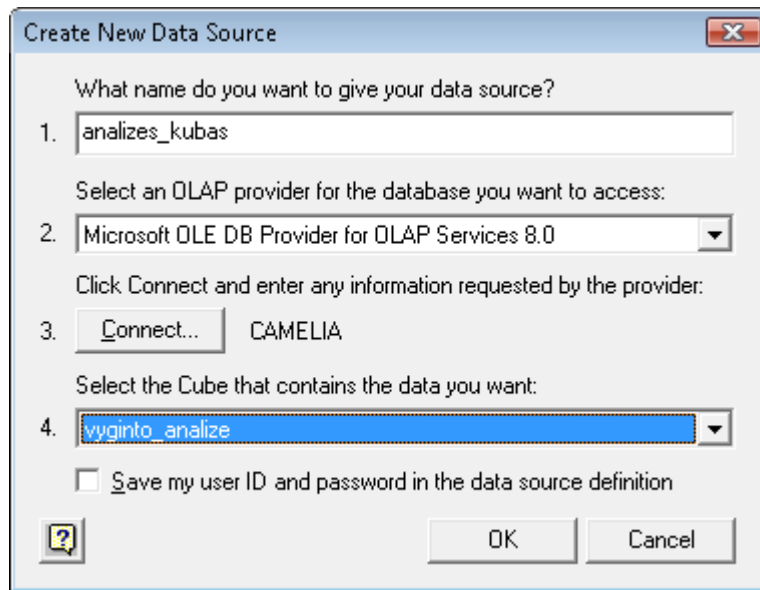
4. Atsidariusiame lange, pasirenkama *OLAP Cubes* kortelė, pasirenkama *<New Data Source>* ir spaudžiamas mygtukas *OK*.



45 pav. New Data Source duomenų šaltinio paikimas

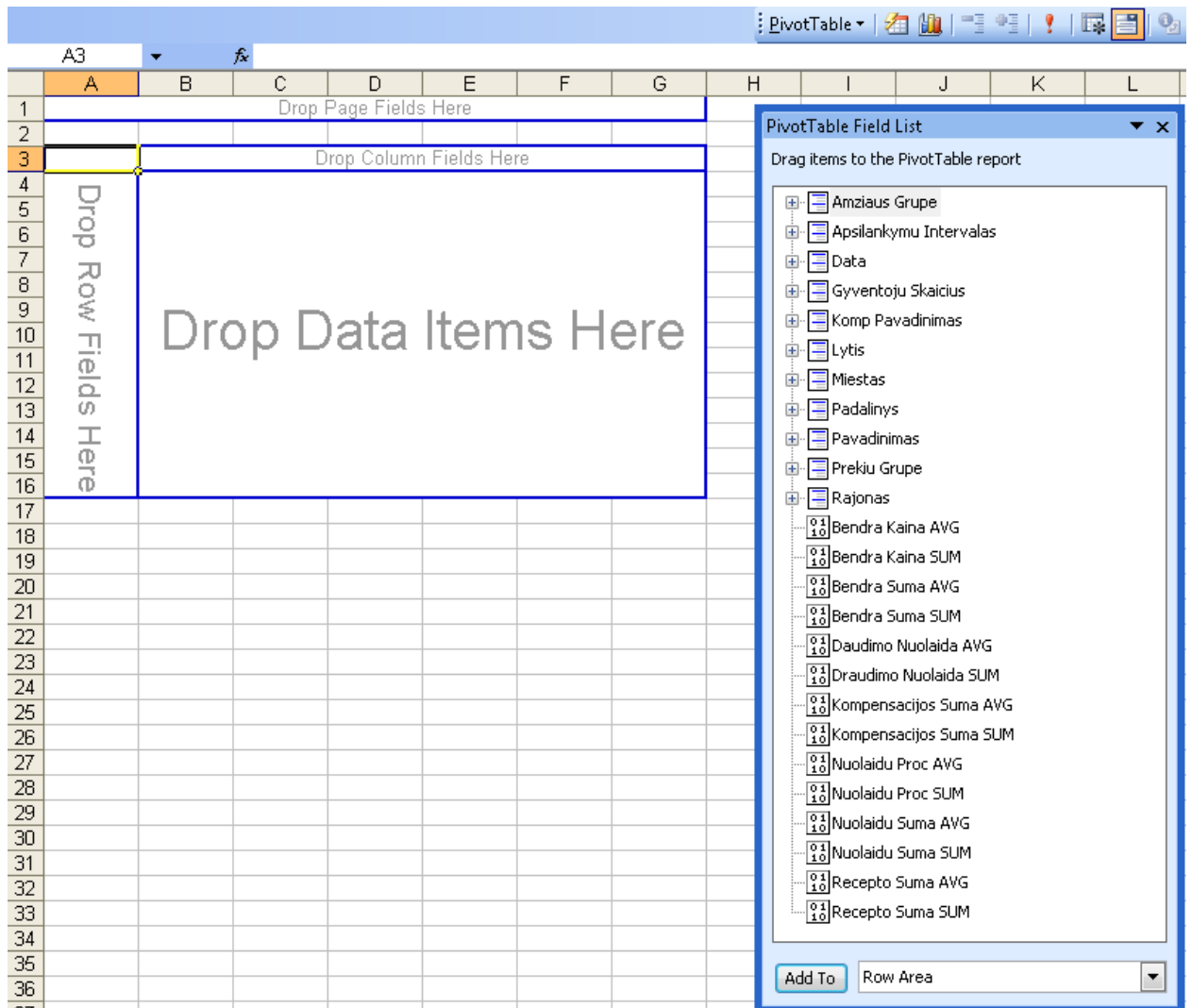
5. Atsidariusiame naujame lange, įvedamas naujo kubo pavadinimas ir pasirenkamas *Microsoft OLE DB Provider for Olap Services 8.0* tiekėjas. Toliau spaudžiamas mygtukas *Connect*, toliau atsiradusiame lange nurodomas

OLAP serverio adresas, pasirenkama duomenų bazė ir pasirenkamas norimas duomenų kubas.



46 pav. Analizės kubo pasirinkimas

6. Pasirinkus analizės kubą ir paspaudus *OK* mygtuką, *Microsoft Excel* pateikia darbo su analizės kubu langą.



47 pav. Darbo su analizės kubu lagas MS Excel aplinkoje

Viską atlikus, MS Excel priemonėmis galima kurti norimas ataskaitas, grafikus. Keletas sukurtų ataskaitų pavyzdžių pateikta žemiau esančiuose paveikslėliuose. Visų ataskaitų duomenys yra paimti nuo 2008-12-01 iki 2008-12-10.

Pirmajame ataskaitos pavyzdyje galima matyti vidutinę vienos prekės, kurią perka tam tikro amžiaus klientai, suskirstyti pagal tai, koku dažnumu apsilanko vaistinėje.

Amžiaus Grupe	Apsilankymu Intervalas	Bendra suma AVG
		12.94075949
		12.94075949
---		12.56565391
	1-2 men	12.80973684
	3-4 men	11.89
	iki 1 men.	13.53925721
	iki 1 sav.	8.794051724
	vienkartinis	12.426875
18-36		5.473333333
	iki 1 men.	5.473333333
37-44		10.66
	iki 1 sav.	10.66
45-62		13.64071429
	iki 1 men.	13.64071429
63+		18.16257143
	iki 1 men.	19.91384615
	iki 1 sav.	14.14
	vienkartinis	4.81
Grand Total		12.58398443

48 pav. Ataskaitos pavyzdys - vidutinė vienos prekės kaina pagal amžiaus grupę bei apsilankymo intervalą

Sekančioje ataskaitoje galima matyti išleistas sumas vaistams pagal prekių grupes miestuose pagal gyventojų skaičių.

Bendra Suma SUM	Prekiu Grupe					Grand Total
Gyventoju Skaicius	A1	B1	C1	L1	R1	Grand Total
0-6000		144.07	307.55	1680.73	20.88	2153.23
120000-300000	283.43	448.6	1721.29	1995.15	880.06	5328.53
300000-500000		57.34	922.05	2215.91	233.48	3428.78
500000+		15.17	1034.42	236.08	202.13	1487.8
50000-120000		141.86	559.29	871.71	8.65	1581.51
6000-15000		74.11	256.9	413.64	247.1	991.75
8000-50000		87.2	290.1	465.67	164.07	1007.04
Grand Total	283.43	1159.92	5731.53	8614.89	1991.4	17781.17

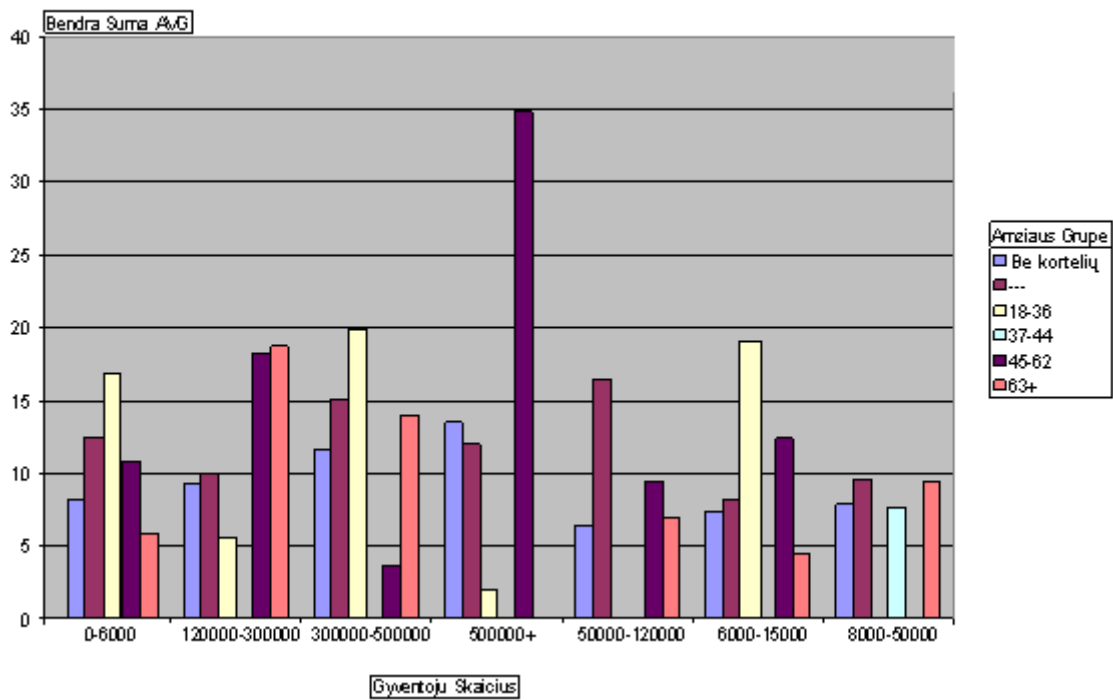
49 pav. Ataskaitos pavyzdys – išleistos sumos vaistams pagal prekių grupes bei gyventojų tankumą

Dar vienoje žemiau esančioje ataskaitoje galima matyti klientų, suskirstytų pagal lytį, bendrai gautas kompensacijų sumas pagal kompensacijos rūšis.

Komp Pavadinimas	Lytis	Kompensacijos Suma	SUM
			0
	-		0
	M		0
	V		0
Invalidui			560.65
			39.13
	-		276.2
	M		245.32
Kita			712.06
	-		496.78
	M		215.28
Kitos			3733.33
			260.09
	-		2013.65
	M		1371.91
	V		87.68
Liga pagal sąrašą			1606.97
	-		1120.87
	M		486.1
Pensininkui			2.38
	-		2.38
Grand Total			6615.39

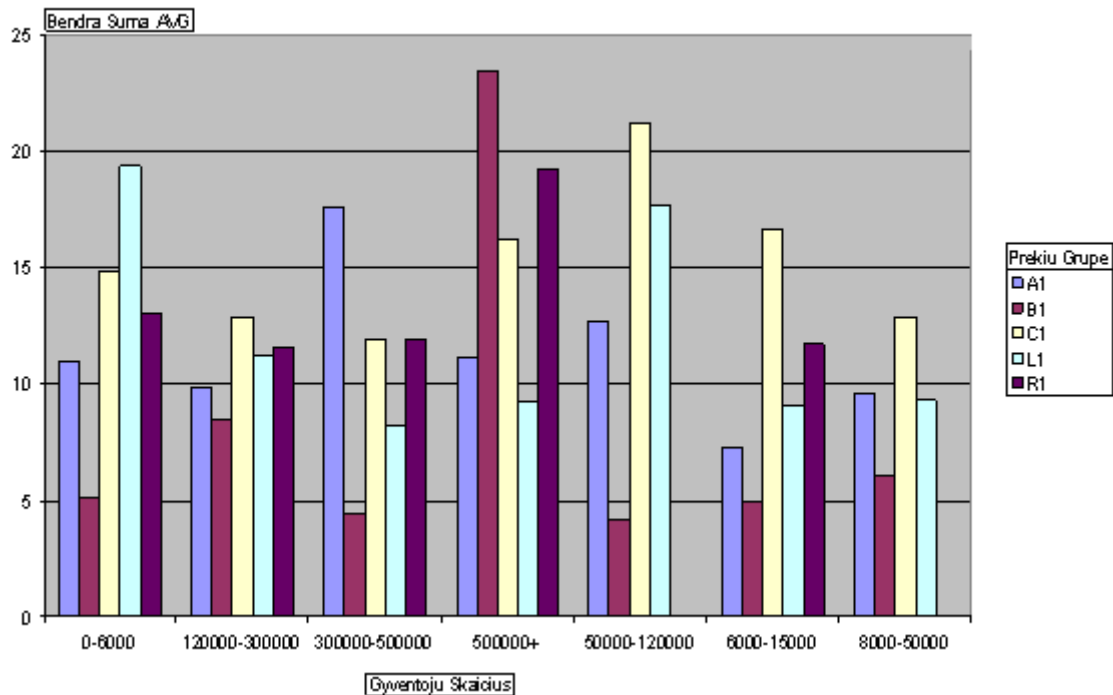
50 pav. Ataskaitos pavyzdys – gautos kompensacijos pagal lytį ir kompensacijos rūšį

Be ataskaitų rengimo, Microsoft Excel priemonėmis galima braižyti įvairius grafikus, kurie leidžia vaizdžiai ir suprantamai vartotojui pateikti informaciją. Žemiau esančiame paveikslėlyje pateiktas grafikas, kuriame galima matyti, kokio dydžio miestuose ar miesteliuose pagal gyventojų skaičių, klientai daugiausia išleidžia pinigų medicininėms prekėms, o kur – mažiausiai.



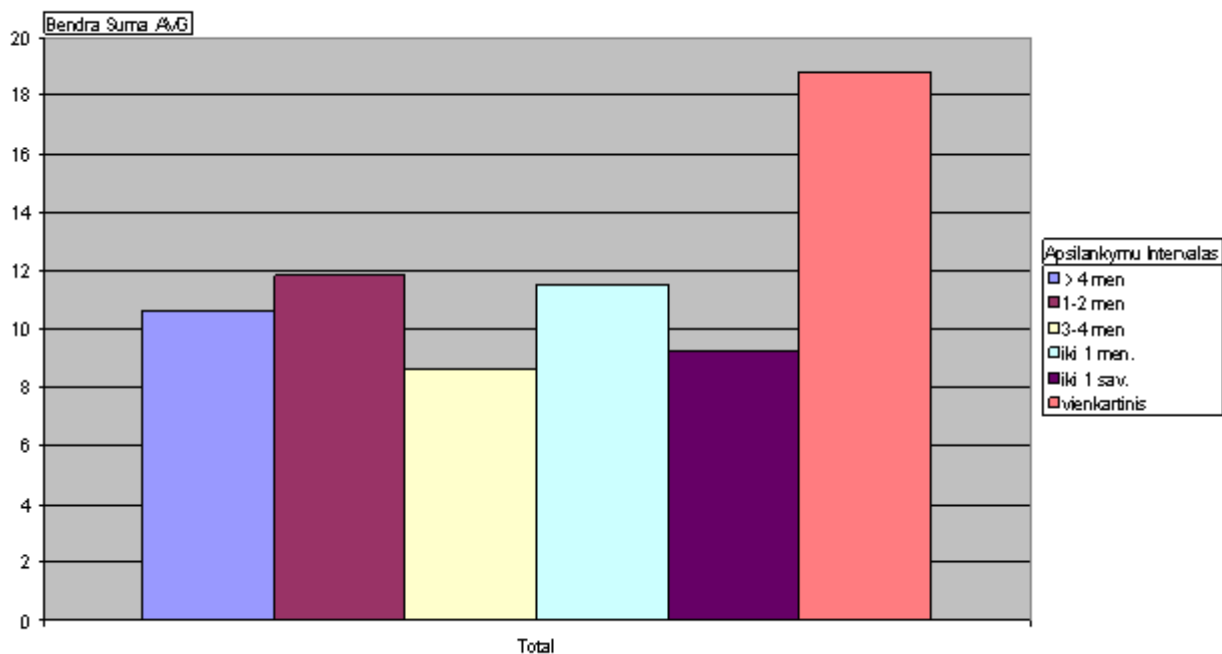
51 pav. Diagramos pavyzdys – išleistos vidutinės sumos dydis pagal gyventojų skaičių mieste ir amžiaus grupę

Žemiau esančiame paveikslėlyje pavaizduota pavyzdinė diagrama, kurioje galima matyti vidutines išleidžiamas sumas tam tikroms prekių grupėms pagal gyventojų skaičių mieste.



52 pav. Diagramos pavyzdys – išleistos vidutinės sumos dydis pagal gyventojų skaičių mieste ir prekių grupę

Sekančioje diagramoje galima matyti klientų apsilankymo vaistinėje ir išleidžiamos vidutinės sumos diagramą. Galima išvelgti tendenciją, jog klientai, vieną kartą pasinaudoję kortele, išleidžia vidutiniškai didžiausią sumą pinigų. Dažniausiai (kas savaitę) apsilankantys klientai – mažiausiai pinigų išleidžia apsipirkimo metu.



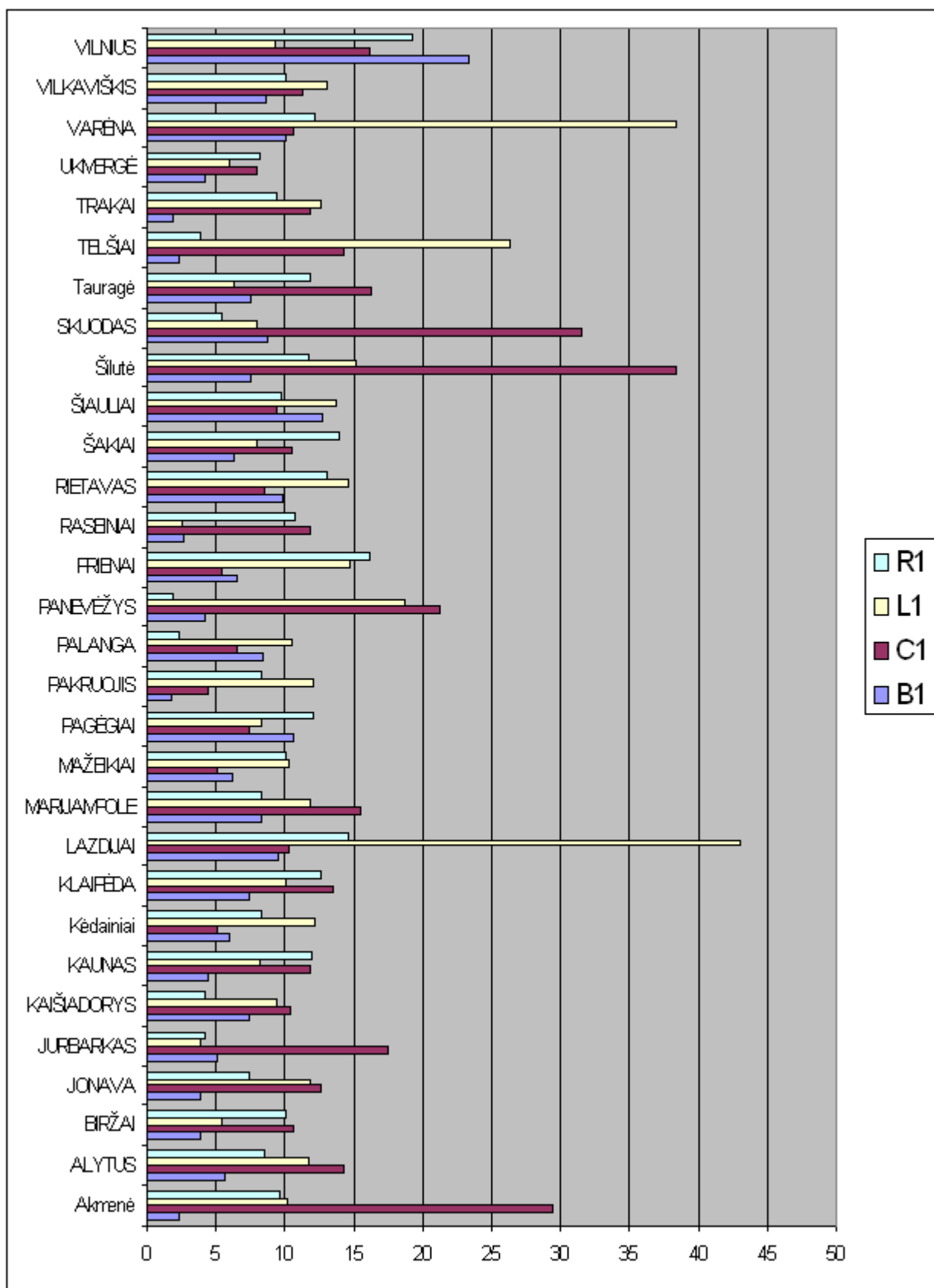
53 pav. Diagramos pavyzdys – išleista pinigų suma pagal apsilankymo vaistinėje intervalą

Nemažiau svarbi yra pardavimų analizė pagal regionus, prekių grupes. Viena iš tokių sugeneruotų ataskaitų pateikta 54 paveikslėlyje. Šioje ataskaitoje galima matyti vidutinės išleistas sumas atskiruose regionuose kiekvienai prekių grupei.

Bendra Suma AVG	Prekiu Grupe				
Rajonas	B1	C1	L1	R1	Grand Total
Akmenė	2.31	29.38	10.17222	9.61	12.86805556
ALYTUS	5.67	14.22	11.78	8.54	10.0525
BIRŽAI	3.86	10.59	5.44	10.11	7.5
JONAVA	3.865	12.57143	11.885	7.44	8.940357143
JURBARKAS	5.11625	17.4781	3.912857	4.22	7.681800595
KAIŠIADORYS	7.46125	10.41	9.34875	4.22	7.86
KAUNAS	4.3852	11.88677	8.221034	11.90111	9.098529947
Kėdainiai	5.96	5.14	12.12	8.32	7.885
KLAIPĖDA	7.456888889	13.45121	10.04388	12.5755	10.8818704
LAZDIJAI	9.54	10.28333	43.07	14.55	19.36083333
MARIJAMPOLE	8.33	15.44	11.885	8.32	10.99375
MAŽEIKIAI	6.24	5.11625	10.33	10.12	7.9515625
PAGĖGIAI	10.6	7.46125	8.32	12.07	9.6128125
PAKRUOJIS	1.76	4.3852	12.09	8.32	6.6388
PALANGA	8.4	6.55	10.50333	2.34	6.948333333
PANEVĖŽYS	4.175	21.19588	18.72294	1.83	11.48095588
PRIENAI	6.5698	5.44	14.758	16.15	10.72945
RASEINIAI	2.6	11.885	2.59	10.75333	6.957083333
RIETAVAS	9.8876	8.52	14.55	13.05	11.5019
ŠAKIAI	6.299375	10.496	7.964286	13.92	9.669915179
ŠIAULIAI	12.75272727	9.393636	13.67484	9.774545	11.39893695
Šilutė	7.54	38.37	15.11	11.78	18.2
SKUODAS	8.725	31.54	7.972857	5.44	13.41946429
Tauragė	7.472857143	16.255	6.340909	11.885	10.48844156
TELŠIAI	2.34	14.26	26.31	3.912857	11.70571429
TRAKAI	1.83	11.885	12.585	9.34875	8.9121875
UKMERGĖ	4.175	7.95	5.975	8.221034	6.580258621
VARĖNA	10.12	10.66	38.37	12.12	17.8175
VILKAVIŠKIS	8.6	11.24	13.05	10.04388	10.73347015
VILNIUS	23.39125	16.19649	9.2445	19.23167	17.01597579
Grand Total	6.914439943	13.32168	12.87801	9.670589	10.69618196

54 pav. Ataskaitos pavyzdys - išleistos vidutinės sumos pagal regionus bei prekių grupes

Šios ataskaitos grafinis vaizdas pateiktas 55 paveikslyje.



55 pav. Diagramos pavyzdys - išleistos vidutinės sumos pagal regionus bei prekių grupes

5.4. Sistemos veikimo ir savybių analizė, kokybės kriterijų įvertinimas

Kalbant apie duomenų saugyklas, duomenų saugyklų kokybė ir efektyvumas dažnai yra tiesiogiai susiję su teikiamų duomenų kokybe. Norint išmatuoti ir užtikrinti šių duomenų kokybę, reikia išmatuoti daugelį įvairių parametrų bei rodiklių, apibūdinančių duomenų saugyklos darbą.

Yra nurodoma daug įvairių duomenų saugyklų kokybės kriterijų: sistemų našumas, saugumas, pasiekiamumas, patogumas ir pan. Dauguma šių kriterijų negali būti pamatuoti objektyviai – todėl vertinti duomenų saugyklą tokiais kriterijais yra gana sudėtinga ir pakankamai netikslu.

Norint objektyviai įvertinti duomenų saugyklas, reikia remtis konkrečiais, išmatuojamais kriterijais. Tokius kriterijus galima suskirstyti į dvi pagrindines grupes:

- Vidinės duomenų saugyklos metrikos. Tai dažniausiai gana tiksliai išmatuojami kriterijai (serverio procesoriaus apkrovimas, atminties panaudojimas, sistemos prieinamumo duomenys ir pan.). Pasitelkiant vidinės metrikos rodiklius, galima gana nesunkiai nustatyti, ar tenkinami, ar ne duomenų saugyklos nefunkciniai reikalavimai (sistemos pasiekiamumas, našumas ir pan.).
- Išorinės metrikos. Tai rodikliai, nurodantys vartotojų naudojimosi sistema duomenų statistiką. Pasitelkiant išorinės metrikos rodiklius, galima atsakyti į įvairius klausimus (Kas naudoja sistemą? Kaip dažnai sistema yra naudojama? Kokių vartotojų sistema yra dažniausiai naudojama? ir pan.).

Siekiant įvertinti sukurtą sistemą, reikia atsakyti į kelis svarbiausius klausimus:

1. Išmatuoti sukurtos sistemos darbo bei jos našumo rodiklius.
2. Nustatyti, ar sistema atitinka visus analizės etapo metu iškeltus funkcinius ir nefunkcinius reikalavimus.

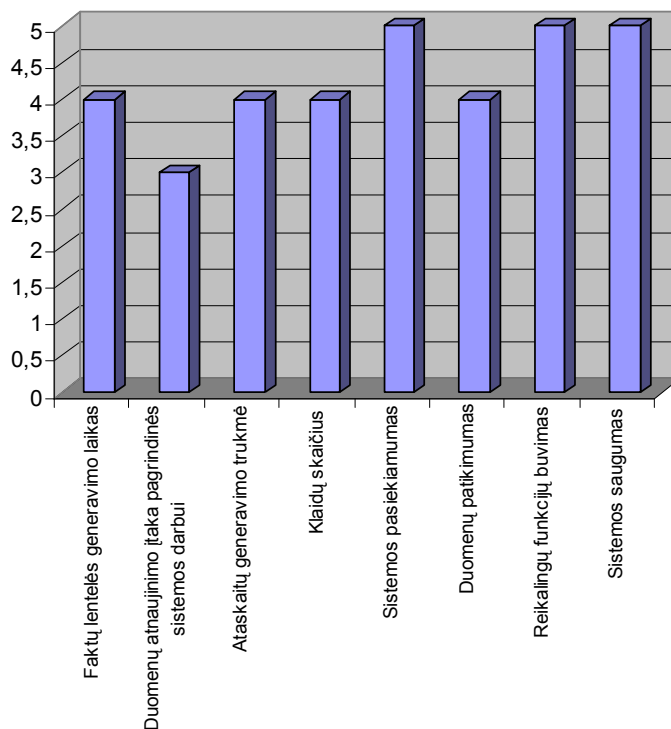
Sistemos kokybę įvertinti pasitelkta kriterijų ir jų įvertinimo lentelė (15 lentelė). Visi kriterijai vertinami penkiabalėje sistemoje, kur 5 balai atitinka 5vertinimą puikiai – t.y. kriterijus pilnai atitinka vartotojų lūkesčius, 1 – blogai – kriterijus neatitinka iškeltų reikalavimų).

15 lentelė. Sistemos kokybės įvertinimų lentelė

Kriterijus	Balai	Komentarai
Faktų lentelės generavimo laikas	4	Faktų lentelės generavimas vidutiniškai trunka 45 minučių. Kadangi duomenys paprastai atnaujinami naktį, sistemos vartotojus toks rodiklis pilnai tenkina.
Duomenų atnaujinimo įtaka pagrindinės sistemos darbui	3	Duomenų atnaujinimo metu, pagrindinės duomenų bazės apkrovimas vidutiniškai padidėja 15%. Įvertinant duomenų apimtį,

		rodiklis vertinamas patenkinamai.
Ataskaitų generavimo trukmė	4	Naudojimo metu, norima ataskaita sugeneruojama vidutiniškai 7 sekundes. Rodiklis vertinamas labai gerai.
Klaidų skaičius	4	Sistemą eksploatuojant vieną mėnesį, pastebėta apie 15 duomenų klaidų. Visos klaidos – dėl klaidingų duomenų pagrindinėje sistemoje. Rodiklis įvertintas labai gerai.
Sistemos pasiekiamumas	5	Sistemą eksploatuojant vieną mėnesį, sistema visuomet veikė ir buvo pasiekiamas, jokių sutrikimų nepastebėta. Rodiklis vertinamas puikiai.
Duomenų patikimumas	4	Sistemą eksploatuojant vieną mėnesį, vieną kartą pastebėtas duomenų neatitikimas dėl nepavykusio duomenų atnaujinimo nakties metu. Sutrikimas pašalintas tą pačią dieną, vėliau įvesta pranešimų sistemų administratoriui sistema apie duomenų atnaujinimo progresą. Rodiklis įvertintas gerai.
Reikalingų funkcijų buvimas	5	Sistema pilnai atitinka esamų vartotojų reikalavimus. Rodiklis įvertintas labai gerai.
Sistemos saugumas	5	Saugumas pilnai užtikrinamas, pasitelkiant Windows Active Directory priemones ir atitinka iškeltus reikalavimus. Rodiklis įvertintas labai gerai.

Sistemos kokybės kriterijai taip pat pavaizduoti 56 paveikslėlyje esančioje stulpelinėje diagramoje.



56 pav. Sistemos kokybės parametrų įvertinimo diagrama

5.5. Sistemos taikymo rekomendacijos

Šio darbo metu sukurta sistema – tai vaistinių tinklo klientų analizės modelis, leidžiantis analizuoti įmonės klientų įpročius, perkamas prekes, lankomumo statistiką ir pan.

Iki šiol UAB „Nemuno vaistinė“ buvo naudojamos klientų nuolaidų kortelės. Klientui tai galimybė nusipirkti reikalingus vaistus pigiau, o įmonei – klientų lojalumas. Tačiau realiai jokios sistemos, leidžiančios peržiūrėti įmonės klientų elgseną, lojalumą, nebuvo įdiegta. Taigi, be sukurto sistemos, nuolaidų kortelės akivaizdžios naudos nedavė.

Sukurtoji sistema leidžia pardavimų analitikams analizuoti įmonės klientų elgseną, apsipirkimo įpročius, išleidžiamas vidutines sumas vaistams, klientų lojalumą, prekių paklausą priklausomai nuo metų laiko, vietovės, klientų amžiaus.

Sukurtas klientų analizės modelis yra daugiau skirtas vaistinių tinklams, nes yra naudojama dalis specifinių duomenų, tokių kaip kompensacijų rūšys, draudimo nuolaidos ir pan. Farmacija – gana specifinė mažmeninės rinkos dalis, kur be prekių grupių, akcijų, yra ir kompensuojamųjų vaistų receptai, draudimo atvejai bei kiti specifiniai atvejai. Tad sukurta sistema ir skirta būtent vaistinių tinklui.

5.6. Tolimesnio sistemos plėtojimo, tobulinimo galimybės

Kuriant įmonės klientų analizės sistemą, buvo keli pagrindiniai kūrimo etapai – tai būsimų sistemos vartotojų reikalavimų išgavimas, jų analizė, sistemos projekto kūrimas, analizė, sistemos realizavimas, parametrų stebėjimas, optimizavimas bei kokybės tyrimas.

Sukurtos sistemos optimizavimas – svarbus etapas, siekiant pagerinti sukurtos sistemos savybes, sumažinti kitų sistemų apkrovimą, paspartinti veikimą ir sukurti dar priimtinesnę sistemą vartotojui.

Taip pat apibendrinus sistemos rodiklius, galima teigti, jog sukurtoji sistema pilnai tenkina vartotojų poreikius. Sukurtoji sistema minimaliai veikia pagrindinės sistemos darbą, sistemos vartotojai to netgi nepajuto.

Kalbant apie sukurtos sistemos tobulinimo galimybes, sukurtąją sistemą planuojama tobulinti, ją papildant funkcionalumu, optimizuojant duomenų atnaujinimo mechanizmą, pasitelkiant naujus verslo analizės sistemų įrankius.

Verslo analizės sritis paskutiniaisiais metais yra ypač sparčiai plėtojama, sukurti verslo analizės sistemų įrankiai yra nuolat tobulinami, ieškoma naujų sprendimų.

6. Išvados

1. Vaistinių tinklo duomenų analizės poreikių tyrimas parodė, kad farmacija skiriasi nuo įprastos mažmeninės prekybos, kadangi šioje srityje yra specifinių duomenų: kompensuojamųjų vaistų receptai, draudimo nuolaidos ir kt., todėl jų analizei reikia sukurti individualių duomenų saugyklos modelį, pritaikytą būtent vaistinių tinklo pardavimų ir klientų analizės poreikiams.

2. Įmonėje egzistuojančios priemonės šių duomenų analizei yra neišvystytos, nelanksčios, jų naudojimas apkrauna visą sistemą ir mažina jos našumą.

3. Literatūros šaltinių analizė parodė, kad duomenų saugyklų kūrimo procesas nėra universalus ir jį reikia pritaikyti konkrečioms atvejams, todėl šiame darbe sukurtas duomenų modelis ir atitinkama programinė įranga vaistinių klientų duomenų analizei.

4. Didžiausios problemos saugyklų kūrime yra ETL efektyvumas ir duomenų kokybės užtikrinimas. Atlikta duomenų saugyklų kūrimo priemonių analizė parodė, kad nors pagrindinė DB yra Oracle, saugyklai tikslinga pasirinkti Microsoft SQL Server priemones dėl jų didesnio tinkamumo. Kadangi šie įrankiai įmonėje jau egzistuoja, dėl to sumažėja bendros projekto išlaidos.

5. Kuriamam įmonės klientų analizės kubui pasirinkta MOLAP tipo duomenų saugykla, kadangi MOLAP tipo saugyklos leidžia efektyviai vykdyti užklausas ir greitai atrinkti duomenis. Tai optimalus duomenų saugyklos modelis atminties išnaudojimo ir veikimo efektyvumo atžvilgiais. Visi skaičiavimai yra atliekami kubo sukūrimo metu ir naudojami užklausų vykdymo metu. Kubas yra atnaujinamas kiekvieną naktį, todėl atnaujinimas mažiausiai veikia pagrindinės sistemos darbą.

6. Sukurta sistema leidžia analizuoti pardavimų duomenis, įmonės klientų įpročius, lojalumą. Pagal analizės rezultatus galima kurti specialius pasiūlymus įmonėms klientams, išlaikyti esamus klientus ir pritraukti naujus..

7. Įgauta patirtis ir žinios verslo duomenų analizės srityje bus panaudota ateityje, kuriant kitas verslo duomenų analizės sistemas.

7. Literatūra

- [1] H-J. Lenz , A. Shoshani. Summarizability in OLAP and Statistical Data Bases.
- [2] W. H. Inmon. OLAP And Data Warehouse. 2000.
- [3] Ralph Kimball, Margy Ross. The Data Warehouse Toolkit.
- [4] Vivant Technologies white paper. The Evolution of ETL: From Hand-coded ETL to Tool-based ETL. 2007.
- [5] Embarcadero Technologies white paper. An Open ETL Architecture for Enterprise Data Integration.
- [6] T.B. Pedersen. Aspects of data modeling and query processing for complex multidimensional data. *PhD thesis, Faculty of Engineering and Science, Aalborg University, Danija*, 2000, 180 p., prieiga per internetą
<http://www.aub.aau.dk/phd/department08/text/pedersen-torben.pdf>.
- [7] Embarcadero Technologies white paper. An Open ETL Architecture for Enterprise Data Integration.
- [8] Donald Feinberg, Mark A. Beyer. Magic Quadrant for Data Warehouse Database Management Systems, 2006
- [9] Arun Sen and Atish P. Sinha. Toward Developing Data Warehousing Process Standards: An Ontology-Based Review of Existing Methodologies.
- [10] Oracle Corporation. Oracle9i OLAP User's Guide, Release 2 (9.2), 2002.
- [11] An Oracle White Paper (July 2007). Oracle Database 11g for Data Warehousing and Business Intelligence.
- [12] JACOBSON R.. Microsoft SQL Server 2000 Analysis Services Step by Step. Microsoft, 2000.
- [13] Eric Thomsen. OLAP Solutions, 2002 John Wiley & Sons Inc.
- [14] Arun Sen and Atish P. Sinha. Toward Developing Data Warehousing Process Standards: An Ontology-Based Review of Existing Methodologies.
- [15] THOMSON E.. OLAP Solutions. Building Multidimensional Information Systems, 2nd. ed. John Wiley & Sons, 2002.
- [16] MICROSOFT. Designing and Implementing OLAP Solutions with Microsoft SQL Server 2000. 2000.
- [17] Microsoft Corp. *Microsoft SQL Server 2000 Analysis Services Performance Guide*. 2003.

- [18] PAUL S., GUATAM N., BALINT R.. Preparing and Mining Data with Microsoft SQL Server 2000 and Analysis Services. Microsoft, 2002.
- [19] An Oracle White Paper (January 2006). Climbing to the OLAP Summit with Oracle Warehouse Builder 10gR2.
- [20] Ralph Kimball, Margy Ross. The Data Warehouse Toolkit, Second Edition, 2002 John Wiley & Sons Inc.

8. Terminų ir santrumpų žodynis

Šiame darbe naudotų terminų bei santrumpų paaiškinimai pateikti žemiau esančioje (16) lentelėje.

16 lentelė. Terminų ir santrumpų žodynis

Santrumpa, terminas	Paaiškinimas
SQL (angl. <i>Structured Query Language</i>)	– struktūrizuota užklausų kalba
DB (angl. <i>Database</i>)	– duomenų bazė.
OLAP (ang. <i>On-Line Analytical Processing</i>)	- priemonė greitai ir patogiai išgauti reikiamus duomenis įvairiais pjūviais
Analysis Services	- Microsoft SQL Server 2000 platformos OLAP komponentas
DBVS	- duomenų bazės valdymo sistema
DTS (ang. <i>Data Transformation Services</i> ,)	- duomenų transformavimo įrankiai
ETL (angl. Extract, Transform and Load)	- išgavimas, transformavimas ir išsaugojimas
Aggregations	- iš anksto paskaičiuojamos OLAP duomenų laukų reikšmės
Windows Active Directory	- Windows tinklų vartotojų autentifikavimo paketas

Pharmacy Data Warehouse Model and Practice

Summary

This work investigates Pharmacy Data Warehouse model and practice. The development and design of data warehouse is faced with various problems. One of the biggest and most time consuming data warehousing design and development issues - is the data extraction from existing systems, transformations and load (ETL process) to target systems.

Analyzed company uses Oracle database as operational data store. Examination of the current situation, the technical base, for the implementation of the data warehouse is selected the Microsoft SQL Server 2000 Analysis Services tools. Because these tools already exist in the company, thus reducing the overall costs needed.

During this work is realized the cube of Pharmacy customers analysis, tested overall system performance, in addition, an example schedules for cube updating are created.

This enables the analysts to more effectively analyze customer behavior, sales and offer customers the most appropriate solutions, while retaining existing customers and attracting new ones.

9. Priedai

1 priedas. Techninė užduotis

Tvirtinu
Vedėjas
Suderinta
Vadovas
Data

TECHNINĖ UŽDUOTIS

1. TEMA:
Vaistinių tinklo duomenų saugyklos modelis ir taikymas
2. ANALITINIS IR TIRIAMASIS DARBAS:
 - 2.1. Užsakovo įmonėje vykdomų procesų analizė.
 - 2.2. Vartotojų poreikių duomenų analizei analizė.
 - 2.3. Integracijos poreikių analizė.
 - 2.4. Nefunkcinių reikalavimų analizė.
 - 2.5. SUPROJEKTUOTI, REALIZUOTI IR PARUOŠTI VARTOJIMUI ĮMONĖS KLIENTŲ OLAP DUOMENŲ ANALIZĖS KUBĄ
3. PARUOŠTI SISTEMOS PANAUDOJIMO DOKUMENTUS:
 - 3.1. Trumpą vartotojo vadovą
 - 3.2. Trumpa atnaujinimų nustatymo vadovą
 - 3.3. Ataskaitų pavyzdžius
4. REIKALAVIMAI PROJEKTAVIMUI, PROGRAMINEI IR TECHNINEI ĮRANGAI:
 - 4.1. Esama duomenų bazė realizuota Oracle ir Microsoft SQL Server 2000 programine įranga
 - 4.2. Kubus realizuoti Microsoft SQL Server 2000 Analysis Services priemonėmis
 - 4.3. Nauja sistema turi lengvai integruotis į esamą įmonės infrastruktūrą; veikti esamuose vartotojų kompiuteriuose, naudoti esamą tinklo apsaugą (Windows Active Directory)
5. REIKALAVIMAI DARBO PRISTATYMOUI:
 - 5.1. Pateikti spausdintą ataskaitos variantą
 - 5.2. Paruošti skaidres

IFM-3/4 grupės studentas
Vygintas Jakimavičius

2 priedas. Pažyma iš įmonės apie sistemos įdiegimą

3 priedas. Realizuotos užklaustos, funkcijos

Šiame priede pateikiama dalis darbo metu sukurtų užklausių, funkcijų bei jų aprašymai.

- Rodinio Vartotojai formavimo užklausa:

```
select
  pard.data,
  --pard.id,
  kort_tranz.korteles_id,
  nuol_kort.apsilankymu_intervalas,
  nuol_kort.amziaus_grupe,
  nuol_kort.lytis,
  pard.padaliny,
  reg.rajonas,
  reg.miestas,
  reg.gyventoju_skaicius,
  prek.pavadinimas,
  prek_gr.prekiu_grupe,
  pard.nuolaidu_suma,
  pard.nuolaidu_proc,
  pard.bendra_suma,
  pard.recepto_suma,
  pard.draudimo_nuolaida,
  rec.bendra_kaina,
  komp.komp_pavadinimas,
  rec.kompensacijos_suma
from

(select
  pardavimai.id,
  pardavimai.data,
  pardavimai.padalinio_id,
  pardavimo_sudetis.prekes_id,
  pardavimo_sudetis.nuolaidu_suma,
  pardavimo_sudetis.nuolaidu_proc,
  pardavimo_sudetis.bendra_suma,
  pardavimo_sudetis.recepto_suma,
  pardavimo_sudetis.draudimo_nuolaida,
  pardavimo_sudetis.recepto_id,
  padaliniai.id as padal_id,
  padaliniai.pavadinimas as padaliny,
  padaliniai.region_id
from
  pardavimai, pardavimo_sudetis, padaliniai
where
  pardavimai.id=pardavimo_sudetis.pardavimo_id
  and
  pardavimai.padalinio_id=padaliniai.id
  and
  pardavimai.data>'2008.09.01'
) pard

left outer join
(select
  id, rajonas, miestas,
  case
    when gyventoju_sk<8000 then '0-6000'
    when gyventoju_sk>=8000 and gyventoju_sk<15000 then '6000-15000'
    when gyventoju_sk>=15000 and gyventoju_sk<50000 then '8000-50000'
    when gyventoju_sk>=50000 and gyventoju_sk<120000 then '50000-120000'
```

```

        when gyventoju_sk >= 120000 and gyventoju_sk < 300000 then '120000-300000'
        when gyventoju_sk >= 300000 and gyventoju_sk < 500000 then '300000-500000'
        when gyventoju_sk >= 500000 then '500000+'
        else '---'
    end as gyventoju_skaicius
from
    regionai) reg
on reg.id = pard.region_id

left outer join
(select
    id, pavadinimas, prekiu_grupes_id
from
    prekes) prek
on prek.id = pard.prekes_id

left outer join
(select
    id, pavadinimas as prekiu_grupe
from
    prekiu_grupes) prek_gr
on prek.prekiu_grupes_id = prek_gr.id

left outer join
(select
    id, bendra_kaina, kompensacijos_id, kompensacijos_suma, mokejimo_suma
from receptai) rec
on rec.id = pard.recepto_id

left outer join
(select
    id, pavadinimas as komp_pavadinimas, procentai
from kompensacijos) komp
on komp.id = rec.kompensacijos_id

left outer join
(select
    korteles_id, pardavimo_id
from
    korteliu_tranzakcijos) kort_tranz
on pard.id = kort_tranz.pardavimo_id

left outer join
(select
    id,
    case
        when gimimo_data >= sysdate - 365 * 18 then '0-18'
        when gimimo_data < sysdate - 365 * 18 and gimimo_data >= sysdate - 365 * 36 then '18-
36'
        when gimimo_data < sysdate - 365 * 37 and gimimo_data >= sysdate - 365 * 44 then '37-
44'
        when gimimo_data < sysdate - 365 * 45 and gimimo_data >= sysdate - 365 * 62 then '45-
62'
        when gimimo_data < sysdate - 365 * 63 then '63+'
        else '---'
    end as amziaus_grupe,
    regiono_id,
    case
        when lytis = 'M' then 'M'
        when lytis = 'V' and vardas = ' ' then '-'

```

```

        when lytis='V' and vardas!=' ' then 'V'
        else '-'
    end as lytis,
    lp_gauti_apsilankymu_intervala(id) as apsilankymu_intervalas
from
    nuolaidu_korteles) nuol_kort
on kort_tranz.korteles_id=nuol_kort.id

```

- Kliento vidutiniam apsilankymų intervalui vaistinėje gauti, yra naudojama atskira sukurta funkcija. Ji pateikta žemiau.

```

CREATE OR REPLACE FUNCTION VYGIS.lp_gauti_apsilankymu_intervala (p_kortelesid
IN NUMBER)
RETURN VARCHAR
IS
    v_dienos    NUMBER;
BEGIN
    SELECT vidurkis
        INTO v_dienos
    FROM
        (select
            --korteles_id,
            AVG(skirtumas) as vidurkis
        from
            (select
                a.korteles_id,
                a.data_laikas,
                CAST (min(b.data_laikas - a.data_laikas ) as NUMBER ) as skirtumas
            from
                (select
                    distinct data_laikas, korteles_id
                from
                    korteliu_tranzakcijos
                where
                    korteles_id=p_kortelesid
                ) a
            inner join
                (select
                    distinct data_laikas, korteles_id
                from
                    korteliu_tranzakcijos
                ) b
                on a.korteles_id = b.korteles_id and a.data_laikas < b.data_laikas
            GROUP BY a.korteles_id, a.data_laikas
        )
        --GROUP BY korteles_id
    );

    IF v_dienos<=7 then return 'iki 1 sav.'; END IF;
    if v_dienos>7 and v_dienos<=31 then return 'iki 1 men.'; end if;
    if v_dienos>31 and v_dienos<=61 then return '1-2 men'; end if;
    if v_dienos>62 and v_dienos<=122 then return '3-4 men'; end if;
    if v_dienos>122 and v_dienos<=180 then return 'iki 6 men'; end if;
    if v_dienos>180 and v_dienos<=365 then return '6-12 men'; end if;
    if v_dienos>365 then return 'virs 1 metu'; end if;
    return 'vienkartinis';

    --return v_dienos;
END;

```