

KAUNO TECHNOLOGIJOS UNIVERSITETAS

INFORMATIKOS FAKULTETAS

INFORMACIJOS SISTEMŲ KATEDRA

Žygimantas Vasiliauskas

**ĮMONĖS DUOMENŲ STATISTINĖS ANALIZĖS,
PANAUDOJANT DBVS, GALIMYBIŲ TYRIMAS**

Magistro darbas

Darbo vadovas

prof. dr. R. Butleris

KAUNAS, 2010

KAUNO TECHNOLOGIJOS UNIVERSITETAS

INFORMATIKOS FAKULTETAS

INFORMACIJOS SISTEMŲ KATEDRA

**ĮMONĖS DUOMENŲ STATISTINĖS ANALIZĖS,
PANAUDOJANT DBVS, GALIMYBIŲ TYRIMAS**

Magistro darbas

Vadovas

prof. dr. R. Butleris

2010-06

Konsultantas

lekt. T. Danikauskas

Recenzentas

dr. S. Drąsutis

2010-06

Atliko

IFM 4/4 gr. stud

Žygimantas Vasiliauskas

2010-06-03

KAUNAS, 2010

Turinys

Įvadas	6
1. Analitinių DBVS priemonių analizė	7
1.1. Analizės tikslas	7
1.2. Tyrimo sritis, objektas ir problema	7
1.3. Tyrimo objekto analizė.....	7
1.4. Vartotojų analizė.....	10
1.4.1. Vartotojų aibė, tipai ir savybės.....	10
1.4.2. Vartotojų tikslai ir problemos.....	11
1.5. Esamų sprendimų analizė.....	12
1.6. Siekiamas sprendimas	15
1.7. Analizės išvados	24
2. Sistemos reikalavimai.....	25
2.1. Reikalavimų specifikacija	25
2.2. Dalykinės srities modelis.....	35
2.3. Reikalavimų analizės apibendrinimas.....	35
3. Sistemos projektas.....	36
3.1. Sistemos architektūros projektas	36
3.1.1. Sistemos loginė architektūra	36
3.1.2. Vartotojo paslaugos	37
3.1.3. Veiklos paslaugos	37
3.2. Detalus projektas.....	38
3.3. Sistemos elgsenos modelis	40
3.4. Duomenų bazės schema	42
3.5. Realizacijos modelis	43
4. Statistinės analizės, panaudojant DBVS funkcijas, realizacija.....	45
4.1. Realizacijos veikimo aprašymas.....	45
4.2. Testavimo modelis ir duomenys, kontrolinis pavyzdys	51
5. Eksperimentinis tyrimas	54
5.1. Eksperimento uždavinių sprendimas	54
5.2. Analizės atlikimo palyginimas naudojant DBVS priemones su Oracle Discoverer ir Microsoft Excel 2007.....	60
5.3. Taikymo rekomendacijos	65

6. Darbo rezultatų apibendrinimas.....	66
Išvados.....	68
Santrumpų ir terminų žodinėlis	69
Literatūra	70
1. PRIEDAS. KONFERENCIJOJE PRISTATYTAS IR IŠSPAUSDINTAS STRAIPSNIS.....	72

Research of Statistical Analysis of Enterprise Data using DBMS

Summary

This work reviews advantages and disadvantages of existing statistical analysis tools usage for small and mid sized enterprises, explores usage of statistical analysis of enterprise data in data warehouse, reviews integrated statistical analysis functions in existing database management systems and integrated graphical software for graphical statistical analysis usage. In this work were analyzed Oracle, Microsoft SQL Server and DB2 DBMS. The new statistical analysis solution was offered. This solution allows statistical analysis of existing data using integrated statistical functions of the database management systems and integrated graphical tool. Proposed solution has been designed and realized for the statistical analysis of insurance company's data. Oracle DBMS was selected for statistical analysis, because this DBMS is used by insurance company. Oracle has a large number of integrated statistical analysis functions; it ensures more diverse and rapid analysis. Graphic depiction of the selected Oracle Discoverer tool to optimally exploit data analysis potential of Oracle DBMS. The proposed statistical analysis process is versatile and suitable for different business areas and can be applied to other DBMS, which has an integrated analytical functions and graphical tools for the results display.

Ivadas

Veiklos analizė - tai verslo valdymo terminas, kurio esmė - reikalingų duomenų surinkimas iš pradinių šaltinių, norimos informacijos parengimas (pagal surinktus duomenis) ir jos teikimas vartotojams, kurie priima svarbius sprendimus, verslo efektyvumui didinti.

Veiklos analizavimas yra labai svarbus dabartinėje įmonių sferoje. Šiandien daugelis stambių įmonių turi išvystytą informacinę infrastruktūrą. Visa įmonės disponuojama informacija tuo pačiu metu yra aktyvas, kuris turi būti visiškai panaudotas verslui plėtoti ir planuoti, klientams segmentuoti ir rizikai valdyti, sąnaudoms, veiklos procesams bei resursams optimizuoti. Tačiau daugelis įmonių nesuvokia kokią valdymo informaciją galima gauti iš įmonės disponuojamų duomenų. Įmonės kaupia ir naudoja informaciją verslo plėtrai ir naujų sprendimų priėmimui, tačiau daugelis mato, jog kaupiama informacija tampa tik duomenimis, kuriuos labai sunku panaudoti dėl jų vis didėjančio kiekio, o neturėdamos žinių ir įgūdžių įmonės negali išspręsti šių problemų. Didelius kiekius duomenų labai sunku analizuoti, kad galima būtų priimti svarbius verslui sprendimus.

Prieš keletą dešimtmečių buvo sukurti pirmieji veiklos analizės įrankiai, kurie padėjo analizuoti svarbius duomenis duomenų saugyklose, taip prisidėdami prie verslo plėtos. Veiklos analizės metodikos ir įrankiai yra svarbios toms įmonėms, kurios nori, kad jų darbuotojai, partneriai ir klientai priimdami sprendimus vadovautųsi ne tik savo intuicija ir patyrimu, bet ir aktualia bei išsamia informacija apie viską, kas yra susiję su priimamais sprendimais. Šiandieniniame žinių ekonomikos amžiuje, labai greitai augant naudojamos informacijos kiekiui, įmonėms integruojant veiklą ir teikiant paslaugas vis didesniam klientų skaičiui, didėjant reikalavimams paslaugų kokybei ir sprendimų priėmimo greičiui, veiklos analizės technologijos yra jau būtinybė, norint išlikti.

Veiklos analizavimo įrankiai tapo vienu iš svarbiausių priemonių kokybiškai ir efektyviai atlikti šią dieną verslo analizę, tačiau su kiekvienais metais šie įrankiai tampa vis sunkiau panaudojami ir brangesni. Įrankiai sudėtingėja, jų teikiamų funkcijų gausa didėja, dalis jų tampa labai integralūs su didelėmis sistemomis taip palikdami mažas ir vidutines įmones be pasirinkimo.

Įmonių veiklos duomenys dažniausiai yra saugomi pasirinktose duomenų bazių valdymo sistemose, kuriose yra sukurta funkcijų, kurių pagalba galima analizuoti duomenis, taip taupant resursus ir lėšas. Todėl alternatyva veiklos analizės įrankiams – naudoti DBVS integruotas funkcijas, kurias vėliau panaudojus užklausoje su integruotais grafiniais įrankiais, įmonės gali gauti analogišką rezultatą kaip naudojant esamus veiklos analizės įrankius.

1. Analitinių DBVS priemonių analizė

1.1. Analizės tikslas

Parinkti analizės metodus ir priemones, išanalizuoti esamus problemos sprendimus, išanalizuoti dalykinę sritį ir joje esančią problemą, nustatyti tyrimo objektą ir sritį, suformuluoti tyrimo tikslą ir uždavinius.

1.2. Tyrimo sritis, objektas ir problema

Tyrimo sritis – DBVS (duomenų bazių valdymo sistemos) priemonių panaudojimas analizuojant duomenis duomenų saugyklose.

Tyrimo objektas - DBVS priemonės, padėsiančios išgauti naudingos informacijos. Kompanijos norėdamos analizuoti įmonės duomenis, tam kad darytų sprendimus, kurių pagalba būtų įmanoma optimizuoti veiklą, naudoja brangius įrankius SAS ar SPSS. SAS yra produktas, kuris savo statistinėmis priemonėmis padeda analizuoti įmonės veiklą ir priimti svarbius sprendimus. Tačiau, tokie įrankiai kaip SAS ir SPSS kainuoja labai brangiai ir yra sudėtingi. Tai yra didžiausia problema, su kuria susiduria visos įmonės įsigijusios šiuos analitinius įrankius, taip prarasdamos daug pinigų, be to investuodamos didelius resursus į žmonių apmokymus. Šiai problemai spręsti galima pasirinkti DBVS analitines priemones ir „select“ užklausas, kurių pagalba galima atlikti svarbią veiklos analizę ir priimti reikalingus sprendimus, tam neišleidus didelių resursų. Vėliau „select“ užklausos gali būti patalpintos į grafinio vaizdavimo įrankius.

1.3. Tyrimo objekto analizė

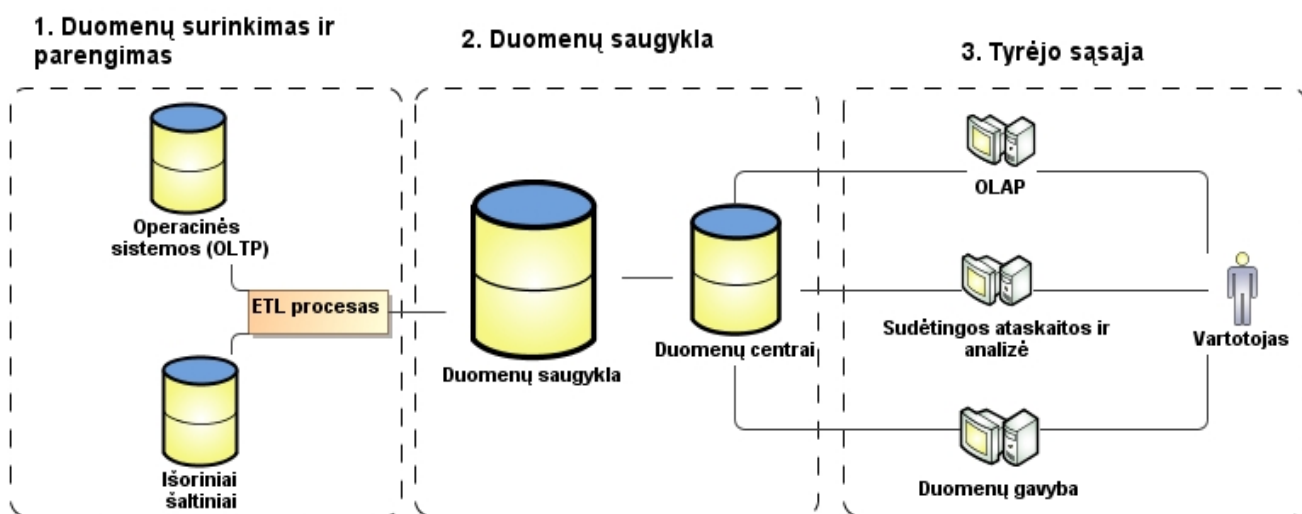
Įmonės veiklos analizės procesas, pavaizduotas paveiksle „Paveikslas 1. Veiklos analizės procesas“(psl. 8).



Paveikslas 1. Veiklos analizės procesas

Įmonės vykdydamos veiklą, nuolat kaupia vykdomos veiklos duomenis, juos panaudoja svarbiai informacijai ir žinioms gauti, tam kad vėliau įmonės darbuotojai galėtų priimti svarbius sprendimus veiklos efektyvumui gerinti.

Duomenys dažniausiai saugomi reliacinėse duomenų bazėse. Jose yra saugomi svarbūs visos įmonės duomenys, kurie yra saugomi vienoje vietoje arba saugomi kiekviename departamente atskirti vieni nuo kitų ir galintys įtakoti įmonės veiklą. Kartais duomenys saugomi kituose šaltiniuose - ne reliacinėje duomenų bazėje. Kai duomenys nėra vienoje vietoje, tuomet atlikti tikslią veiklos analizę tampa sudėtingiau. Todėl priimant labai svarbius sprendimus, prarandamas labai svarbus laikas. Dėl šios priežasties yra naudojamos duomenų saugyklos, pavaizduotos duomenų saugyklos koncepcinėje diagramoje „Paveikslas 2. Duomenų saugyklos koncepcinė diagrama“.



Paveikslas 2. Duomenų saugyklos koncepcinė diagrama

Trys pagrindinės veiklos analizės technologijų dalys [1]:

1. Duomenų surinkimas ir parengimas

Reikalingų informacijai parengti duomenų surinkimas iš pradinių šaltinių, jos apdorojimas ir sutvarkymas.

Informacijos naudotojams skirtos aktualios informacijos parengimas apjungiant (integruojant) surinktus duomenis, reikiama juos transformuojant ir grupuojant, papildant skaičiavimų būdu gautais išvestiniais rodikliais bei pateikiant informacijos ir jos parengimo algoritmų paaiškinimus. Ši veiklos analizės dalis dažnai vadinama ETL procedūra (Extraction, Transformation, Loading). ETL procedūra paprastai yra atliekama atskiroje tarnybinėje stotyje.

Paruoštos informacijos sukėlimas į analitinėms duomenų užklausoms vykdyti skirtą duomenų bazę, kuri yra vadinama duomenų saugykla.

2. Duomenų saugykla

Svarbi veiklos analizės technologijų dalis yra duomenų saugykla, kurioje saugomi sutvarkyti iš pradinių šaltinių paimti duomenys. Duomenų saugykloje yra saugoma aktualaus laikotarpio informacija.

Informacija duomenų saugyklose parengta taip, kad ją būtų galima greitai ir patogiai nuskaityti.

Duomenų saugyklos duomenys reguliariai atnaujinami ETL procedūrų pagalba. Galutinis vartotojas arba tyrėjas duomenų saugyklos duomenų negali keisti ar taisyti, juos gali tik analizuoti.

3. Tyrėjo sąsaja

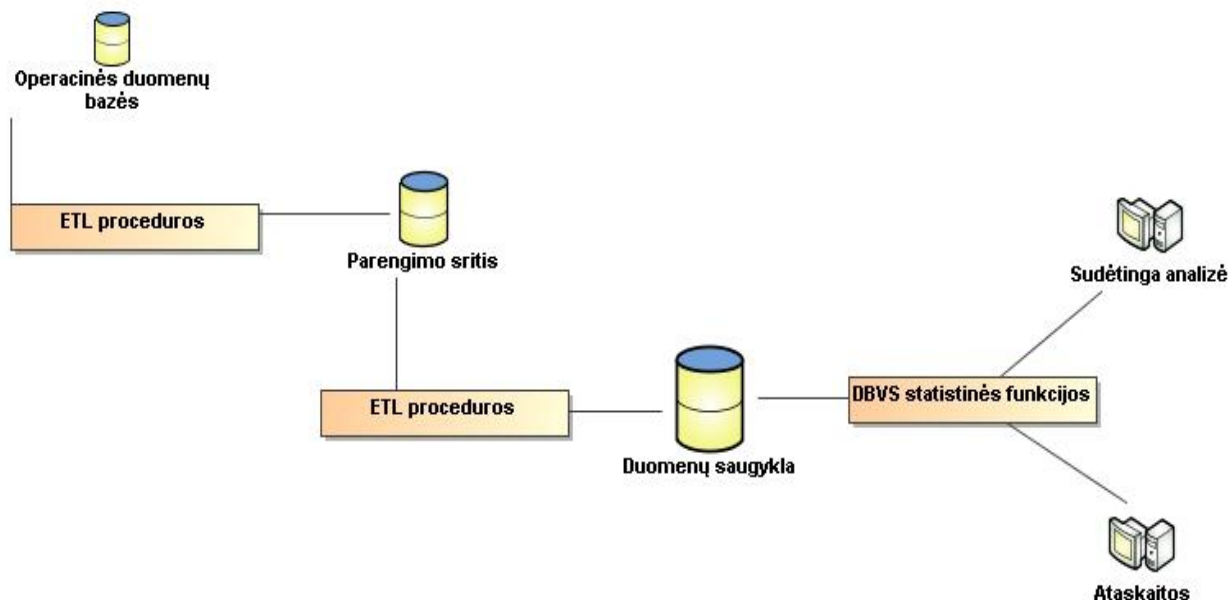
Tai kompiuterinė ir programinė įranga, kuria naudodamiesi tyrėjai iš savo darbo vietų prisijungia prie duomenų saugyklos, nuskaityti iš jos duomenis ir informaciją, juos analizuoja, formatuoja, rengia savo ataskaitas ir dokumentus, atlieka stebėjimo, planavimo, prognozavimo bei sprendimų priėmimo veiksmus.

Daugelis veiklos analizės priemonių yra skirti atlikti statistinę analizę, tyrėjams, kurie priima labai svarbius įmonei sprendimus. Tačiau analitiniai įrankiai yra brangūs ir sudėtingi, kaip paminėjo Gareth Herschel, Gartner group atstovas, savo metinėje ataskaitoje [2].

Visų analitinių įrankių bendras bruožas – jos naudoja *select* užklausas ir jau sudarytus algoritmus tam, kad atliktų analizę.

Daugelyje DBVS gali būti panaudotos analitinės funkcijos. Šios analitinės funkcijos naudojamos dažniausiai duomenų saugyklose, kuriose yra saugoma terabaitai įmonės duomenų ir duomenys yra tikslūs ir paruošti analizei.

Oracle, Microsoft SQL Server ar DB2 turi integruotas analitines funkcijas, kurias galima panaudoti duomenų analizei. Vėliau šios funkcijos gali būti panaudotos įvairiuose įrankiuose, pateikus *select* užklausą su šiomis analitinėmis funkcijomis, kaip pavaizduota paveiksle „Paveikslas 3. Duomenų saugyklos ir analitinių funkcijų koncepcinė panaudojimo schema“.



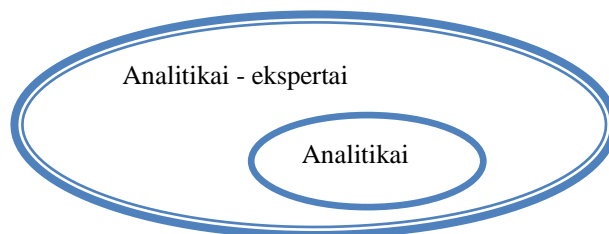
Paveikslas 3. Duomenų saugyklos ir analitinių funkcijų koncepcinė panaudojimo schema

Pateiktoje schemoje, duomenys į duomenų saugyklą patenka iš įmonės operacinių duomenų bazių, kur prieš patekdami į duomenų saugyklą pirmiausiai sutvarkomi parengimo srityje, kurioje nustatomas duomenų validumas. Duomenų saugykloje integruotos DBVS analitinės funkcijos įgalina duomenis analizuoti pačioje duomenų bazėje, o užklausa duomenų analizei ar ataskaitų kūrimui gali būti vykdoma pasitelkus grafinius vaizdavimo įrankius.

1.4. Vartotojų analizė

1.4.1. Vartotojų aibė, tipai ir savybės

Veiklą analizuoja dažniausiai įmonių analitikai - dalykinės srities specialistai. Analitikais taip pat laikomi dalykinės srities ekspertai, puikiai žinantys ir suprantantys statistiką ir puikūs informacinių technologijų žinovai. Tokius vartotojus mes tiesiog įvardinsime kaip tyrėjus.



Paveikslas 4. Vartotojų aibės

Kaip pavaizduota paveiksle „Paveikslas 4. Vartotojų aibės“, matome ryšį tarp analitikų – ekspertų ir analitikų. Analitikai priklauso analitikų – ekspertų aibei, kurie yra pagrindinei veiklos analizės specialistai, priimančys efektyvius sprendimus. Analitikai yra savo dalykinės srities specialistai, tačiau jie neturi reikiamų įgūdžių statistikoje ir informacinėse technologijose, todėl norint jiems tapti aibės analitikai – ekspertai dalimi, reikia turėti bent šių sričių pagrindus.

Tyrėjai analizuodami duomenis atlieka labai svarbius sprendimus, lemiančius įmonės sėkmingą veiklą. Pagal turimus sukauptus duomenis, tyrėjai ją bando analizuoti ir versti į informaciją, o informaciją į žinias tam, kad būtų priimtos tinkamos išvados ir būtų palyginta su pasiūlyta hipoteze apie vykdomą veiklą.

Vienu metu analizuoti duomenis gali ne vienas tyrėjas. Taip galima atlikti įvairių analizę laiku, analizuojant pardavimus, prekybą, žmoniškųjų išteklių panaudojimą ir kitą pasirinktą dalykinę sritį.

1.4.2. Vartotojų tikslai ir problemos

Pagrindinis vartotojų tikslas – naudojant analitinius analizės įrankius iš sukauptų duomenų gauti dalykinės srities informaciją, o informacijai tampant žiniomis, priimti svarbius sprendimus, padėsiančius įmonei ateityje pagerinti veiklos naudą. Vartotojų problemos, išskylančios naudojantis esamus sprendimus, aprašytos lentelėje „Lentelė 1. Vartotojų esamos problemos dirbant su jau esamais sprendimais“.

Lentelė 1. Vartotojų esamos problemos dirbant su jau esamais sprendimais

Problema	Kaip viskas vyksta dabar
Analizės atlikimas laiko atžvilgiu yra lėtas	Vartotojai analizuodami duomenis analitinėmis priemonėmis sugaišta daug laiko nurodant rodiklius bei kitus svarbius parametrus, tam kad gautų reikiamą rezultatą.
Analizės atlikimas yra sudėtingas procesas	Vartotojai analizuodami dalykinę sritį turi puikiai žinoti visus statistinius rodiklius, jų parametrus ir norimo rezultato vaizdą, todėl parinktas netikslus parametras gali sukelti visiškai priešingą analizės rezultatą nei buvo tikėtasi. Analitinės priemonės labai sudėtingos, todėl nėra tiek daug analitikų ar statistikų, kurie efektyviai galėtų atlikti veiklos analizę[3].

Šioms problemoms spręsti galima pasitelkti integruotas DBVS funkcijas, kurios bus naudojamos užklausomis *select* sakiniuose įvairiuose paketuose.

1.5. Esamų sprendimų analizė

SAS(Statistical Analysis System)

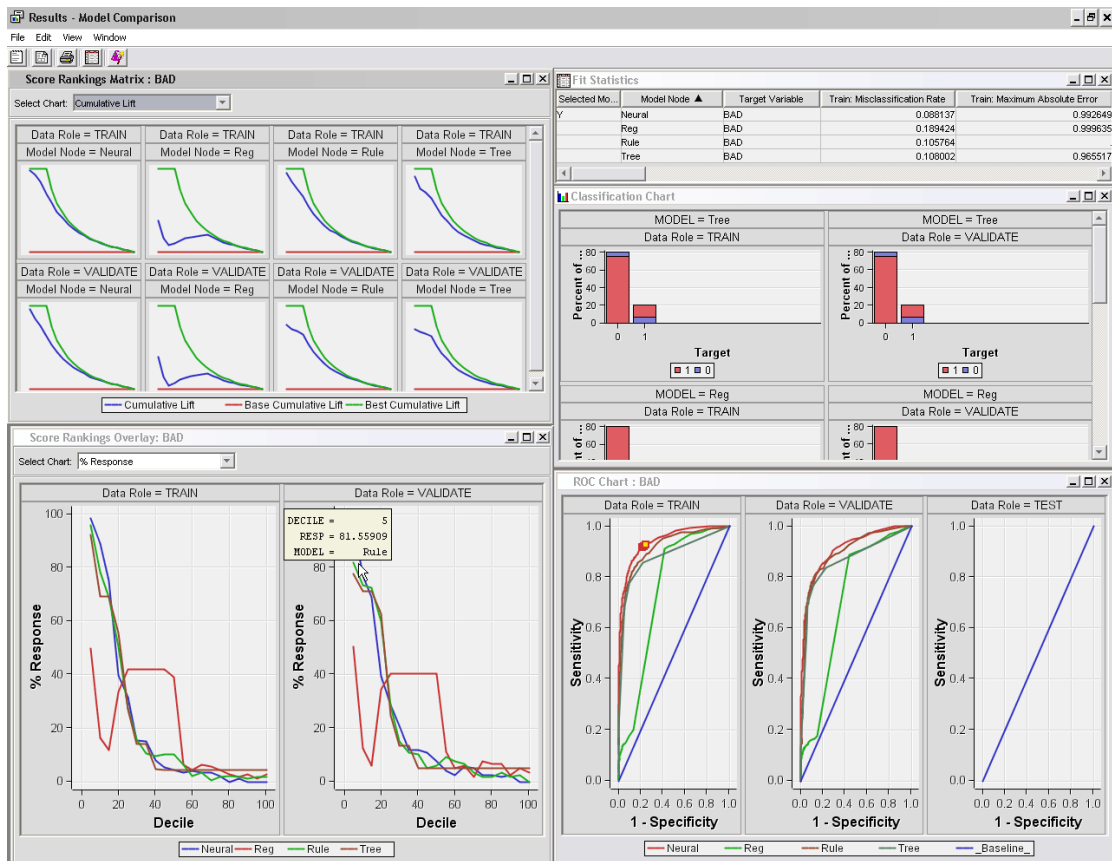
Sprendimas priimtas todėl, kad yra viena iš populiariausių priemonių, spręsti statistinių uždavinių analizę [2].

1976 m. įkurta kompanija, kuri rūpinasi SAS produkto kūrimu bei tobulinimu. Per šį laiką ši priemonė tapo viena žinomiausių visame pasaulyje analitinių įrankių rinkoje. Pagrindinis informacinis šaltinis apie šį sprendimą – <http://www.sas.com>.

Šios priemonės pagrindinis tikslas – leisti vartotojui analizuoti įmonių veiklą ir priimti reikalingus sprendimus, naudojant labai sudėtingus statistinius algoritmus. Ši idėja nuo sukūrimo pradžios iki šiandien yra šios priemonės stiprioji pusė. Šis sprendimas naudojamas tada, kai įmonėms reikia atlikti veiklos analizę pagal tam tikrus statistinius rodiklius arba atlikti sudėtingą duomenų gavybą numatant tolimesnę įmonės veiklą.

Šio sprendimo vartotojai, dažniausiai didžiosios pasaulio įmonės, kurios gali išleisti milijonus tam, kad galėtų įsigyti šio sprendimo licenziją. SAS sprendimai yra panaudoti 45000 žiniatinkliuose, 109 šalyse, įskaitant 91 įmonę, kuri priklauso FORTUNE 100 (didžiausių pasaulio įmonių šimtukas, renkamas kasmet, pagal atnešamą įmonių pelną).

SAS įrankis orientuotas į veiklos analizės sprendimus ir kasmet produktas funkcionalumu vis didėja, tuo pačiu kyla ir produkto kaina. Šiuo metu produktas gali pasiūlyti ne tik minėtas veiklos analizės priemones, bet ir duomenų integracijos sprendimus bei statistinės analizės sprendimus. SAS vartotojo sąsaja pavaizduota paveiksle „Paveikslas 5. SAS vartotojo sąsaja“(psl. 13).



Paveikslas 5. SAS vartotojo sąsaja

SAS priemonė veikia Windows, UNIX operacinėse sistemose, tad problemų vartotojams dėl operacinės sistemos pasirinkimo neiškyla, nes 95% vartotojų naudojami būtent Windows arba UNIX operacinėmis sistemomis.

SAS naudoja unikalią kalbą aprašyti procedūroms bei funkcijoms, todėl išmokti kaip naudotis reikia keletą mėnesių, o efektyviai naudotis – metų. SAS taip pat turi daug jau aprašytų algoritmų, kuriuos galima naudoti atliekant statistinę analizę. SAS taip pat galima naudoti SQL procedūras.

SAS yra geriausias duomenų gavybos įrankis rinkoje, atliekantis vartotojo pačias sudėtingiausias ir kritiškiausias duomenų gavybos užduotis [4]. SAS priemonės privalumai ir trūkumai pateikti lentelėje „Lentelė 2. SAS sprendimo privalumai bei trūkumai“.

Lentelė 2. SAS sprendimo privalumai bei trūkumai

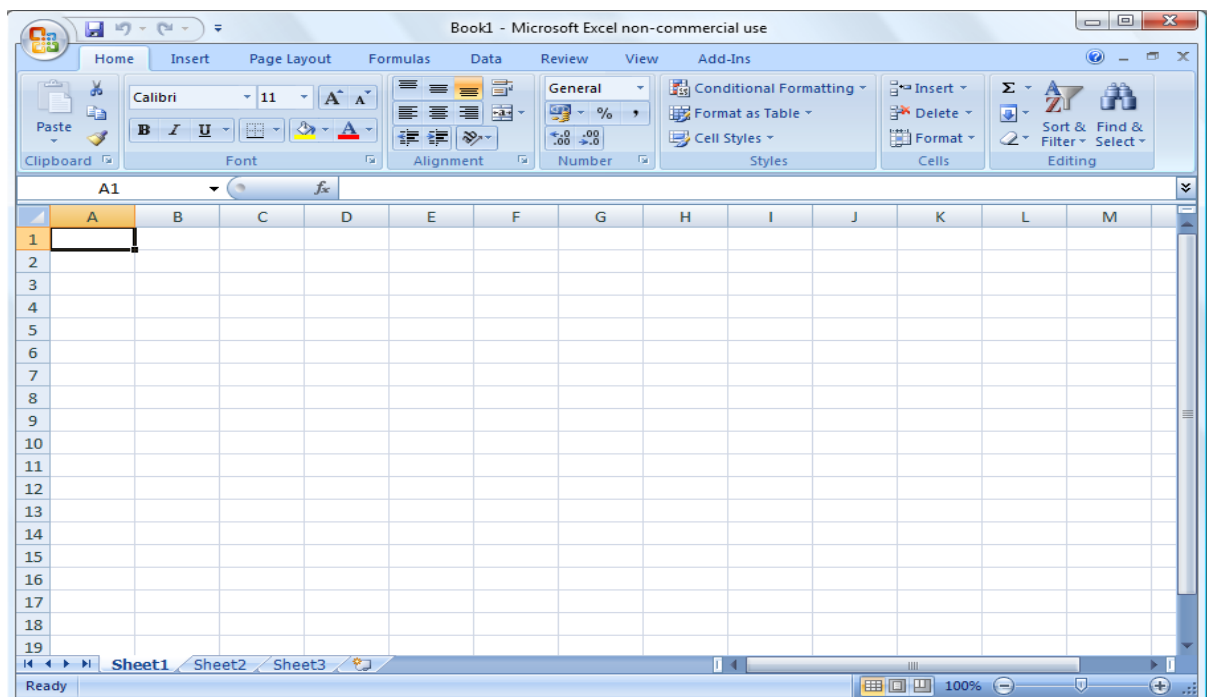
Privalumai	Trūkumai
Gali atlikti sudėtingas duomenų gavybos užduotis.	Didelė kaina.
SAS įrankių gausa leidžia efektyviai panaudoti sprendžiat įvairius statistinius uždavinius.	Labai sudėtingas panaudojimas.

Vienu metu analizuoti duomenys gali daug vartotojų.	Norint efektyviai panaudoti įrankį, reikia gerai mokėti programuoti SAS unikalio kalba.
Veikia tiek Windows tiek UNIX operacinėse sistemose.	Reikia keletą mėnesių norint efektyviai išmokti naudotis priemone.

Microsoft Excel

Kadangi sprendimas plačiai taikomas statistinei analizei atlikti, būtina jį nuodugniai išanalizuoti. Microsoft kompanijos produktas, dar 1993 m. tapęs Microsoft Office paketo dalimi. Vienas iš populiariausių produktų, naudojamų Windows OS terpėje, matematiniais bei statistiniams skaičiavimams atlikti. Beveik kiekvienas vartotojas naudojasi šiuo įrankiu kasdienėje veikloje sprendžiant matematinius uždavinius [5]. Daugelyje įmonių naudojamas įrankis, kuris gali pagelbėti daugumai vartotojų priimant sprendimus, atliekant duomenų analizę. Šis įrankis yra skirtas atlikti nedidelius statistinius analizės uždavinius, todėl dirbant su dideliu kiekiu duomenų, šis įrankis nėra tinkamas naudoti.

Microsoft Excel turi labai daug papildomų programos praplėtimų, įgalinančių atlikti žymiai sudėtingesnę analizę. Apie šiuos papildomus įskiepius galima pasiskaityti kūrėjų puslapyje <http://www.lumenaut.com/>. Jie skirti praplėsti, statistinei duomenų analizei skirtų, algoritmų rinkinį, tačiau šie papildomi įskiepiai kainuoja brangiai ir jie neištaiso svarbių, efektyvumą užtikrinančių, veiksnių. Microsoft Excel vartotojo sąsaja pateikta paveiksle „Paveikslas 6. Microsoft Excel vartotojo sąsaja“.



Paveikslas 6. Microsoft Excel vartotojo sąsaja

Įrankis puikiai tinka atlikti kasdienes duomenų analizės uždavinius mažuose įmonės padaliniuose, tačiau rezultatas gali būti ir nelabai tikslus.

Vienas didžiausių Microsoft Excel trūkumų – saugumas. Įmonės personalas, analizuodamas svarbius įmonei duomenis, gali išsaugoti Excel bylą ir pernešti ją į kita vietą, todėl Microsoft Excel atsisakoma kaip vienos iš nesaugiausių verslo analitikos priemonių.

Įrankis yra nesudėtingas naudoti, tačiau norint atlikti analizę pagal sudėtingesnius algoritmus, reikia žinoti Microsoft Visual Basic kalbą [6]. Microsoft Excel nebuvo kuriamas tam, kad būtų geras statistinės analizės įrankis [7]. Šis įrankis skirtas atlikti nesudėtingus įmonės veiklos apskaičiavimus.

Privalumai bei trūkumai Microsoft Excel aprašyti „Lentelė 3. Microsoft Excel įrankio privalumai ir trūkumai“.

Lentelė 3. Microsoft Excel įrankio privalumai ir trūkumai

Privalumai	Trūkumai
Atlieka nesudėtingų statistinių uždavinių analizę.	Lėtai atlieka statistinę analizę.
Lengvai įsisavinamos žinios.	Duomenis gali analizuoti tuo pačiu metu tik vienas vartotojas.
Daug sukurtų funkcijų, kurių kasdienėje įmonės veikloje pakanka nesudėtingam uždaviniui spręsti.	Papildomų funkcijų kūrimas reikalauja gilių Microsoft Visual Basic žinių.
Paprastas naudoti ir nėra brangus.	Nesaugus sprendimas atliekant analizę su svarbiais įmonei duomenimis.
	Įrankis skirtas tik Windows operacinei sistemai.

1.6. Siekiamas sprendimas

Atlikus esamų sprendimų analizę, lentelėje „Lentelė 4. Lyginimo sprendimų kriterijai“ analizuojami konkretūs sprendimų palyginimo kriterijai.

Lentelė 4. Lyginimo sprendimų kriterijai

Lyginimo kriterijai	SAS	Microsoft Excel
Kaina	Didelė	Maža
Sudėtingumas ir panaudojimas	Įrankis sudėtingas atlikti net nesudėtingą statistinę analizę. Norint efektyviai dirbti, reikia mokėti SAS kalbą. Sukurtos papildomos funkcijos veiks tik SAS įrankio aplinkoje.	Įrankiu lengva naudotis. Tačiau norint sukurti sudėtingesnes funkcijas, reikia mokėti naudoti Microsoft Visual Basic kalbą. Papildomos sukurtos funkcijos bus panaudojamos Microsoft Excel aplinkoje.
Funkcijų pasirinkimas uždaviniams spręsti	Funkcijų kiekis didelis statistiniams uždaviniams spręsti.	Statistinės funkcijos yra standartinės.

Darbas su dideliu kiekiu duomenų	Gali analizuoti terabaitinius duomenų masyvus.	Gali efektyviai dirbti su nedaugiau nei 1000000 eilučių duomenų.
Resursų sunaudojimas	Analizei paimti duomenys talpinami SAS laikinojoje srityje, taip užimdami vietą fiziniėje atmintyje.	Paimti duomenys užima daug vietos fiziniėje atmintyje, o vykdant ir operatyvioje atmintyje.
Duomenų prieinamumas tuo pačiu metu keliems vartotojams	Tuo pačiu metu gali dirbti iki 20 vartotojų. Norint, kad analizę atliktų daugiau nei 20 žmonių, reikia papildomai mokėti.	Analizuoti duomenis gali tik vienas analitikas. Tuo pačiu metu norintis analizuoti duomenis kitas analitikas redaguoti sukurtos bylos negali.
Saugumas	Naudojimasis įrankiu yra saugus, kadangi analizuojami duomenys yra šifruojami.	Duomenys gali būti išsaugomi byloje ir pernešami arba persiunčiami.
Duomenų analizavimas naudojant skirtingas duomenų saugyklos	Gali paimti duomenis tiek iš Oracle, DB2 ar Microsoft SQL Server duomenų saugyklos.	Duomenis analizei gali paimti iš Microsoft SQL Server, Oracle duomenų saugyklos.

Palyginus abu sprendimus, galima daryti išvada, kad nors ir Microsoft Excel sprendimas yra ganėtinai suprantamas vartotojui ir nepasižymi didele kaina, tačiau sudėtingų statistinių uždavinių sprendimui nėra tinkamas. Analizuojant svarbius įmonės duomenis iškyla pavojus informacijos konfidencialumui, kadangi Microsoft Excel byla gali būti išsaugota ir pernešta kitiems vartotojams. Analizės duomenų kiekis turi būti nedidelis norint efektyviai išnaudoti įrankį, todėl net mažų įmonių sudėtingesnei statistinei analizei įrankio pajėgumų gali nepakakti.

SAS įrankis yra brangus ir sudėtingas, tačiau gali pasiūlyti labai tikslią ir efektyvią statistinę analizę. SAS įrankiu efektyviai naudotis gali tik labai patyręs statistikos specialistas, todėl daugeliui įmonių šis sprendimas daug naudos neatneša, kadangi rezultatų reikia būtent dabar ir čia. Investuoti į SAS įrankį gali tik didelės ir turtingos įmonės, kurios, naudodamos šį įrankį, rezultatų gali laukti keletą mėnesių.

Siūlomas sprendimas, kuris gali papildyti aptartus sprendimus, yra DBVS funkcijos, kurios gali būti vykdomos duomenų saugykloje ir grafiškai atvaizduojamos įvairiuose integruotuose įrankiuose.

DBVS analitines funkcijas naudinga naudoti, nes [8]:

- Saugumas;
- SQL analitinių funkcijų naudojimas duomenų saugyklos viduje, todėl galima sukurti tokias užklausas, kokios reikalingos veiklos analizei pagal vartotojo poreikius [9];
- Realus laiko skaičiavimai;

- Sudėtingų užklausų sudarymas ir panaudojimas kuriant dar sudėtingesnes;
- DBVS sukurtų funkcijų optimizavimas ir modernizavimas;
- DBVS funkcijų panaudojimas grafiniuose analizės įrankiuose.

DBVS analitinių funkcijų integracija versle:

- Užklausas integruoti į Oracle Discoverer, Oracle XML Publisher, Microsoft SQL Server Analysis Services, IBM Cognos;
- Užklausas integruoti įrankiams, kurie gali bendrauti su DBVS sistemomis – Business Objects, QlickView.

Duomenų saugyklose galima sukurti vartotojo aprašytas funkcijas kartu naudojant ir SQL užklausas, kurių pagalba analizavimas pačioje duomenų saugykloje tampa efektyvesnis nei naudojant vien tik SQL kalbą [10].

Didžiausios duomenų saugyklų sistemos yra Microsoft SQL Server, Oracle ir DB2. Kiekviena šių sistemų turi skirtingas integruotas analitines funkcijų aibes, todėl reikia teisingai pasirinkti DBVS, kurioje efektyviausiai būtų pritaikomos analitinės funkcijos.

Microsoft SQL Server

Microsoft SQL Server funkcijų gausa nėra didelė. Tai dažniausiai standartinės analitinės funkcijos, tačiau keletą funkcijų yra efektyvesnių, bandant atlikti sudėtingesnę duomenų analizę. Microsoft SQL Server analitinių funkcijų aprašymas pateiktas lentelėje „Lentelė 5. Microsoft SQL Server analitinių funkcijų aprašymas“.

Lentelė 5. Microsoft SQL Server analitinių funkcijų aprašymas

Funkcijos aprašymas	Funkcijos
Rangas tam tikrai grupei	RANK
Tvarkingo duomenų rinkinio padalinimas į kubus.	NTILE
Imties standartinis nuokrypis	STDEV
Kvadratinė šaknis iš duomenų dispersijos	STDEVP
Duomenų dispersijos rinkinys	VARP

Microsoft SQL Server analitinių funkcijų rinkinys nėra didelis, tačiau atlikti svarbią pradinę analizę galima ir su šiomis funkcijomis. Ateityje Microsoft kompanija žada praplėsti DBVS analitinių funkcijų paketą Microsoft SQL Server produktui, todėl ateityje atlikti veiklos analizę bus įmanoma ir naudojant Microsoft SQL Server DBVS.

Oracle

Oracle DBVS analitines funkcijas pirmasis aprašė Thomas Kyte [8] dar 2000 metais Oracle 8 versijai. Dar tada buvo skeptiškai žiūrima į šias funkcijas, tačiau vis kylant analitinių įrankių kainai bei sudėtingėjant analitiniams įrankiams naudinga pradėti naudoti šias DBVS integruotas funkcijas, pradiniai duomenų analizei atlikti[11].

Oracle DBVS 11g turi sudėtingų analitinių funkcijų, kurios gali padėti atlikti analizę nenaudojant didelių resursų. Funkcijos, kurios gali būti naudojamos: kompleksinė duomenų transformacija, duomenų gavyba, vaizdinių požymių išskyrimas, linijinė algebra, nuspėjamoji duomenų analitika, erdvinė analitika, analitinės funkcijos, teksto gavyba, Pareto analizė, Chi kvadrato statistika. Oracle analitinių funkcijų aprašymas pateiktas lentelėje „Lentelė 6. Oracle analitinių funkcijų aprašymas“.

Lentelė 6. Oracle analitinių funkcijų aprašymas

Funkcijos aprašymas	Funkcijos
Rangas tam tikrai grupei	RANK
Tvarkingo duomenų rinkinio padalinimas į kubus.	NTILE
Imties standartinis nuokrypis	STDDEV
Kvadratinė šaknis iš duomenų dispersijos	STDDEV_POP
Duomenų dispersijos rinkinys	VAR_POP
Koreliacijos koeficientas	CORR
Imties kovariacija nustatyta porai	COVAR_COUNT
Kaupiamasis paskirstymas grupėje	CUME_DIST
Suteikia priėjimą prie eilutės rinkinio	LAG
Suteikia priėjimą prie eilutės rinkinio	LEAD
Grįžtamoji paskirstymo funkcija	PERCENTILE_CONT
Linijinės regresijos funkcija	REGR_AVGX
Linijinės regresijos funkcija	REGR_AVGY
Linijinės regresijos funkcija	REGR_COUNT
Linijinės regresijos funkcija	REGR_INTERCEPT
Linijinės regresijos funkcija	REGR_R2
Linijinės regresijos funkcija	REGR_SLOPE
Linijinės regresijos funkcija	REGR_SXX
Linijinės regresijos funkcija	REGR_SXY
Linijinės regresijos funkcija	REGR_SYY
Duomenų gavybos funkcijos	

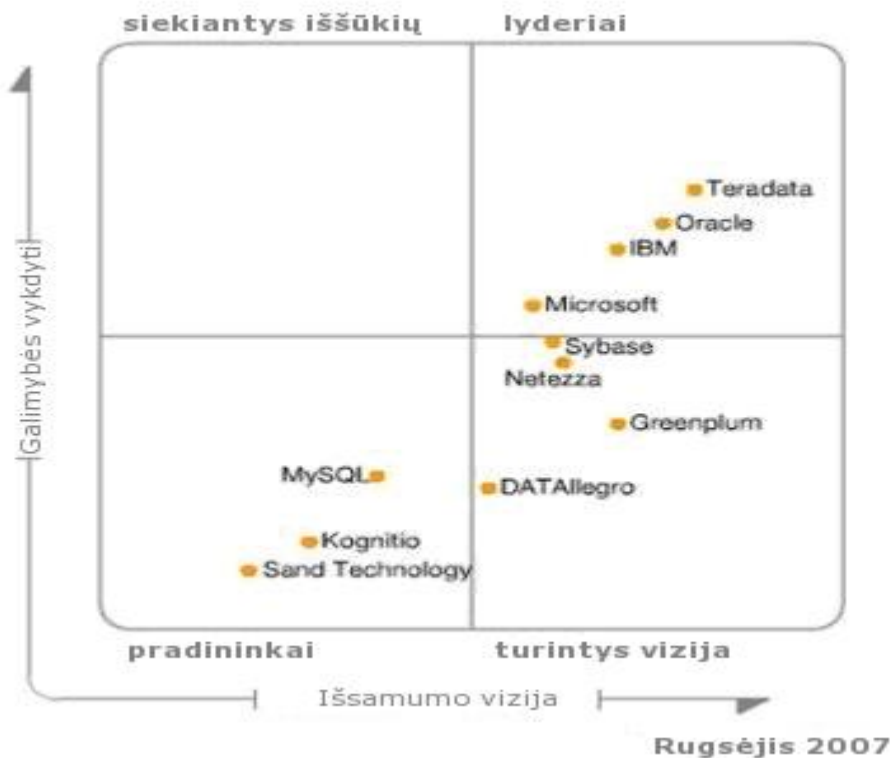
Įsitikinimo laipsnis narystės eilėje	CLUSTER_PROBABILITY
Masyvas objektų, galimų įvykių	CLUSTER_SET
Tam tikro bruožo ID, kurio didžiausia koeficiento vertė	FEATURE_ID
Geriausia prognozė konkretaus modelio	PREDICTION
Sąnaudų matas konkrečiai prognozei	PREDICTION_COST
Duotos prognozės tikimybė	PREDICTION_PROBABILITY

Tai vienos iš labiausiai naudojamų Oracle DBVS analitinių funkcijų [12], kurios padeda efektyviai atlikti duomenų analizę panaudojant „*select*“ užklausas. Taip pat nuo Oracle 9 versijos atsirado ir duomenų gavybos funkcijos (Data Mining), kurių pagalba galima bandyti manipuluoti esamais duomenimis, tam, kad nuspėtumėme ateities įvykius.

Pagal analitinių funkcijų įvairumą ir kiekį Oracle turi įvairesnę analitinių funkcijų aibę nei Microsoft SQL Server, todėl tolimesnėje analizėje bus naudojamos šio produkto analitinės funkcijos.

Apibendrinimas

Renkantis duomenų bazės sistemą, reikia atsižvelgti į tam tikrus poreikius, kurie gali iškilti analizuojant sukauptus duomenis. „Gartner“ grupė, atlikusi 2007 metų duomenų bazių valdymo sistemų analizę [13], pristatė savo kasmetinėje ataskaitoje ir nurodė geriausių duomenų bazių valdymo sistemų kūrėjus, kurie pavaizduoti paveiksle „Paveikslas 7. Duomenų bazių valdymo sistemų kūrėjų pasiskirstymas, pagal "Gartner" grupės atliktus tyrimus“(psl. 20).



Paveikslas 7. Duomenų bazių valdymo sistemų kūrėjų pasiskirstymas, pagal "Gartner" grupės atliktus tyrimus

Kaip matome geriausių kūrėjų ketvertukas išsidėstė tokia tvarka:

1. Teradata;
2. Oracle;
3. IBM;
4. Microsoft.

Išanalizavus ekspertų pateiktus duomenis ir pagrindinių kūrėjų duomenų saugyklų analitines funkcijas, patikimiausia rinktis Oracle. Oracle turintis didelį kiekį analitinių funkcijų ir grafinių analitinių priemonių padės efektyviai analizuoti veiklą. Jeigu ateityje Microsoft SQL Server DBVS bus integruotos alternatyvios analitines funkcijos ir bus įmanoma pasinaudoti grafinėmis analitinėmis priemonėmis, atlikti veiklos analizę bus įmanoma ir Microsoft SQL Server įrankiu.

Lietuvoje Oracle naudoja daugelis valstybinių institucijų, telekomunikacijų bendrovės ir kitos vidutinės bei didelės verslo bendrovės [14]. Šis faktas įrodo Oracle populiarumą, todėl daugeliui Oracle panaudojimo galimybės gali atnešti dar didesnės naudos, atliekant veiklos analizę su turimomis priemonėmis. Labai svarbus ir tas faktas, kad naudojant Oracle produktus kartu su SAS įrankiu, kyla resursų ir aptarnavimo kaina, todėl sprendimas naudoti SAS įrankį labai neefektyvus, kai galima naudoti Oracle priemonėse integruotus sprendimus.

Analitinių funkcijų panaudojimas

Analitinių funkcijų panaudojimas pritaikomas įvairiose srityse, kuriose galimi statistinių uždavinių sprendimai. Dažniausiai tai gali būti prekybos įmonės prekių analizės sprendimai, telekomunikacijų bendrovėse vartotojų analizės sprendimai, mažų ir vidutinių įmonių klientų veiklos analizės sprendimai, draudimo įmonių vartotojų statistinės analizės sprendimai.

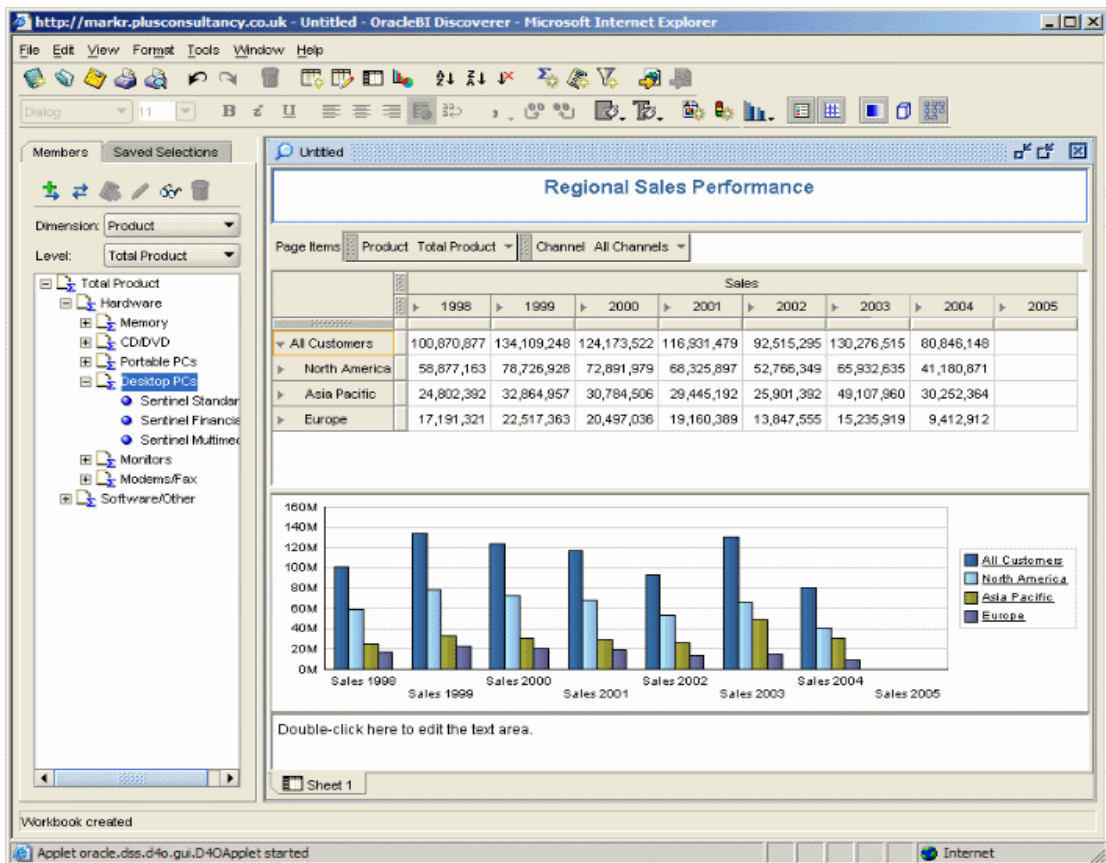
Atliekant veiklos analizę galima pasirinkti įvairią sritį ir joje pritaikyti analitines funkcijas įvairiems uždaviniams spręsti. Sprendimą pritaikyti galima konkrečiai sričiai, kadangi analitinių funkcijų įvairovė ir panaudojimas labai platus, todėl panaudoti galima įvairiose veiklose tik iškeliant vis kitokius statistinius uždavinius ir juos sprendžiant panaudojus užklausas ir analitines funkcijas [15,17].

Keletas uždavinių formulavimų duomenų saugyklos aplinkoje:

1. Ekstremumo paieška pagal dimensija - kokia dimensija daro didžiausią įtaką rodikliui (koreliacija)?
2. TOP-N užklausa įrašo lygyje, ekstremumo paieška pagal konkretaus rodiklio reikšmę, trejetukas didžiausių pasirašytų įmokų grupuojant pamėnesiui.
3. TOP-N užklausa grupės lygyje, ekstremumo paieška pagal konkretaus rodiklio suminę reikšmę (funkcija SUM), trejetukas produktų, kuriems pasirašyta daugiausiai įmokų pamėnesiui.
4. LAG ir LEAD funkcijų pavyzdys palyginamosiose ataskaitose: kiek pasirašyta ši mėnesį, kiek buvo praeitą mėnesį. Augimo ataskaita pagal ataskaitinio periodo pradžią - sausio mėnesį.
5. Aprašomoji statistika, tipinių - netipinių reikšmių analizė. Sugrupuota pagal mėnesį, produktą.

Štai tokie ir panašūs uždaviniai gali būti naudojami įvairiose veiklose, tik vienoje uždaviniai formuluojami vienaip išskiriant svarbesnius rodiklius, kitose atsižvelgiama į visai kitus rodiklius, kurie parodo vis kitokią veiklos efektyvumą.

Šių uždavinių sprendimui užklausas galima panaudoti ne tik duomenų bazės formate, bet ir įdėti į vaizdavimui skirtas priemones su grafikais tam, kad pabrėžtumėme esamų uždavinių siekiamą rezultatą. Oracle turi integruotas priemones, kuriomis galima pasinaudoti – tai Oracle Discoverer įrankis, kuris prisijungia prie esamos duomenų bazės ir vykdo sukurtą užklausą bei pateikia grafinį rezultatą, kaip pateikta paveiksle „Paveikslas 8. Oracle Discoverer užklausų vaizdavimas“ (psl. 22).



Paveikslas 8. Oracle Discoverer užklausų vaizdavimas

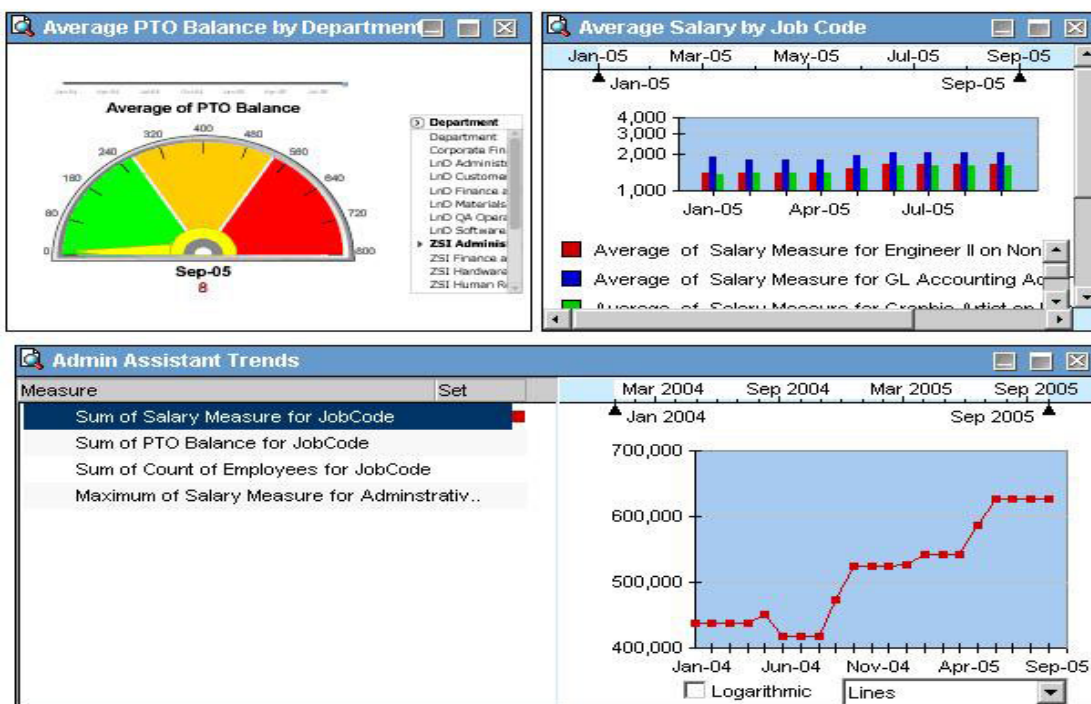
Oracle įrankis Oracle BI Publisher taip pat palaiko užklausų įvedimą ir teikia įvairius grafikos atvaizdavimo variantus, kaip pateikta paveiksle „Paveikslas 9. Oracle BI Publisher grafinis atvaizdavimas“(psl. 23).



Paveikslas 9. Oracle BI Publisher grafinis atvaizdavimas

Dar vienas įrankis, tačiau jau nebe Oracle priemonių linijai priklausantis – Business Objects (BO) paketas, priklausantis SAP kompanijai. BO paketas pasižymi dideliu verslo analizės (*business intelligence*) priemonių rinkiniu – įvairūs grafikai, skalės, efektyvių rodiklių identifikacija, valdymo langai greitai tendencijų identifikacijai (*scorecard & dashboard*).

Šis produktas labai populiarus tarp verslo analizės specialistų ir puikiai tinkamas naudoti su Oracle duomenų saugyklomis, kadangi įrankio sąsaja pritaikyta dirbti su šia duomenų bazės valdymo sistema. Business Objects vartotojo sąsaja parodyta paveiksle „Paveikslas 10. Business Objects grafinis vaizdavimas“ (psl. 24).



Paveikslas 10. Business Objects grafinis vaizdavimas

1.7. Analizės išvados

Išanalizuota bendra įmonės veikla ir jai daranti įtaką veiklos analizė, panaudojant analitines priemones.

Išanalizuoti sprendimai SAS ir Microsoft Excel, kurie padeda analizuoti įmonės duomenis ir priimti sprendimus, tačiau abu sprendimai turi savybių, kurias galima pagerinti naudojant užklausas su integruotomis analitinėmis funkcijomis Oracle aplinkoje, o ateityje ir Microsoft SQL Server aplinkoje.

Atliekant analitinių funkcijų panaudojimo analizę, buvo rasta informacijos apie šių funkcijų realų taikymą veiklos analizėje, todėl šis tyrimas turėtų parodyti, kokias galimybes galima pasiekti su įprastomis DBVS analitinėmis funkcijomis konkrečioje įmonės veikloje.

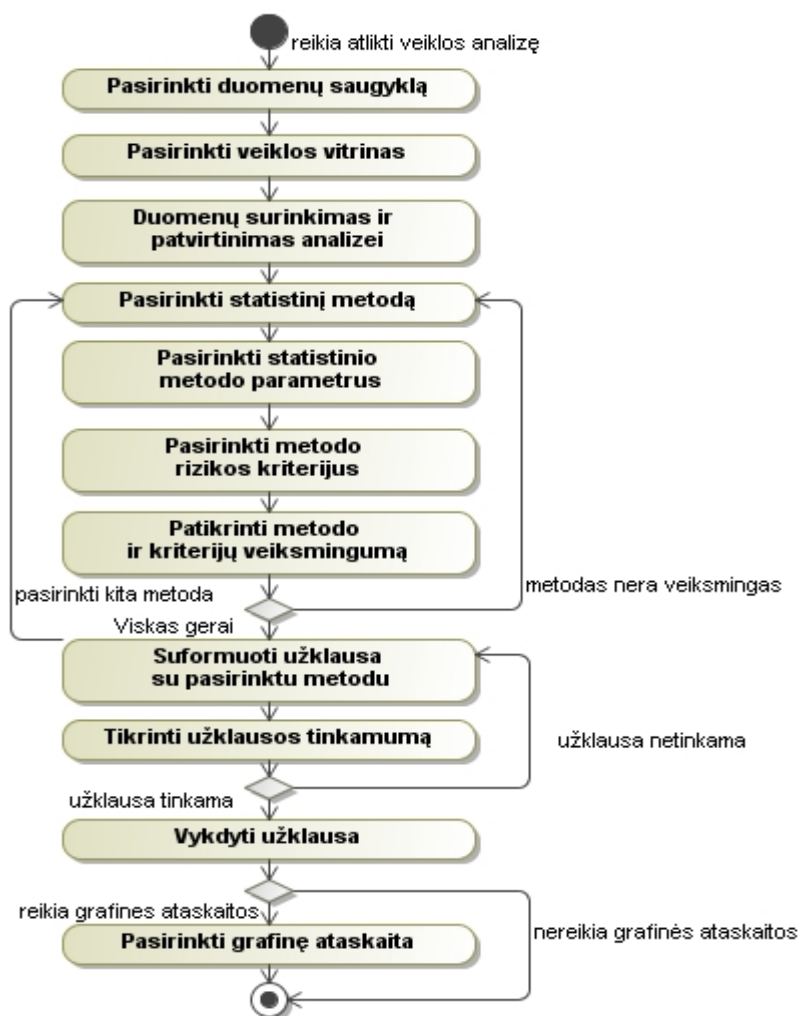
Atlikta išsami analizė analitinių funkcijų ir panaudojimas Oracle duomenų saugykloje, pateikti statistinių uždavinių pavyzdžiai ir funkcijų panaudojimas verslo analizėje.

Atlikta SQL užklausų integravimo analizė SQL kalbą palaikančiuose įrankiuose ir buvo nustatyta, jog daugelyje produktų SQL užklausas galima atvaizduoti grafiškai.

2. Sistemos reikalavimai

2.1. Reikalavimų specifikacija

Šiuo metu veiklos analizė atliekama sudėtingai ir užima daug laiko. Analitiniai įrankiai sukurti pasitelkiant sudėtingą metodiką ir norint puikiai dirbti analitiniais įrankiais reikia keletą metų nuolatinio darbo. Veiklos procesas, panaudojant šiuolaikinius analitinius įrankius, pavaizduotas paveiksle „Paveikslas 11. Veiklos analizės procesas, naudojant šiuolaikinius analitinius įrankius“.



Paveikslas 11. Veiklos analizės procesas, naudojant šiuolaikinius analitinius įrankius

Pavaizduotame veiklos procese vieni iš pagrindinių ir vieni iš svarbiausių veiksmų yra „Duomenų surinkimas ir patvirtinimas analizei“, „Pasirinkti statistinį metodą“, „Pasirinkti metodo rizikos kriterijus“ ir „Suformuoti užklausa su pasirinktu metodu“.

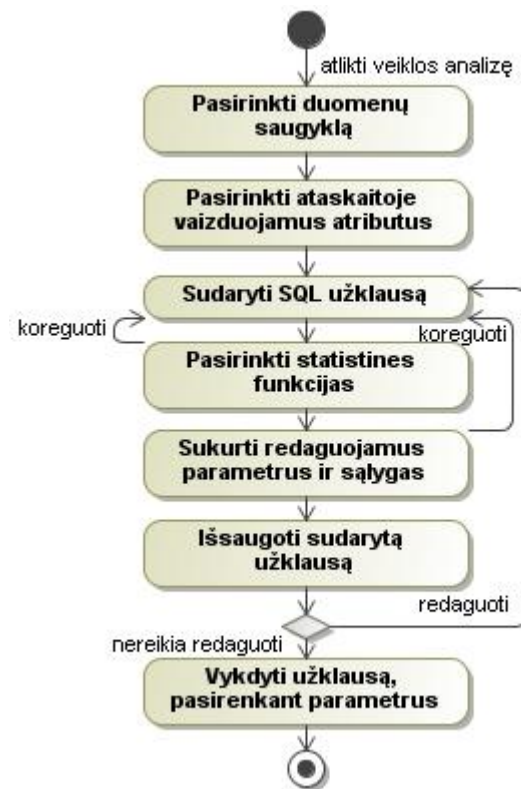
Šiuolaikiniai analitiniai įrankiai vis sudėtingėja ir su šiais įrankiais įmanoma analizuoti duomenis kurie yra saugomi tiek Oracle ar MS SQL Server duomenų bazėse, todėl iškyla

problemų vartotojui pasirenkant duomenų masyvus duomenų pasirinkimo punkte „Duomenų surinkimas ir patvirtinimas analizei“. Prie šio veiksmo daugelis vartotojų turi suprasti pačių analitinių įrankių galimybes nuskaityti tam tikrą kiekį duomenų iš skirtingų duomenų laikmenų ir tuos duomenis apdoroti bei saugoti fiziniėje atmintyje. Jei vartotojas nori analizuoti keleto mėnesių informacija, tuomet tokie įrankiai kaip SAS norimus analizuoti duomenis saugo laikinoje fiziniėje atmintyje, tad jei duomenų saugykloje yra keliasdešimt milijonų įrašų duomenų, tuomet saugomas kiekis fiziniėje atmintyje smarkiai išauga.

Atlikus veiksmus ir patikrinus duomenų teisingumą, vartotojas turi pasirinkti koku statistiniu metodu bus atliekama pasirinktų duomenų analizė. Šiuolaikiniuose įrankiuose statistinių metodų pasirinkimas yra labai didelis, todėl norint pasirinkti tinkamą metodą reikia taip pat būti puikiai susipažinus su įrankio funkcijomis bei mokėti tas funkcijas taikyti reikiamu momentu. Tai yra viena didžiausių analitinių įrankių problemų, kadangi norint greitai ir tiksliai išanalizuoti tam tikrus analitinius uždavinius reikia žinoti kokia analitinė funkcija duoda tinkamą rezultatą. Funkcijų specifika žinoti būtina, kadangi nustatant metodo veiksmingumo ribas punkte „Pasirinkti metodo rizikos kriterijus“ reikia nurodyti procentinę metodo paklaidą. Šis punktas nurodo kokia yra statistinio metodo galima paklaida, ir jei vartotojas parenka per didelę arba per mažą paklaidą, rezultatas tampa netikslus. Šis punktas reikalingas analitiniams įrankiams, kurie operuoja sudėtingais algoritmais prieš pradėdant skaičiavimus. Jei įrankių vidiniams algoritmams nustačius, jog paklaida yra per didelė arba per mažą, metodo pasirinkimo procesą vartotojui gali tekti kartoti.

Jei vartotojui pavyko pasirinkti norimą statistinį metodą, tuomet vartotojas turi pradėti konstruoti užklausą punkte „Suformuoti užklausa su pasirinktu metodu“. Tokiuose įrankiuose kaip SAS užklausos formavimas atliekamas unikalia SAS kalba, todėl nežinant šios kalbos pagrindų vartotojui suformuoti teisingą užklausą bus sudėtinga. Kadangi statistinis metodas, kuriuo bus operuojami duomenys, jau pasirinktas ir patikrintas, tuomet įrankių algoritmai patys pritaiko šį statistinį įrankį sukurtai vartotojo užklausiai įvykdyti. Jeigu užklausa sudaroma teisinga, tuomet gaunamas ir rezultatas, kurio tikslumas paaiškėja vėliau, kai matomi veiklos analizės rezultatai, pagal kuriuos vartotojai gali spręsti ar rezultatai teisingi ar ne.

Pateiktas veiklos procesas naudojant šiuolaikinius įrankius yra ganėtinai sudėtingas ir reikalaujantis naudojamo įrankio žinių. Norint gauti analizės rezultatą, konkrečiam uždaviniui išspręsti, tenka atlikti daug veiksmų, kurių įgyvendinimas užima nemažai laiko. Siūlomas veiklos procesas, naudojant statistines DBVS funkcijas, pavaizduotas paveiksle „Paveikslas 12. Veiklos analizės procesas, panaudojant statistines DBVS funkcijas“.



Paveikslas 12. Veiklos analizės procesas, panaudojant statistines DBVS funkcijas

Pavaizduotas veiklos analizės procesas panaudojant DBVS funkcijas kartu su integruotu vaizdavimo įrankiu yra trumpesnis ir efektyvesnis nei analizuojant tą pačią veiklą panaudojus šiuolaikinius įrankius (žr. paveikslėlį „Paveikslas 11. Veiklos analizės procesas, naudojant šiuolaikinius analitinius įrankius“).

Veiklos analizės proceso veiksmi:

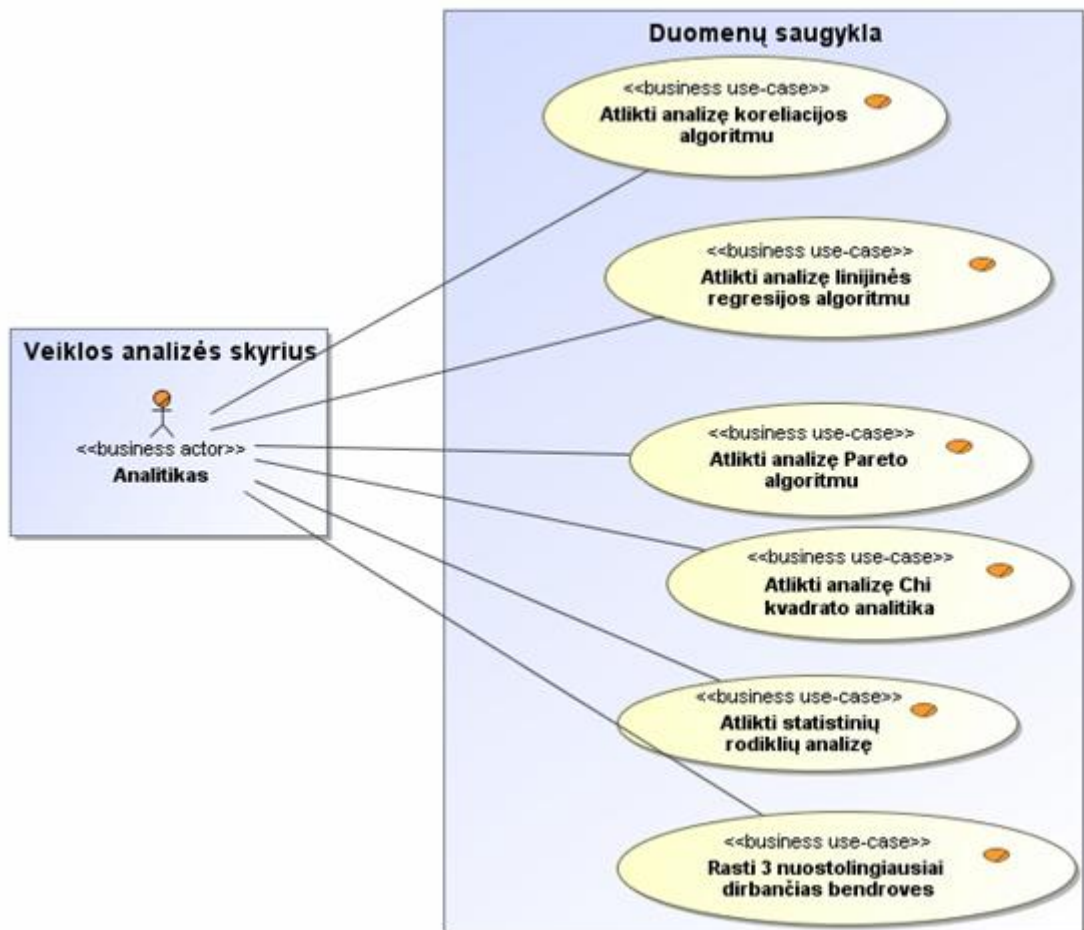
1. Vartotojas atlikdamas „Pasirinkti duomenų saugyklą“ veiksmą, naudojantis integruotu vaizdavimo įrankiu, pasirenka duomenų saugyklą, kurioje randasi reikalingi analizei duomenys. Pačioje saugykloje ir bus atliekama analizė, tai efektyvu, nes duomenys nėra paimami ir saugomi fizinėje atmintyje, o atliekama analizė pačioje duomenų saugykloje. Skaičiavimai atliekami pačioje duomenų saugykloje, kurioje yra integruotos DBVS statistinės funkcijos, todėl rezultatas yra gaunamas per mažesnę laiko tarpą. Pasirinktoje duomenų saugykloje bus saugoma visa informacija apie sukurtą užklausą ir ataskaitos tipą, todėl šią ataskaitą bus galima panaudoti keliose duomenų saugyklose, pakeičiant tik reikiamus atributus.
2. Vartotojui sėkmingai pasirinkus duomenų saugyklą, toliau atliekamas „Pasirinkti ataskaitoje vaizduojamus atributus“ veiksmas. Vartotojas pasirenka ataskaitos tipą ir pagrindinius ataskaitos atributus, kurie yra vaizduojami lentelės formoje. Šiuo atveju

atributai - tai duomenų saugyklos lentelių stulpelių pavadinimai, kuriuos vartotojas gali pasirinkti pagal vaizdavimo pobūdį ir sprendžiamo uždavinio tipą.

3. Sėkmingai pasirinkus atributus toliau vykdomas veiksmas „Sudaryti SQL užklausa“. Vartotojas sudaro užklausa, atitinkančią uždavinio formulavimą. Užklausa sudaroma panaudojant SQL kalbos standartus. Kadangi tai plačiai naudojamas ir pripažįstamas užklausių formavimo būdas, vaizdavimo įrankiuose tos integruotos standartinės užklaustos panaudojamos funkcijose.
4. Tarp integruotų funkcijų yra ir statistinės analizės funkcijos, kurias vartotojas pasirenka „Pasirinkti statistines funkcijas“ veiksme. Statistinių užklausių gausa nėra tokia didelė, kaip naudojant įrankį SAS, tačiau pradinei statistinei analizei funkcijų pakanka.
5. Pasirinkus statistines funkcijas, sudaromos užklaustos sąlygos ir vartotojo pasirenkami parametrai punkte „Sukurti redaguojamus parametrus ir sąlygas“. Sąlygos užklausoje naudojamos apibrėžti analizuojamos srities režius, o vartotojo nurodomi parametrai naudojami keisti užklaustos formatą ir padaryti pačią užklausa dinaminę, įvedant parametrus užklaustos vykdymo metu.
6. Išsaugojus užklausa punkte „Išsaugoti sudarytą užklausa“, užklausa su ataskaitos atributais ir duomenimis apie užklausa yra saugomi pačioje duomenų saugykloje, todėl galima sukurtas užklausas viešai publikuoti ir naudoti tik toje DBVS, kurioje buvo kurta užklausa.
7. Sėkmingai išsaugota užklausa įvykdoma „Vykdyti užklausa, pasirenkant parametrus“ punkte, kuriame, paleidžiant ataskaitos tipą prašoma pasirinkti užklaustos parametrus, kurie buvo nustatyti punkte „Sukurti redaguojamus parametrus ir sąlygas“.

Pateiktas veiklos analizės procesas, panaudojant statistines DBVS funkcijas užklausose, bus naudojamas šiame tyrime, kuriame bus naudojami draudimo veiklos duomenys.

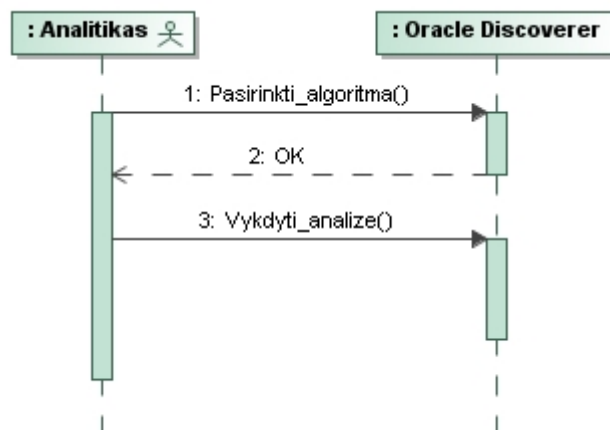
Šis tyrimas atliekamas siekiant pagerinti, paspartinti, atpiginti nedidelės draudimo bendrovės veiklos procesus, siekiant pagerinti darbo rezultatus. Atliekant draudimo statistinę analizę, galime naudotis panaudojimo atvejų diagrama, kuri pavaizduota paveiksle „Paveikslas 13. Draudimo įmonės veiklos statistinės analizės sistemos panaudojimo atvejų diagrama“(psl. 29).



Paveikslas 13. Draudimo įmonės veiklos statistinės analizės sistemos panaudojimo atvejų diagrama

Pateiktame paveiksle pavaizduoti pagrindiniai kasdieniai draudimo srities panaudojimo atvejai, kuriuos naudoja draudimo srities analitikai, analizuodami veiklos duomenis.

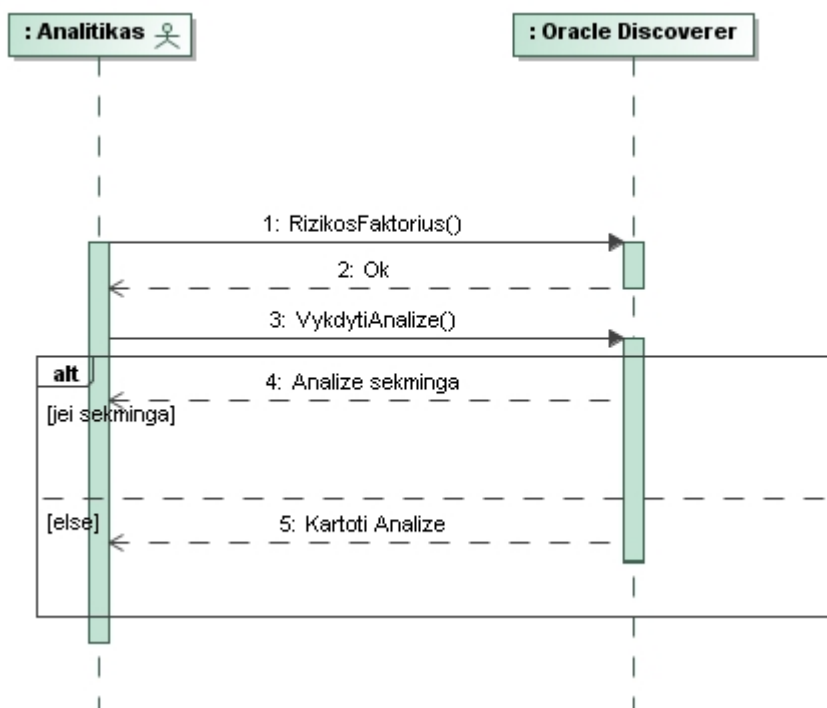
Panaudojimo atvejo „Atlikti analize Chi kvadrato analitika“ sekų diagrama pavaizduota paveiksle „Paveikslas 14. Panaudojimo atvejo "Atlikti analize Chi kvadrato analitika" sekų diagrama“.



Paveikslas 14. Panaudojimo atvejo "Atlikti analize Chi kvadrato analitika" sekų diagrama

Pateiktoje sekų diagramoje vaizduojamas Chi kvadrato analitikai spręsti sukurtos užklauso pasirinkimas. Oracle Discoverer įrankyje sukurtos užklauso saugomas pačioje duomenų saugykloje, todėl galima pasirinkti jau sukurtas užklauso arba kurti naujas.

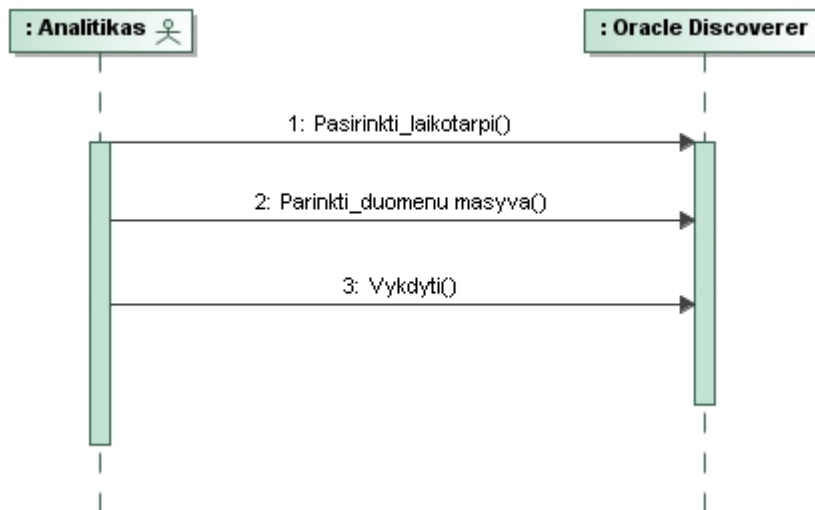
Panaudojimo atvejo „Atlikti analizę koreliacijos algoritmu“ sekų diagrama pavaizduota paveiksle „Paveikslas 15. Panaudojimo atvejo "Atlikti analizę koreliacijos algoritmu" sekų diagrama“.



Paveikslas 15. Panaudojimo atvejo "Atlikti analizę koreliacijos algoritmu" sekų diagrama

Sekų diagramoje vaizduojamas rizikos faktoriaus pasirinkimas Oracle Discoverer įrankyje, taip pasirenkant kintamąjį koreliacijos algoritmui sudaryti. Analizė vykdoma Oracle Discoverer įrankio pagalba, tačiau visi skaičiavimai atliekami duomenų saugykloje.

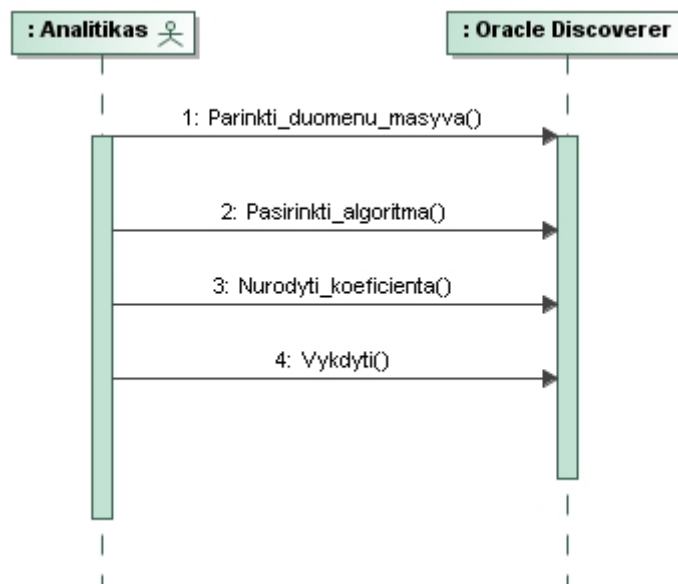
Panaudojimo atvejo „Atlikti analizę linijinės regresijos algoritmu“ sekų diagrama pavaizduota paveiksle „Paveikslas 16. Panaudojimo atvejo "Atlikti analizę linijinės regresijos algoritmu" sekų diagrama“ (psl. 31).



Paveikslas 16. Panaudojimo atvejo "Atlikti analizę linijinės regresijos algoritmu" sekų diagrama

Sekų diagramoje vaizduojamas laikotarpio pasirinkimas įrankyje Oracle Discoverer ir duomenų masyvo pasirinkimas norimam laikotarpiui. Analizė vykdoma Oracle Discoverer įrankio pagalba, tačiau visi skaičiavimai atliekami duomenų saugykloje.

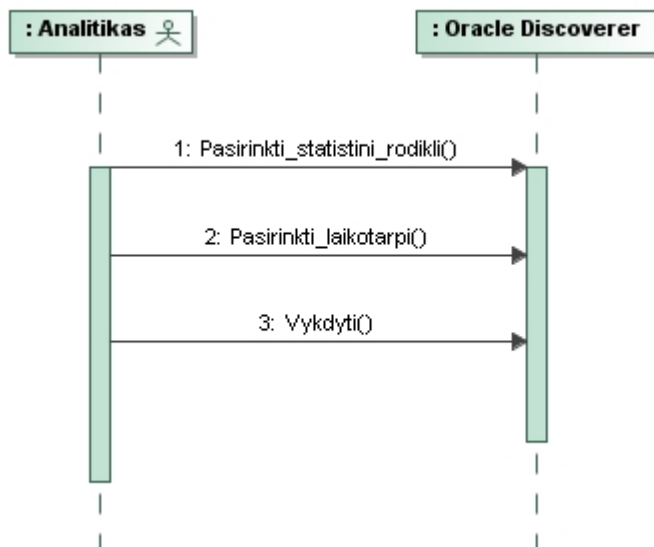
Panaudojimo atvejo „Atlikti analizę Pareto algoritmu“ sekų diagrama pavaizduota paveiksle „Paveikslas 17. Panaudojimo atvejo "Atlikti analizę Pareto algoritmu" sekų diagrama“.



Paveikslas 17. Panaudojimo atvejo "Atlikti analizę Pareto algoritmu" sekų diagrama

Sekų diagramoje pavaizduotas Pareto algoritmo pagalba atliekama analizė. Iš pradžių pasirenkamas duomenų masyvas, vėliau pasirenkamas Oracle Discoverer integruota Pareto algoritmo funkcija, nurodomas koeficientas. Analizė vykdoma Oracle Discoverer įrankio pagalba, tačiau visi skaičiavimai atliekami duomenų saugykloje.

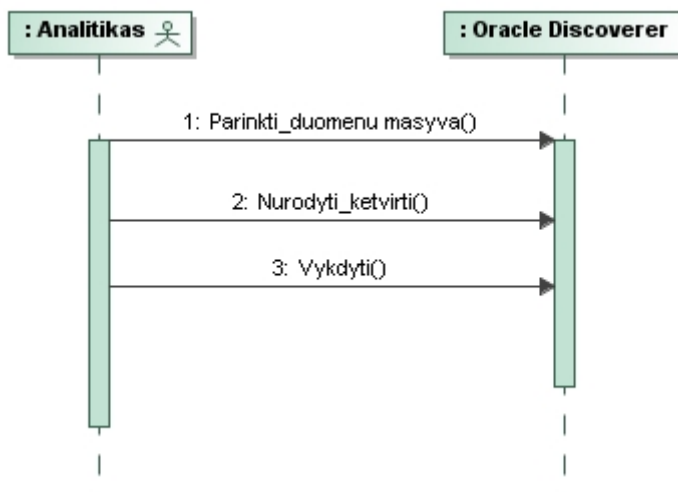
Panaudojimo atvejo „Atlikti statistinių rodiklių analizę“ sekų diagrama pavaizduota paveiksle „Paveikslas 18. Panaudojimo atvejo "Atlikti statistinių rodiklių analizę" sekų diagrama“.



Paveikslas 18. Panaudojimo atvejo "Atlikti statistinių rodiklių analizę" sekų diagrama

Sekų diagramoje vaizduojamas rodiklio ir laikotarpio pasirinkimas, tam kad būtų atlikta statistinių rodiklių analizė. Analizė vykdoma Oracle Discoverer įrankio pagalba, tačiau visi skaičiavimai atliekami duomenų saugykloje.

Panaudojimo atvejo „Rasti 3 nuostolingiausiai dirbančias bendroves“ sekų diagrama pateikta paveiksle „Paveikslas 19. Panaudojimo atvejo "Rasti 3 nuostolingiausiai dirbančias bendroves" sekų diagrama“.



Paveikslas 19. Panaudojimo atvejo "Rasti 3 nuostolingiausiai dirbančias bendroves" sekų diagrama

Sekų diagramoje vaizduojamas duomenų masyvo pasirinkimas ir ketvirčio pasirinkimas, tam kad būtų rasto trys nuostolingiausiai dirbančios bendrovės. Analizė vykdoma Oracle Discoverer įrankio pagalba, tačiau visi skaičiavimai atliekami duomenų saugykloje.

Kadangi visi panaudojimo atvejai yra panašūs ir skiriasi tik statistinių funkcijų panaudojimas, todėl lentelėje „Lentelė 7. Panaudojimo atvejo „Atlikti statistinių rodiklių analizę“ specifikacija“ pateikta specifikacija tinka visiems panaudojimo atvejams.

Lentelė 7. Panaudojimo atvejo „Atlikti statistinių rodiklių analizę“ specifikacija

PA „Atlikti statistinių rodiklių analizę“	
Tikslas. Atlikti draudimo veiklos rodiklių analizę	
Aprašymas. Šis PA rodiklių analizės tyrimą draudimo veikloje	
Prieš sąlyga	Duomenų saugykloje duomenys patalpinti sutvarkyti
Aktorius	Analitikas
Sužadinimo sąlyga	
Susiję panaudojimo atvejai	Išplečia PA
	Apima PA
	Specializuoja PA
Pagrindinis įvykių srautas	
1. Vartotojas pateikia užklausa	Duomenų saugykla pateikia duomenis į nurodytą užklausa ir atspausdina duomenis
2. Sistema baigia PA	
Po sąlyga:	
Alternatyvūs scenarijai	

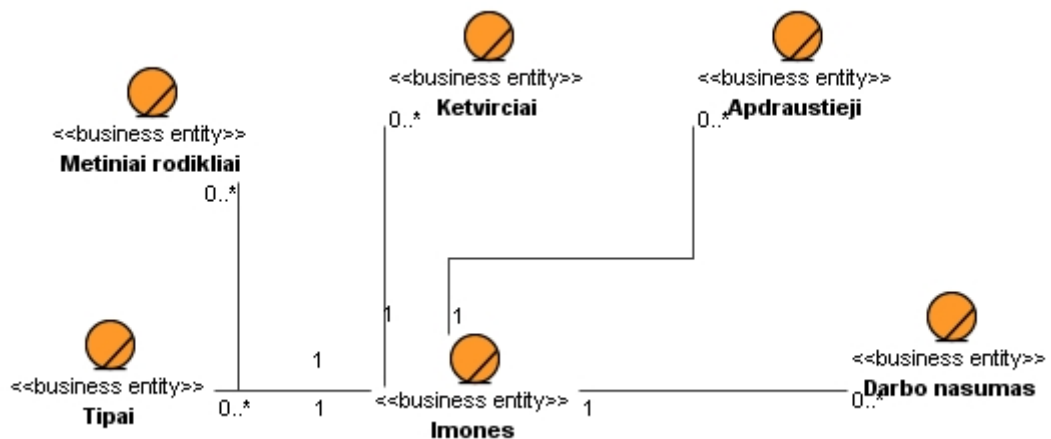
Panaudojimo atvejo „Atlikti statistinių rodiklių analizę“ veiklos procesai pavaizduoti paveiksle „Paveikslas 20. Panaudojimo atvejo "Atlikti statistinių rodiklių analizę" veiklos diagrama“(psl. 34).



Paveikslas 20. Panaudojimo atvejo "Atlikti statistinių rodiklių analizę" veiklos diagrama

Visi kiti panaudojimo atvejai yra tokie patys, tik skiriasi parametų išdėstymas ir kitų funkcijų pasirinkimas pagal analizuojamą veiklos uždavinį, todėl ši veiklos diagrama yra bendra visiems panaudojimo atvejams.

Vaizduojant veiklos konceptų modelį neišanalizavus pilnai realios draudimo srities yra gana sunku, tačiau abstraktų veiklos konceptų modelį galima pavaizduoti paveiksle „Paveikslas 21. Draudimo veiklos konceptų modelis“.

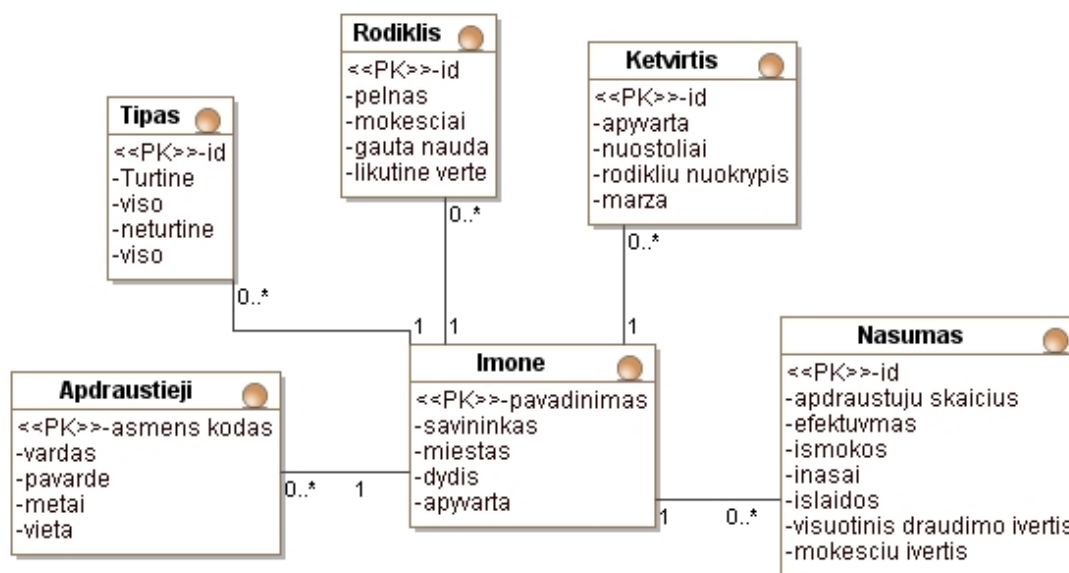


Paveikslas 21. Draudimo veiklos konceptų modelis

2.2. Dalykinės srities modelis

Dalykinės srities esybių klasių modelis reikalavimų specifikavimo etape turi būti visiškai apibrėžtas, tačiau nepriklausomas nuo būsimo projekto.

Koncepcinis dalykinės srities esybių modelis pritaikant stereotipus pavaizduotas paveiksle „Paveikslas 22. Koncepcinis dalykinės srities esybių modelis pritaikant stereotipus“.



Paveikslas 22. Koncepcinis dalykinės srities esybių modelis pritaikant stereotipus

Tolesnių projektavimo etapų metu jį galima papildyti susijusiais konceptais. Detalaus projektavimo etape aprašomas galutinis dalykinės srities modelis, kuris transformuojamas į duomenų bazės schemą.

2.3. Reikalavimų analizės apibendrinimas

Atliekant tyrimą pasirinktoje draudimo srityje buvo išanalizuoti reikalavimai ir dalykinės srities esybės. Pagal dabartinius reikalavimus dirbantys draudimo darbuotojai, analizuodami veiklą statistinėmis priemonėmis, atlieką sudėtingus ir užimančius nemažai laiko veiksmus. Šis tyrimas ir integruotos užklauskos turėtų sumažinti laiko sąnaudas bei supaprastinti atliekamus veiksmus. Išanalizuotas panaudojimo atvejo modelis bendrame ir detalizuotame panaudojimo atvejo modelyje bei pateiktas dalykinės srities esybių modelis, kuris leido suprojektuoti sistemos projektą statistinei analizei atlikti.

3. Sistemos projektas

3.1. Sistemos architektūros projektas

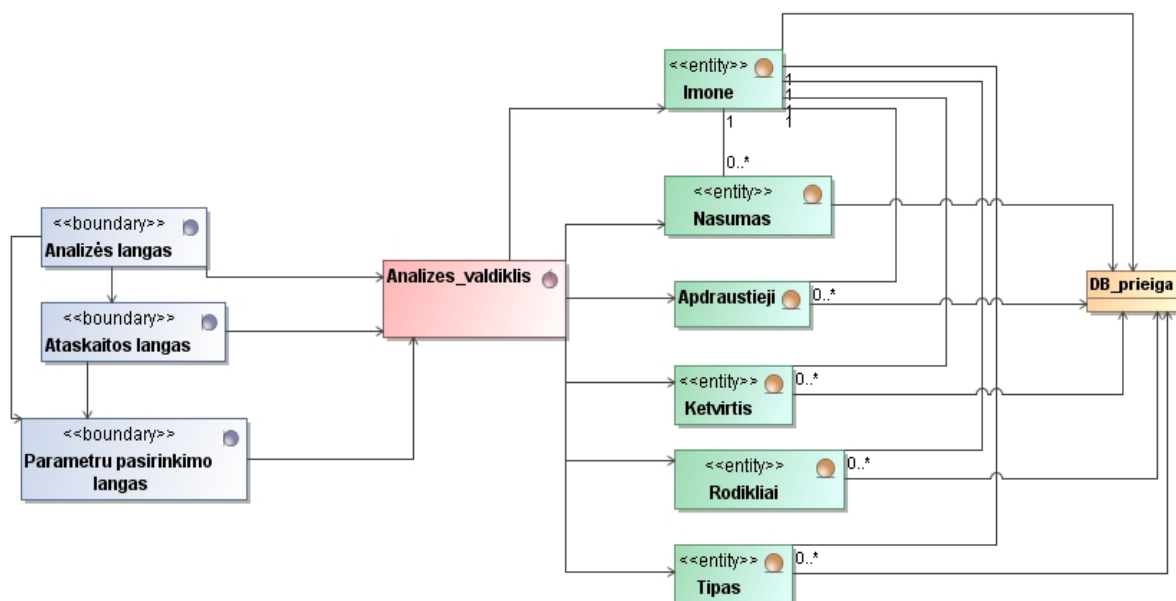
3.1.1. Sistemos loginė architektūra

Sistemos loginė architektūra susideda iš pagrindinio posistemio, statistinės analizės posistemio ir duomenų saugyklos posistemio, kaip pavaizduota paveiksle „Paveikslas 23. Sistemos loginė architektūra“.



Paveikslas 23. Sistemos loginė architektūra

Panaudojimo atvejų realizacija projekto klasėmis pateikta paveiksle „Paveikslas 24. Panaudojimo atvejų realizacija projekto klasėmis“. Diagrama parodo kokios esybės, valdikliai ir ribinės klasės sąveikauja tarpusavyje.

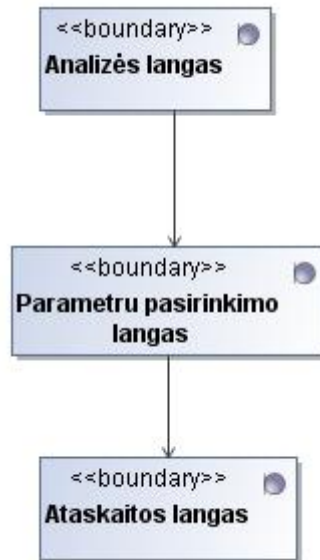


Paveikslas 24. Panaudojimo atvejų realizacija projekto klasėmis

Oracle Discoverer pagrindiniai langai pavaizduoti kaip „Analizės langas“, „Ataskaitos langas“ ir „Parametru pasirinkimo langas“. Analizės valdiklis saugo užklausų formas, kurias kaupia aprašytas savybes, kuriomis naudojasi kreipiantis į duomenų saugyklą.

3.1.2. Vartotojo paslaugos

Vartotojo navigavimo planas naudojant Oracle Discoverer įrankį pavaizduotas paveiksle „Paveikslas 25. Vartotojo sąsajos navigavimo planas“.

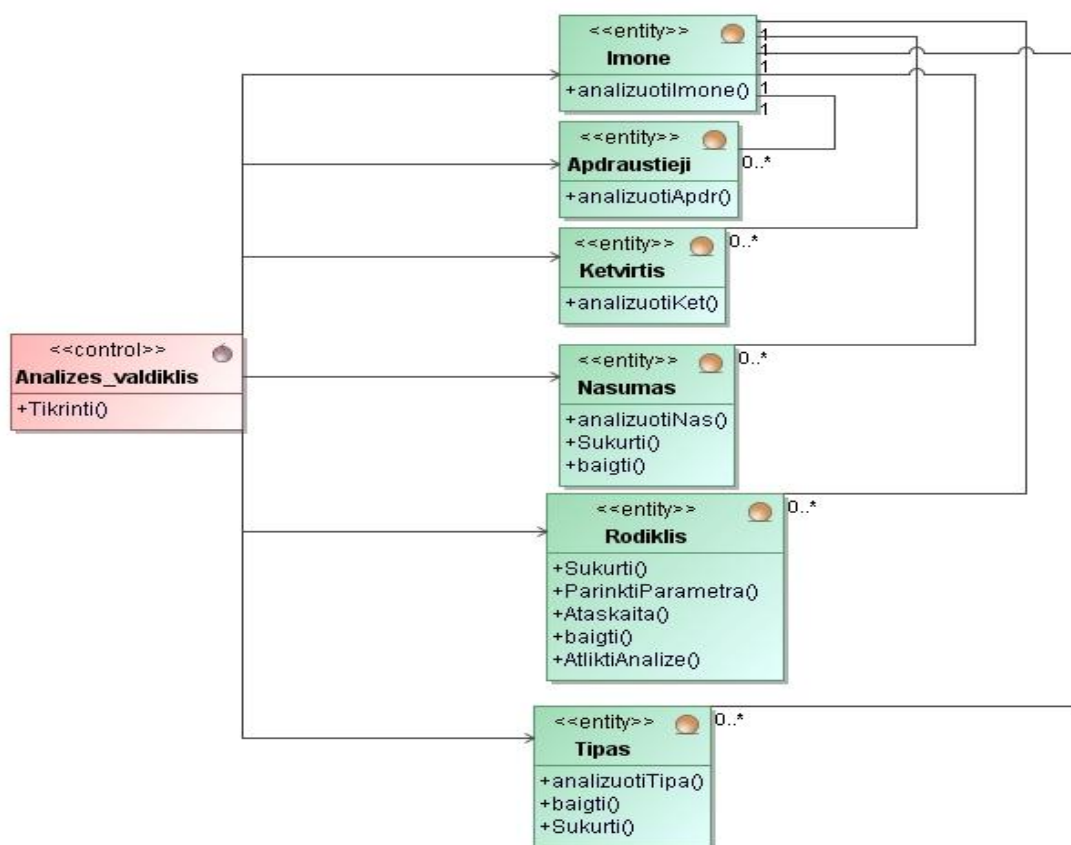


Paveikslas 25. Vartotojo sąsajos navigavimo planas

Vartotojas Oracle Discoverer įrankyje gali pereiti iš analizės lango į parametru pasirinkimo langą ir į ataskaitos langą.

3.1.3. Veiklos paslaugos

Veiklos paslaugos pateiktos paveiksle „Paveikslas 26. Valdiklių ir esybių klasių diagrama“ (psl. 38) parodo kokie valdikliai ir esybės sąveikauja tarpusavyje.

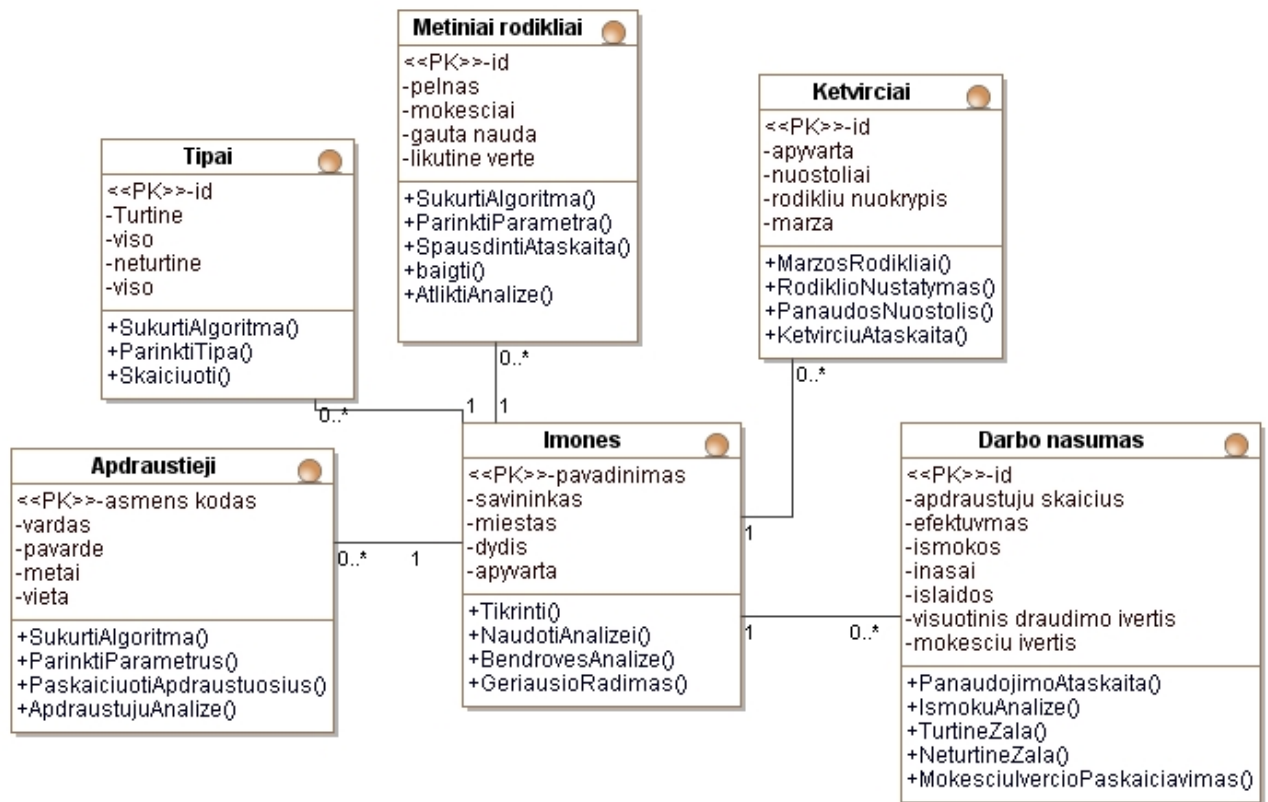


Paveikslas 26. Valdiklių ir esybių klasių diagrama

Analizės valdiklis naudoja „Imone“, „Apdraustieji“, „Ketvirtis“, „Nasumas“, „Rodiklis“, „Tipas“ esybes.

3.2. Detalus projektas

Draudimo srities klasių diagrama su operacijomis pavaizduota paveiksle „Paveikslas 27. Klasių diagrama su operacijomis“ (psl. 39).



Paveikslas 27. Klasių diagrama su operacijomis

Klasių diagrama parodo kokios funkcijos atliekamos kiekvienoje klasėje.

Klasė „Tipas“

Metodas	Paskirtis
SukurtiAlgoritma()	Algoritmo sukūrimas.
ParinktiTipa()	Tipo parinkimas iš pateikto sąrašo.
Skaiciuoti()	Tipo efektyvumui nustatyti atliekami skaičiavimai.

Klasė „Apdraustieji“

Metodas	Paskirtis
SukurtiAlgoritma()	Algoritmo sukūrimas.
ParinktiParametrus()	Parametro parinkimas iš pateikto sąrašo.
PaskaiciuotiApdraustuosius()	Apdraustųjų apskaičiavimas.
ApdraustujuAnalize()	Analizuoti apdraustųjų duomenis.

Klasė „Rodiklis“

Metodas	Paskirtis
SukurtiAlgoritma()	Algoritmo sukūrimas.

ParinktiParametrus()	Parametro parinkimas iš pateikto sąrašo.
SpausdintiAtaskaita()	Ataskaitos su rodikliais spausdinimas.
baigti()	Proceso užbaigimas.
AtliktiAnalize()	Analizuoti apdraustųjų duomenis.

Klasė „Imone“

Metodas	Paskirtis
Tikrinti()	Užklausos tikrinimas.
NaudotiAnalizei()	Pasirinktus duomenis naudoti algoritmo analizei atlikti.
Bendrovesanalize()	Atlikti esamų įmonių analizę.
GeriausioRadimas()	Geriausios įmonės pagal tam tikrą rodiklį radimas.

Klasė „Ketvirtis“

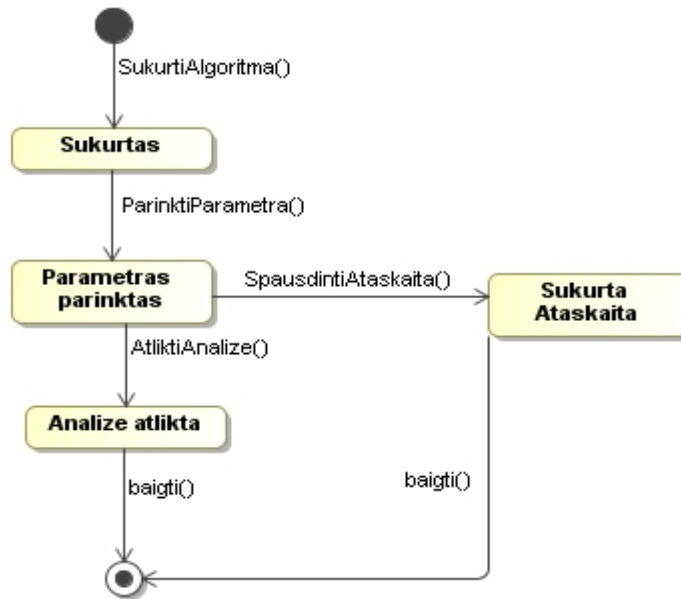
Metodas	Paskirtis
MarzosRodikliai()	Atliekami analizavimai skaičiuojant maržą tam tikram ketvirčiui.
RodiklioNustatymas()	Rodiklio parinkimo režimas.
PanaudosNuostolis()	Panaudos nuostolio analizės atlikimas.
KetvirciuAtaskaita()	Pasirinkto ketvirčio ataskaitos spausdinimas.

Klasė „Nasumas“

Metodas	Paskirtis
PanaudojimoAtaskaita()	Spausdinama ataskaita.
IsmokuAnalize()	Išmokų atlikimo analizė pasirinktam rodikliui.
TurtineZala()	Turtinės žalos radimas pasirinktai reikšmei.
NeturtineZala()	Neturtinės žalos radimas pasirinktai reikšmei.
MokesciuIvercioPaskaiciavimas()	Mokestinių rodiklių paskaičiavimas atsižvelgiant į reikšmes.

3.3. Sistemos elgsenos modelis

Sistemos bendra būsenų diagrama pavaizduota paveiksle „Paveikslas 28. Sistemos būsenų diagrama“(psl. 41).

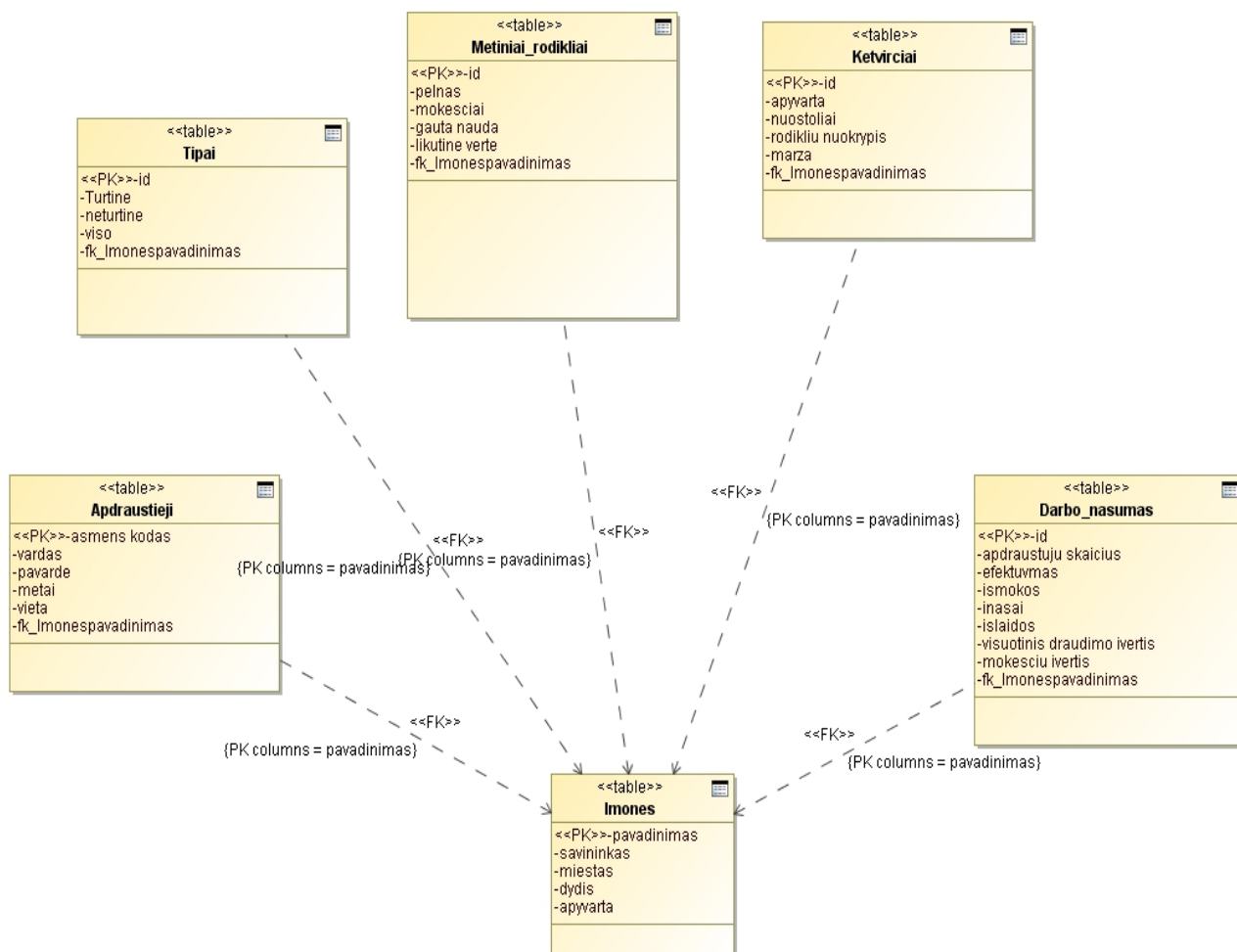


Paveikslas 28. Sistemos būsenų diagrama

Būsenų diagramoje pavaizduotas algoritmo sukūrimas, kuomet kuriama užklausa tam tikram uždaviniui išspręsti.

3.4. Duomenų bazės schema

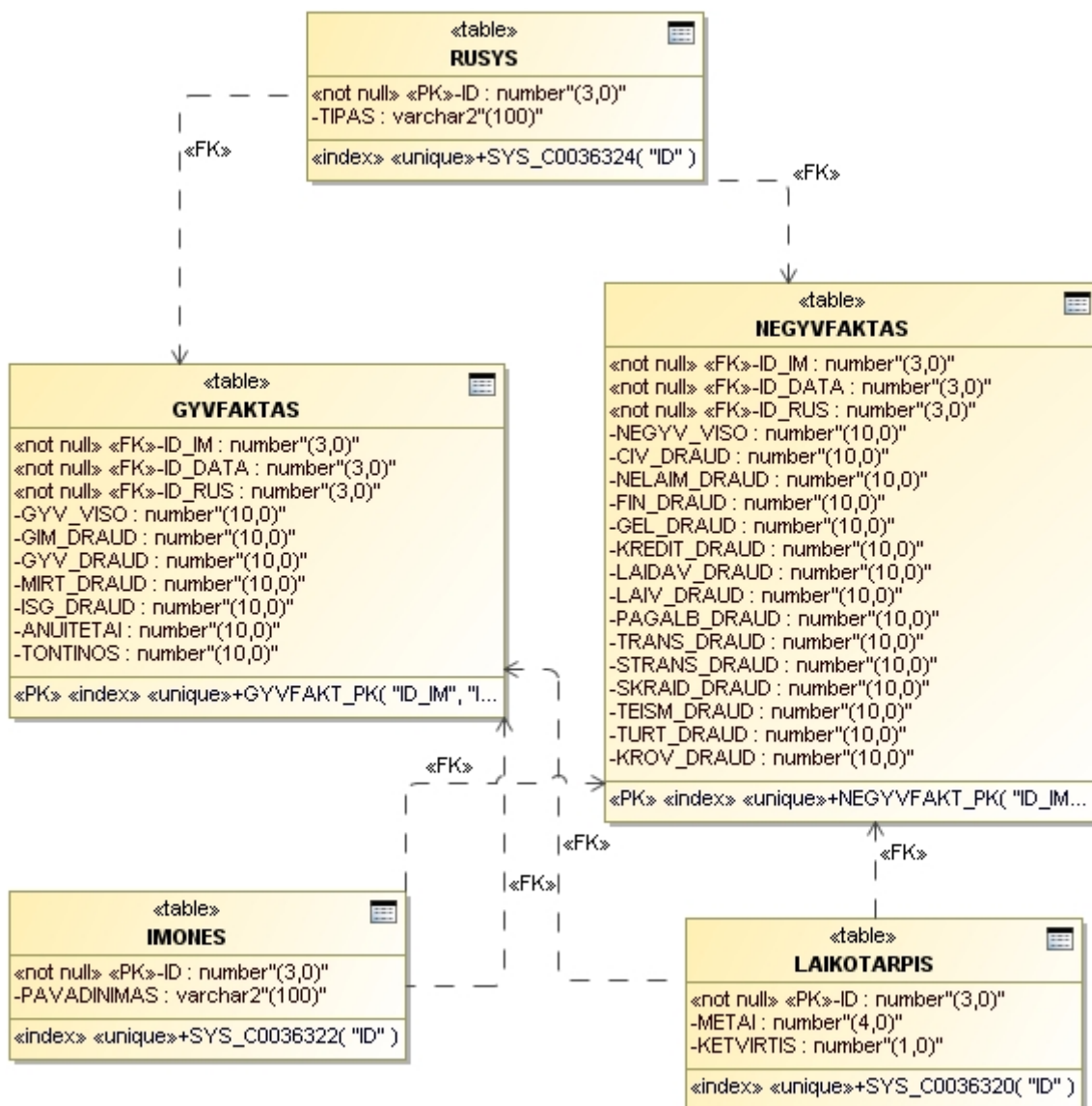
Duomenų bazės schema pavaizduota paveiksle „Paveikslas 29. Duomenų bazės schema“.



Paveikslas 29. Duomenų bazės schema

Duomenų bazė sudaryta iš lentelių „Imones“, „Apdraustieji“, „Tipai“, „Darbo_nasumas“, „Metiniai_rodikliai“ ir „Ketvirčiai“.

Statistiniai analizei atlikti ir šiam tyrimui atlikti reikalinga duomenų saugyklą, kurioje greitai būtų vykdomos užklausos ir saugoma įmonės informacija vienoje vietoje. Duomenų saugyklą sudaro faktai ir dimensijos. Faktai leidžia įvertinti verslo veiklą. Faktinėse lentelėse saugomi faktai apie draudimo srities veiklą, o dimensijų lentelėse – verslo šakos aprašai [16]. Draudimo srities duomenų saugyklos schema pavaizduota paveiksle „Paveikslas 30. Duomenų saugyklos schema“ (psl. 43).

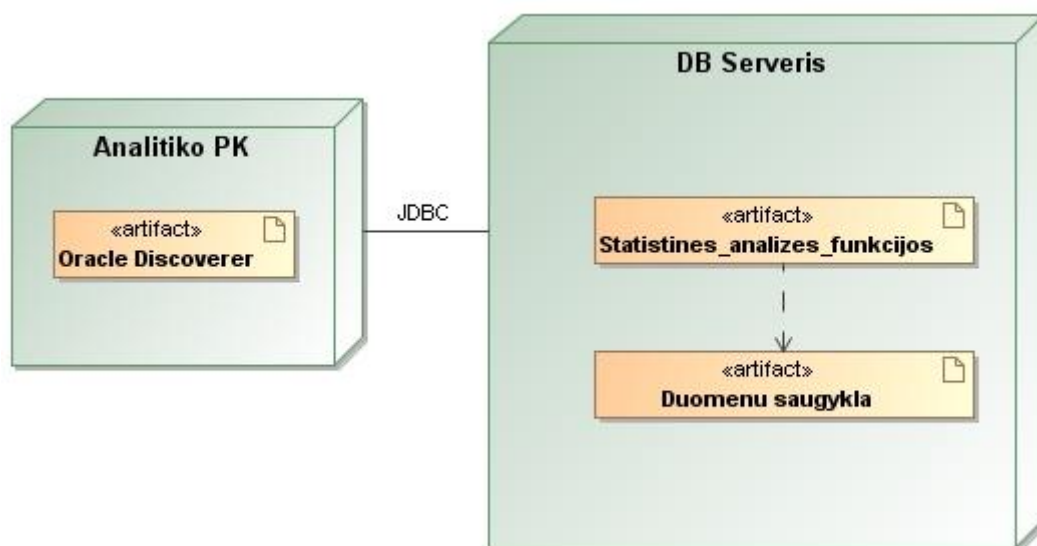


Paveikslas 30. Duomenų saugyklos schema

Duomenų saugyklos schemoje faktinės lentelės yra „Gyv faktas“ ir „Negyv faktas“, kurie nusako gyvybės ir ne gyvybės draudimo faktus. Dimensijų lentelės yra „Imonės“, „Laikotarpis“ ir „Rusys“.

3.5. Realizacijos modelis

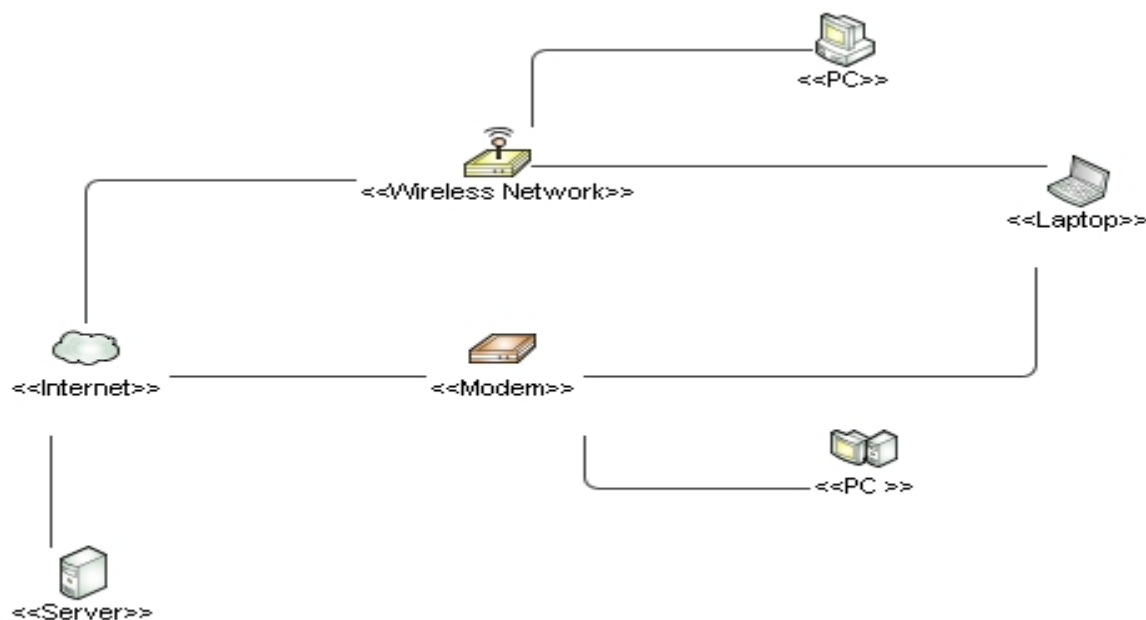
Sistemos diegimo planas pavaizduotas paveiksle „Paveikslas 31. Sistemos diegimo planas“ (psl. 44).



Paveikslas 31. Sistemos diegimo planas

Sistemos diegimo plane pavaizduota, kaip siejasi „Analitiko PK“ ir „DB Serveris“. „Analitiko PK“ yra analitiko darbo vieta – asmeninis kompiuteris. Kompiuteryje analitikas naudoja „Oracle Discoverer“ įrankių, kurio pagalba yra sukuriamos ir naudojamos statistinės analizės funkcijos užklausoje duomenų bazės serveryje, kurių pagalba analizuojami duomenys duomenų saugykloje.

Techninė sistemos architektūra pavaizduota paveiksle „Paveikslas 32. Techninė informacinės sistemos architektūra“.



Paveikslas 32. Techninė informacinės sistemos architektūra

4. Statistinės analizės, panaudojant DBVS funkcijas, realizacija

4.1. Realizacijos veikimo aprašymas

Eksperimentui bus naudojami draudimo srities duomenys, kurie yra realūs ir skelbiami Lietuvos Respublikos draudimo priežiūros komisijos tinklapyje <http://www.dpk.lt>. Reikalingi eksperimentui duomenys yra saugomi Oracle duomenų bazėje, kurioje ir vykdomos užklausos norint atlikti duomenų statistinę analizę. Sprendžiant uždavinius naudojamas Oracle Discoverer įrankis, kurio pagalba vykdoma grafinė duomenų analizė, panaudojant sudarytas užklausas.

Užklausos šablono sudarymas

Užklausos sudaromos naudojantis SQL kalbos standartais. Norint statistiškai analizuoti duomenis nustatytais veiklos srities aspektais, reikia sudaryti SQL užklausa su statistinėmis funkcijomis, kad būtų įvykdytas veiklos srities uždavinys.

Sudarant SQL užklausa galima remtis užklausos sudarymo šablonu, kuris suskirstytas į tris pagrindines dalis:

1. SQL užklausa prasideda žodžiu *SELECT*, po kurio turi būti pasirenkamas reikalingas duomenų masyvas, reikalingų parametrų išskyrimas panaudojant įvairias statistines funkcijas.

SELECT parametras1 Miestas,

LAG(sum(parametras2)) OVER (PARTITION BY parametras1 ORDER BY parametras3)

Rodiklis,

SUM(parametras4) Metinis_nuostolis,

Pateiktame šablono pavyzdyje pasirenkami parametrai *parametras1*, *parametras2*, *parametras3*, *parametras4* esantys reliacinėje duomenų bazėje. Šis pavyzdys parodo kaip galima modifikuoti parametrus naudojant tam tikras statistines funkcijas (Lentelė 6. Oracle analitinių funkcijų aprašymas), taip išskiriant reikalingus duomenų masyvo rodiklius atvaizdavimui lentelės formatu arba grafiškai.

2. Sekantis SQL užklausos elementas – *FROM*. Šiuo žodžiu prasideda reliacinių duomenų lentelių priskyrimas panaudojant ryšius. Duomenys yra paimami iš vienos arba daugiau lentelių.

FROM parametras1 g, parametras2 l,parametras3 m

Pateiktame šablono pavyzdyje lentelės pavadintos *parametras1*, *parametras2*, *parametras3* yra SQL užklausoje duomenų šaltiniai reliacinėje duomenų bazėje.

3. SQL užklausoje taip pat reikalingas sąlygos sakiny, prasidedantis žodžiu *WHERE*. Šis sąlygos sakiny apibrėžia tam tikrus kriterijus, kaip turėtų būti ieškomi duomenys.

WHERE parametras1 = parametras2 and parametras3 = parametras4 and parametras1 = 1 and parametras2 like 'Bon%'

Pateiktas šablono pavyzdys rodo parametrų *parametras1*, *parametras2*, *parametras3*, *parametras4* SQL užklausoje sąlygas prie pasirinktų kriterijų.

Bendra aprašyta šablono struktūra atrodo taip:

SELECT {Pasirenkami parametrai, kurie turėtų būti atvaizduoti tiek lentele tiek grafiškai. Parametrai turi atvaizduoti pateikto uždavinio prasmę. Kiekvienas pasirenkamas parametras gali būti aprašytas statistinėmis funkcijomis }

FROM {Pasirenkamos duomenų lentelės, iš kurių bus imami duomenys.}

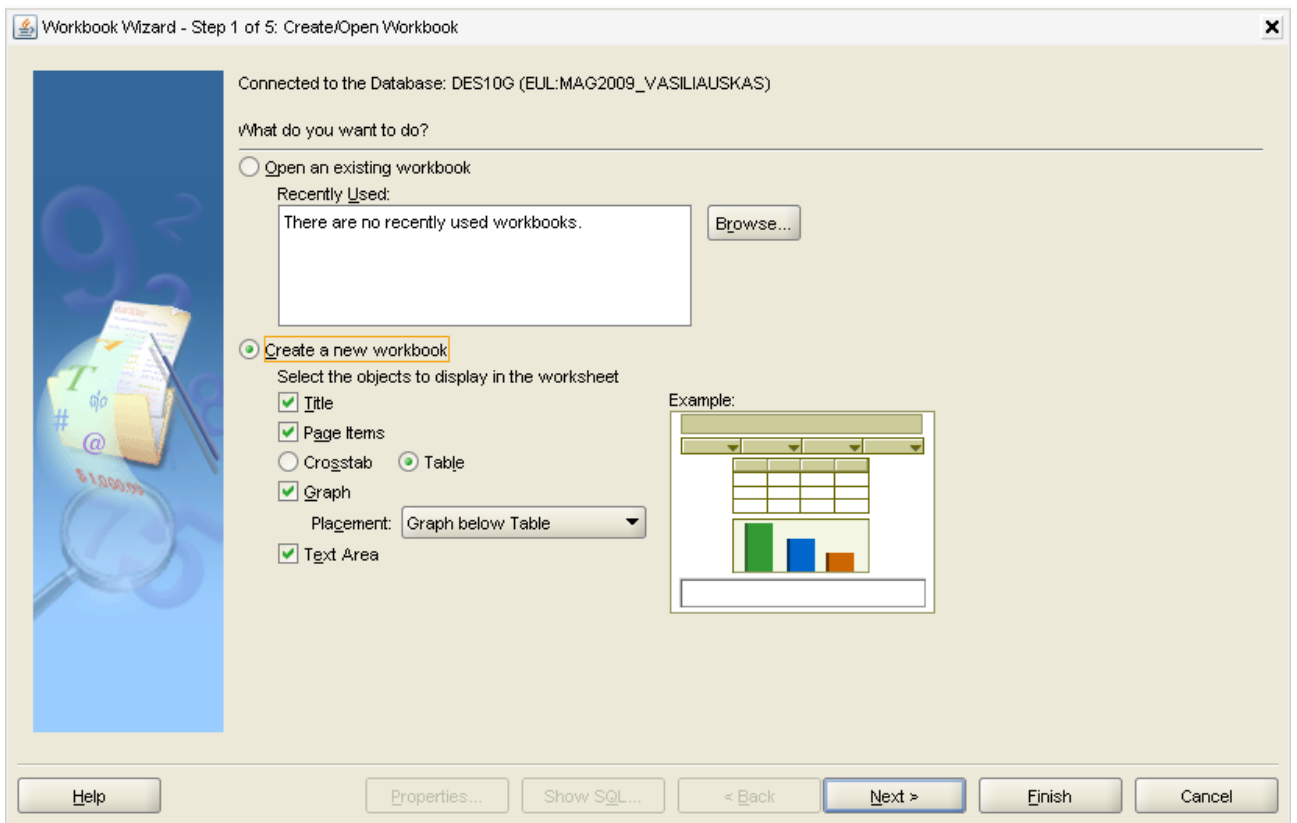
WHERE {Nustatomos pateikto uždavinio sąlygos ir aprašomos.}

Sudarius SQL užklausoje remiantis veiklos pateiktais rodikliais ir suformuluotu klausimu, šią užklausoje galima talpinti Oracle Discoverer įrankyje, rezultatų atvaizdavimui grafiškai ir ataskaitos sukūrimui.

Užklausoje taikymas Oracle Discoverer įrankyje

Įvykdyti ir gauti rezultatą naudojamas Oracle Discoverer įrankis, kuris turi integruotas analitines funkcijas ir vykdo užklausoje ir gražina rezultatą Oracle serveryje, taip efektyviai naudodamas resursus.

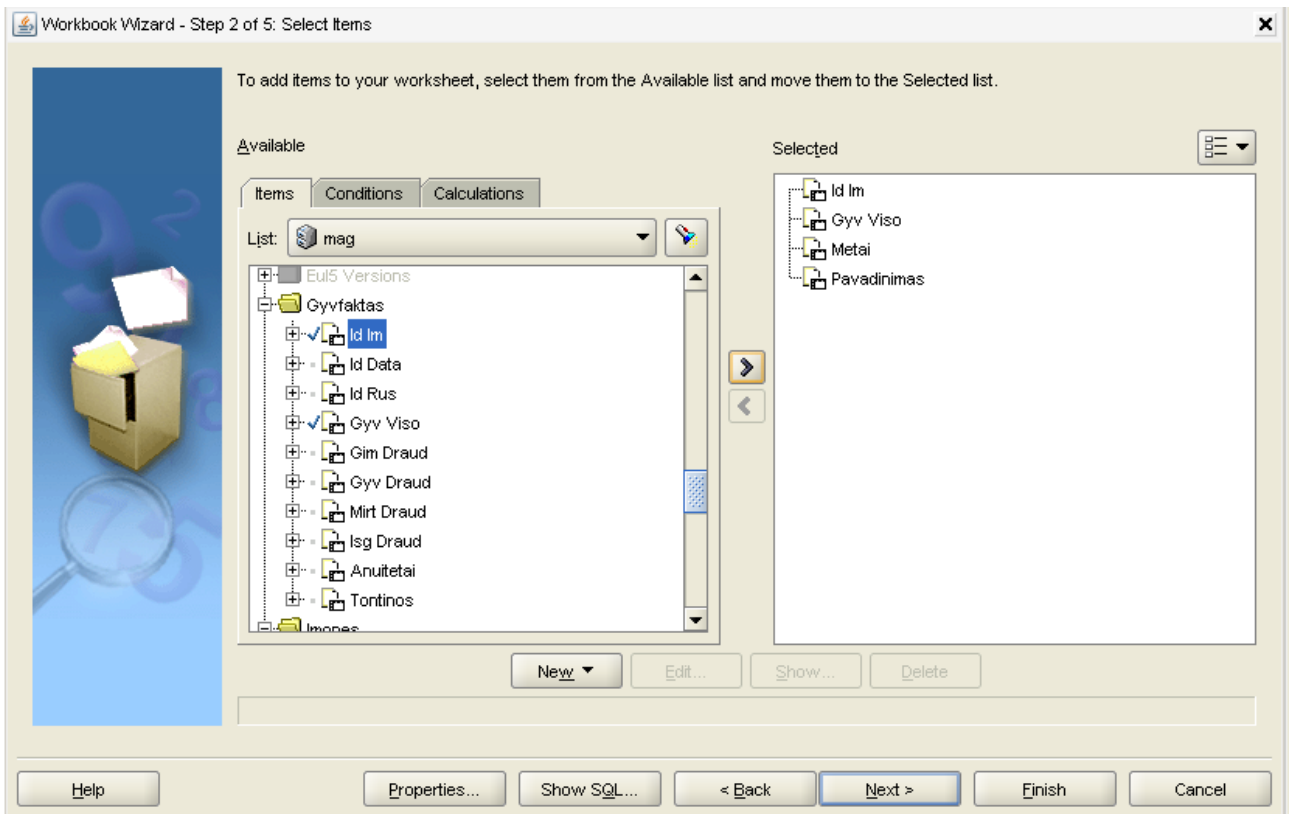
Pradinis Oracle Discoverer įrankio langas pavaizduotas paveiksle „Paveikslas 33. Oracle Discoverer ataskaitos kūrimo langas“ (psl. 47).



Paveikslas 33. Oracle Discoverer ataskaitos kūrimo langas

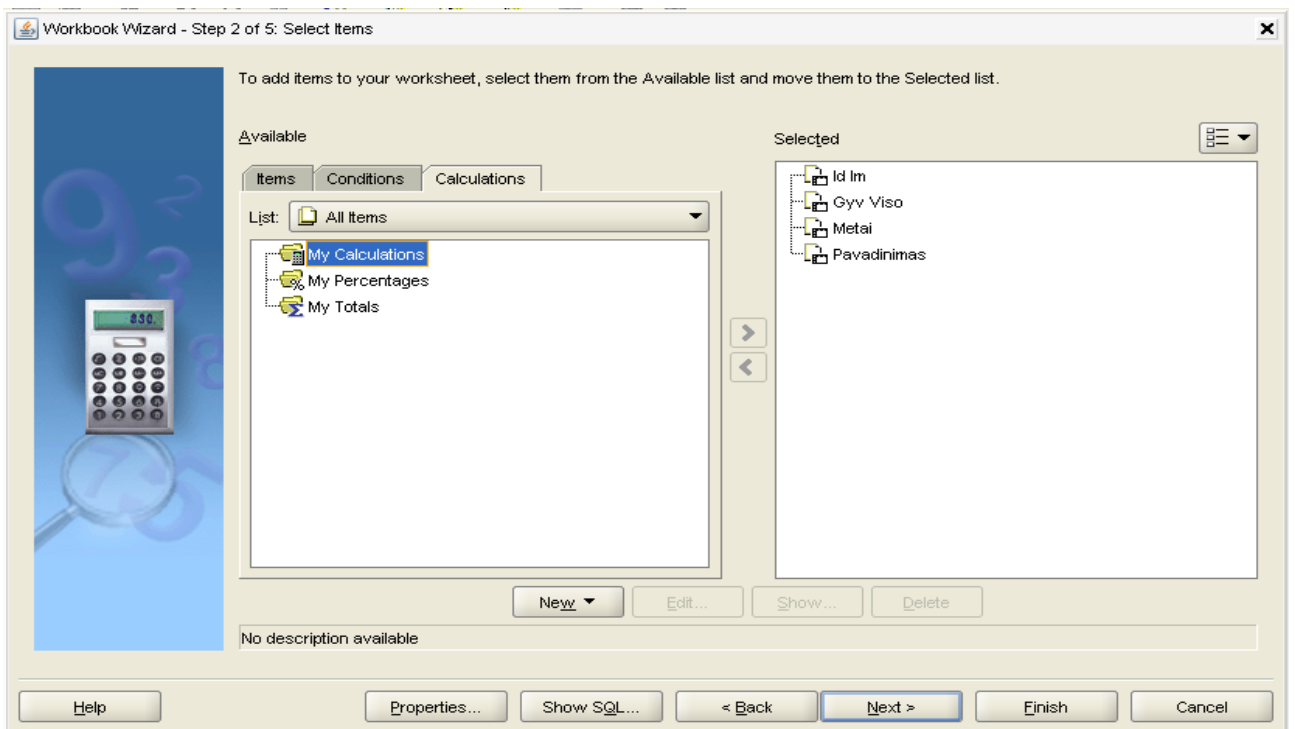
Pasirinkus sukurti naują ataskaitą ir pažymėjus norimus atvaizdavimo ataskaitoje būdus, spaudžiame „Next“.

Atsidariusiame lange „Paveikslas 34. Ataskaitos komponentų parinkimas ir užklauso sudarymas“ (psl. 48), pasirenkame tuos lentelių stulpelius, kurie bus naudojami sudarant SQL užklauso. Šis parinkimas stulpelių atstoja SQL užklauso „Select“ dalį.



Paveikslas 34. Ataskaitos komponentų parinkimas ir užklauso sudarymas

Pasirinkę reikalingus stulpelius, kurie bus naudojami sukuriant užklausą, pereiname į „Calculations“ skiltį pavaizduota paveiksle „Paveikslas 35. Skaičiavimo laukų kūrimas“.

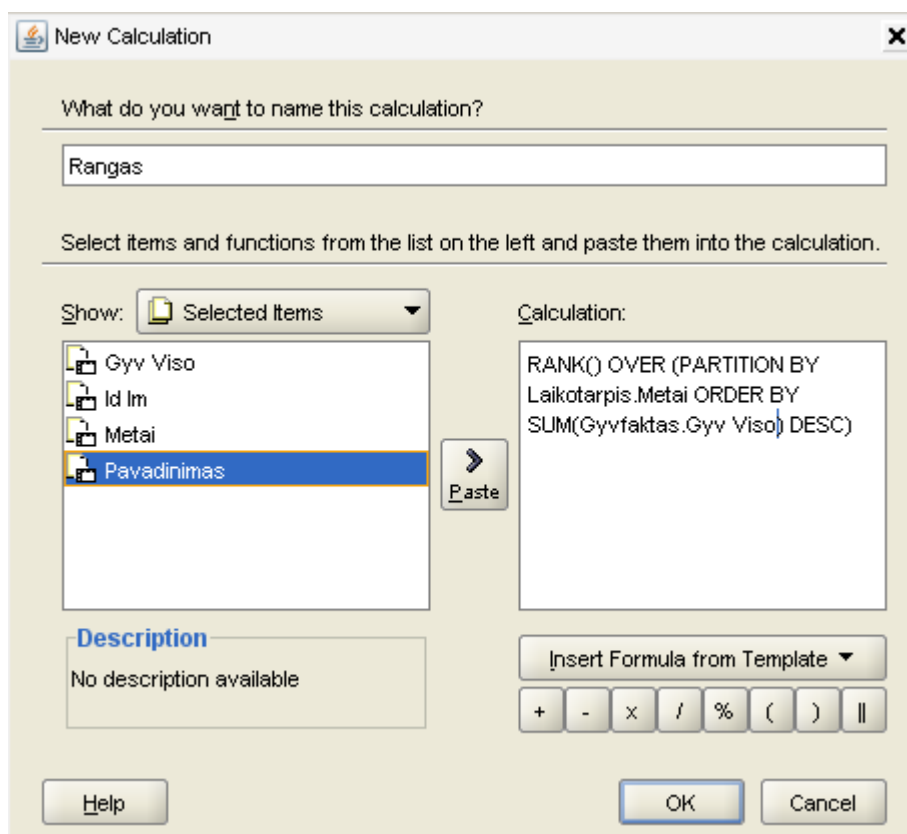


Paveikslas 35. Skaičiavimo laukų kūrimas

Pasirenkame „New->New Calculations“ ir taip kuriame skaičiavimo parametą, kuris bus naudojamas SQL užklausoje. Tarkime turime tokią SQL užklausa:

```
SELECT * FROM (  
    SELECT  
        RANK() OVER (PARTITION BY l.metai ORDER BY SUM(g.GYV_VISO) DESC) AS  
rangas,  
        RANK() OVER (ORDER BY SUM(g.GYV_VISO) DESC) AS metinis_rangas  
    FROM gyvfaktas g, laikotarpis l, imones m where g.id_data = l.id and g.id_im = m.id and  
g.id_rus = 1  
    GROUP BY l.metai, m.pavadinimas  
)  
WHERE rangas <= 3  
ORDER BY metai, rangas asc;
```

Paryškintos užklauskos dalys yra apskaičiuojami parametrai, todėl jie yra talpinami kaip apskaičiuojamos reikšmės Oracle Discoverer aplinkoje. Sukuriami apskaičiuojami parametrai pavaizduoti paveiksle „Paveikslas 36. Apskaičiuojamų parametų kūrimas“.



Paveikslas 36. Apskaičiuojamų parametų kūrimas

Taip yra sukuriami visi apskaičiuojami laukai. „Show“ skiltyje galima pasirinkti ne tik naudojamus parametrus, bet ir naudojamas funkcijas parametrą apskaičiuoti. Sukūrus parametą spaudžiame „OK“ tam, kad grįžtume atgal į užklauskos formavimo langą.

Užklauso „Where“ dalį galime nustatyti „Conditions“ skiltyje, tačiau ataskaitoje ši dalis yra pasirenkama dažniausiai vartotojo, todėl spaudžiame „Next“, kad galėtume nustatyti užklauso parametrus, kuriuos galės modifikuoti pats vartotojas. Taip užklausa taps kintanti pagal vartotojo poreikius. Parametro kūrimo langas pavaizduotas paveiksle „Paveikslas 37. Parametrų kūrimas“.

The screenshot shows the 'New Parameter' dialog box with the following configuration:

- Name: Ruisis
- Base item: Id Rus
- Operator: =
- Condition: (Id Rus = :Rusis)
- Prompt: Kokia pasirenkama imoku rusis?
- Description: (empty)
- Worksheet settings: Allow only one set of parameter values for all worksheets.
- Properties: Require users to enter a value (checked), Enable users to select multiple values (checked), Enable users to select either indexes or values (unchecked).
- Default value: (empty)
- Filtering: Show all available values (selected).

Paveikslas 37. Parametrų kūrimas

Taikomame SQL pavyzdyje mes turime „Where“ dalį:

```
WHERE g.id_data = l.id and g.id_im = m.id and g.id_rus = 1
```

Šiuo atveju vartotojui laisvai pasirenkamas bus „g.id_rus“ parametras, kurį vartotojas galės laisvai keisti, taip gaunant norimą rezultatą. Sukūrus parametrus, kuriuos gali naudoti ir parinkti vartotojas, užklausa bus baigta kurti.

Sukūrus užklausa Oracle Discoverer aplinkoje, užklausa bus sugeneruota ir pradėta vykdyti. Įvedus parametrus „Paveikslas 38. Parametrų įvedimo langas“ (psl. 51) ir paleidus vykdyti užklausa, bus gražinamas rezultatų langas pavaizduotas paveiksle „Paveikslas 39. Oracle Discoverer rezultatų analizės langas“ (psl. 51).

Edit Parameter Values

Select values for the following parameters:

Imones Pavadinimas*:

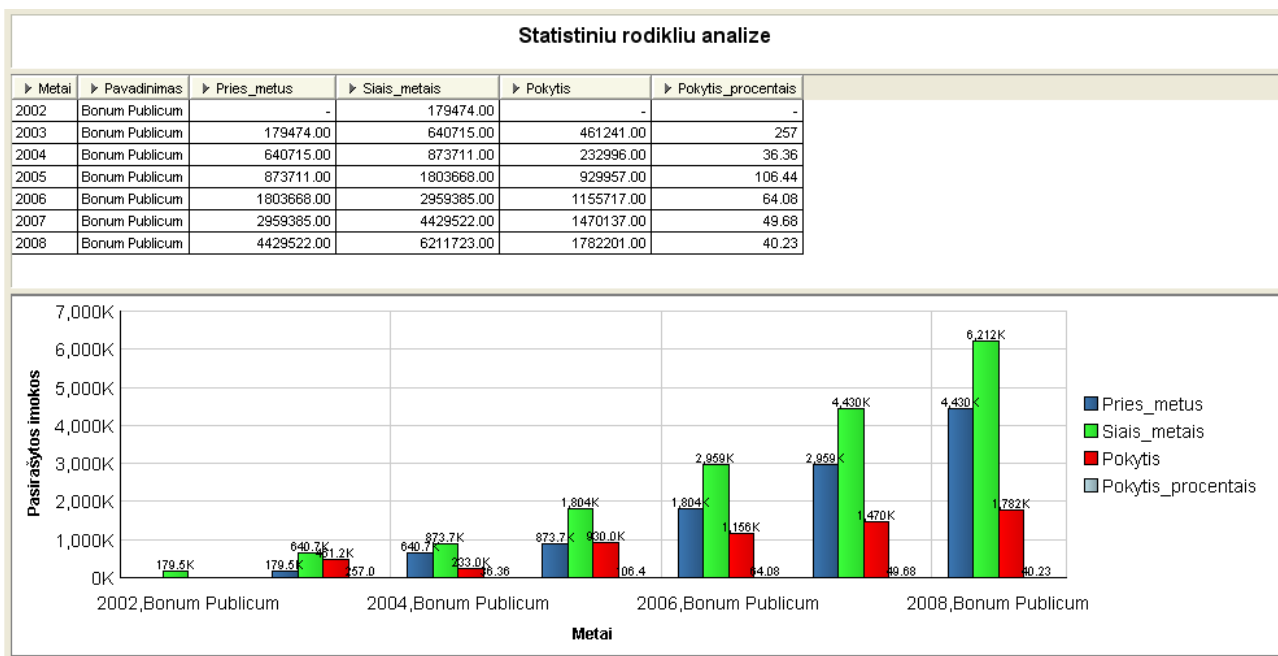
Pasirinkite rusi*:

Description

Use % as a wildcard

* indicates required field.

Paveikslas 38. Parametų įvedimo langas



4.2. Testavimo modelis ir duomenys, kontrolinis pavyzdys

Eksperimentinė duomenų bazė sudaryta iš 5 lentelių, kuriose yra apie 1000000 įrašų. Oracle DBVS funkcijos naudojamos SQL kalbos užklausoje ir procedūrose.

Užklausoje rezultatas bus gražinamas Oracle Discoverer įrankiui, kuris vartotojui pateiks grafinę užklausoje vaizdinį.

Chi kvadrato analitikai naudojamos SQL užklausoje sudarymas pavaizduotas veiklos diagramos paveiksle „Paveikslas 40. Chi kvadrato analitikos užklausoje sudarymo diagrama“ (psl. 52).



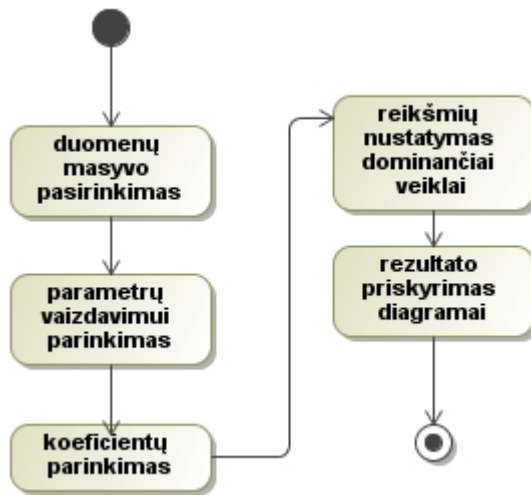
Paveikslas 40. Chi kvadrato analitikos užklauso sudarymo diagrama

Statistinės koreliacijos SQL užklauso sudarymo diagrama pavaizduota paveiksle „Paveikslas 41. Statistinės koreliacijos užklauso sudarymo diagrama“.



Paveikslas 41. Statistinės koreliacijos užklauso sudarymo diagrama

Linijinės regresijos SQL užklauso sudarymo diagrama pavaizduota paveiksle „Paveikslas 42. Linijinės regresijos užklauso sudarymo diagrama“ (psl. 53).



Paveikslas 42. Linijinės regresijos užklausos sudarymo diagrama

Pareto analizės SQL užklausos sudarymas pavaizduotas paveiksle „Paveikslas 43. Pareto analizės užklausos sudarymo diagrama“.



Paveikslas 43. Pareto analizės užklausos sudarymo diagrama

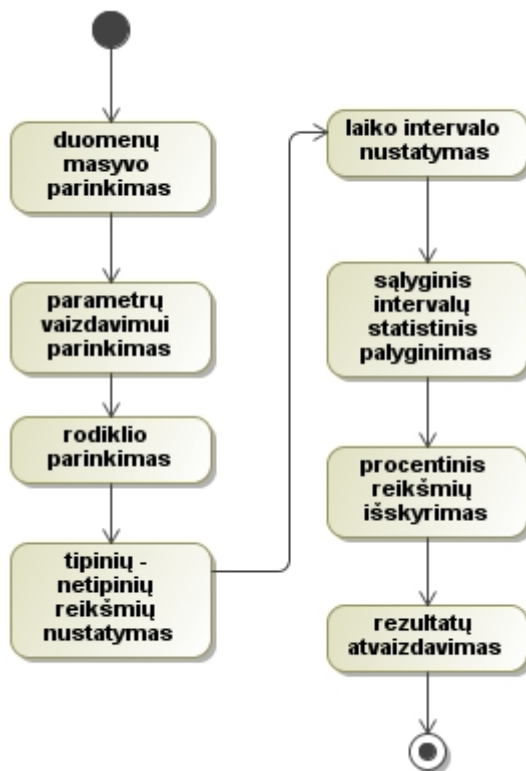
5. Eksperimentinis tyrimas

Siekiant nustatyti statistinės analizės galimybes, panaudojant DBVS priemones, bus sprendžiami statistiniai draudimo srities uždaviniai, panaudojant realius draudimo srities duomenis ir Oracle Discoverer įrankį. Eksperimento metu bus patikrinamas sudarytų uždavinių prieinamumas kitiems vartotojams tuo pačiu metu, analizuojamų duomenų saugumas ir analizės atlikimas laiko ir resursų atžvilgiu. Taip pat eksperimento metu bus tarpusavyje palygintos galimybės atlikti analizę DBVS priemonėmis panaudojant Oracle Discoverer ir Microsoft Excel įrankį.

5.1. Eksperimento uždavinių sprendimas

1. Kiek pasirašyta įmokų šiais metais, kiek buvo pasirašyta praeitais metais. Augimo ataskaita pagal pasirinktą įmonę, pateikiant procentinį pasirašytų įmokų pokytį lyginant su praėjusiais metais.

Šiam uždaviniui spręsti panaudojama statistinių rodiklių analizės SQL užklauso sudarymo diagrama, pavaizduota paveiksle „Paveikslas 44. Statistinių rodiklių analizės užklauso sudarymas“ (psl. 54).



Paveikslas 44. Statistinių rodiklių analizės užklauso sudarymas

Sudaryta užklausa pasirinktai užduočiai spręsti:

```

SELECT l.metai , substr(m.pavadinimas,0,20) imone,
       LAG(sum(g.gyv_viso)) OVER (PARTITION BY m.pavadinimas ORDER BY l.metai)
pries_metus,
       SUM(g.gyv_viso) siais_metais,
       SUM(g.gyv_viso)-(LAG(sum(g.gyv_viso)) OVER (PARTITION BY m.pavadinimas
ORDER BY l.metai)) pokytis,
       TO_CHAR(
       (SUM(g.gyv_viso)-(LAG(sum(g.gyv_viso)) OVER (PARTITION BY m.pavadinimas
ORDER BY l.metai)))/
       (LAG(sum(g.gyv_viso)) OVER (PARTITION BY m.pavadinimas ORDER BY
l.metai))*100,'999.99') pokytis_procentais
FROM gyvfaktas g, laikotarpis l, imones m
WHERE g.id_data = l.id and g.id_im = m.id and g.id_rus = 1 and m.pavadinimas like 'Bon%'
GROUP BY l.metai, m.pavadinimas
ORDER BY m.pavadinimas, l.metai;

```

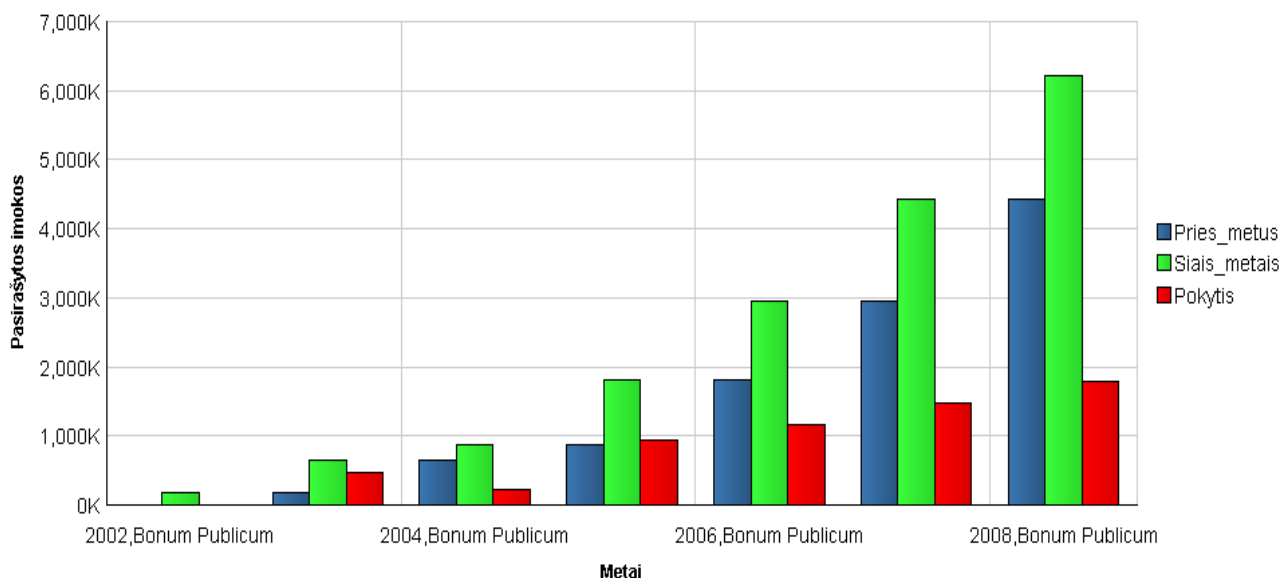
Šiai užklausei įvykdyti ir gauti rezultatą naudojamas Oracle Discoverer įrankis. Užklauso rezultatas pateikiamas Oracle Discoverer lentele paveiksle „Paveikslas 45. Augimo ataskaitos pagal pasirinkta įmone sąrašas, pateiktas lentele Oracle Discoverer įrankio aplinkoje“.

▶ Metai	▶ Pavadinimas	▶ Pries_metus	▶ Siais_metais	▶ Pokytis	▶ Pokytis_procentais
2002	Bonum Publicum	-	179474.00	-	-
2003	Bonum Publicum	179474.00	640715.00	461241.00	257
2004	Bonum Publicum	640715.00	873711.00	232996.00	36.36
2005	Bonum Publicum	873711.00	1803668.00	929957.00	106.44
2006	Bonum Publicum	1803668.00	2959385.00	1155717.00	64.08
2007	Bonum Publicum	2959385.00	4429522.00	1470137.00	49.68
2008	Bonum Publicum	4429522.00	6211723.00	1782201.00	40.23

Paveikslas 45. Augimo ataskaitos pagal pasirinkta įmone sąrašas, pateiktas lentele Oracle Discoverer įrankio aplinkoje

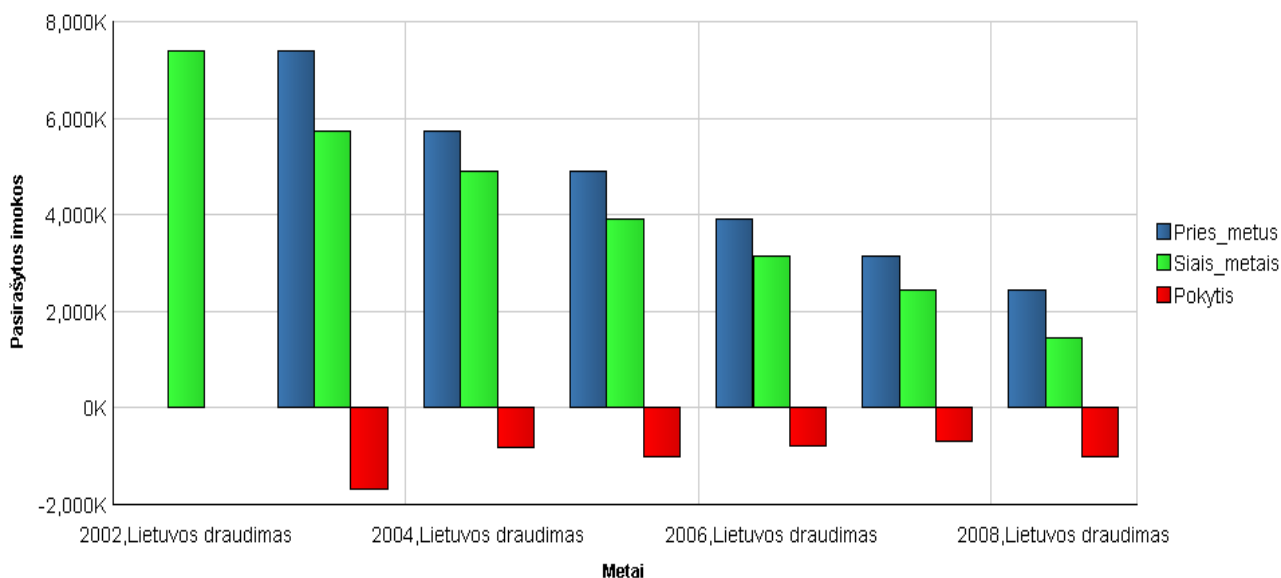
Iš pateiktos lentelės duomenų matome, jog prieš metus rodiklis rodo praėjusių metų pasirašytas įmokas konkrečiai įmonei, rodiklis šiais metais rodo einamų metų pasirašytas įmokas, pokyčio rodiklis rodo pokytį tarp einamų ir praėjusių metų pasirašytų įmokomų skaičių.

Uždavinio grafinis vaizdavimas pateiktas diagrama paveiksle „Paveikslas 46. Augimo ataskaitos pagal pasirinktą įmonę sąrašas, pateiktas diagrama Oracle Discoverer įrankio aplinkoje“.



Paveikslas 46. Augimo ataskaitos pagal pasirinktą įmonę sąrašas, pateiktas diagrama Oracle Discoverer įrankio aplinkoje

Diagramoje galime išvelgti, jog pasirinktos draudimo įmonės „Bonum Publicum“ pasirašytos įmokos kiekvienais metais augo ir nei karto nebuvo neigiamo pokyčio, kas simbolizuotų veiklos neigiamą rodiklį. Šiuo atveju pokytis kiekvienais metais akivaizdus, tačiau matomas nuo 2005 metų mažėjantis šių metų ir praėjusių metų rodiklio skirtumo pokytis, nors pasirašoma įmokų vis daugiau, leidžia daryti prielaidą jog šią sritį įtakoja konkurencija.



Paveikslas 47. "Lietuvos draudimas" įmonės augimo ataskaita pateikta diagrama

Išanalizavus „Lietuvos draudimas“ įmonės duomenis paveiksle „Paveikslas 47. "Lietuvos draudimas" įmonės augimo ataskaita pateikta diagrama“, galime matyti tendencingą pasirašytų įmokų mažėjimą, kuris turi didelę įtaką pajamoms, todėl šis uždavinys gali būti labai svarbus pavyzdys vaizduojant tam tikros srities pokyčius, priklausomai nuo pasirašytų įmokų ar sutarčių.

2. Rasti trejetuką įmonių, kurioms pasirašyta daugiausiai įmokų kasmet, nustatyti kokią įtaką pasirašytoms įmokoms daro gyvybės draudimas. Ekstremumo paieška pagal konkretaus rodiklio suminę reikšmę.

Šiam uždaviniui spręsti galima panaudoti nuostolingai dirbančių bendrovių užklauso sudarymo diagramą, pavaizduota paveiksle „Paveikslas 48. Nuostolingiausiai dirbančių bendrovių užklauso sudarymo diagrama“(psl. 58).



Paveikslas 48. Nuostolingiausiai dirbančių bendrovių užklauso sudarymo diagrama

Sudaroma užklausa pasirinktai užduočiai spręsti:

```

SELECT * FROM (
  SELECT l.metai as metai, m.pavadinimas,
  round(SUM(g.GYV_VISO),2) bendras_kieki,
  round(SUM(g.GYV_DRAUD),2),
  RANK() OVER (PARTITION BY l.metai ORDER BY SUM(g.GYV_VISO) DESC) AS
rangas,
  ROUND(100*(RATIO_TO_REPORT(SUM(g.GYV_VISO)) OVER ()),2) AS
met_proc_suma,
  RANK() OVER (ORDER BY SUM(g.GYV_VISO) DESC) AS met_sumos_rangas
FROM gyvfaktas g, laikotarpis l, imones m where g.id_data = l.id and g.id_im = m.id and
g.id_rus = 1
GROUP BY l.metai, m.pavadinimas
)
WHERE rangas <= 3
ORDER BY metai, rangas asc;
  
```

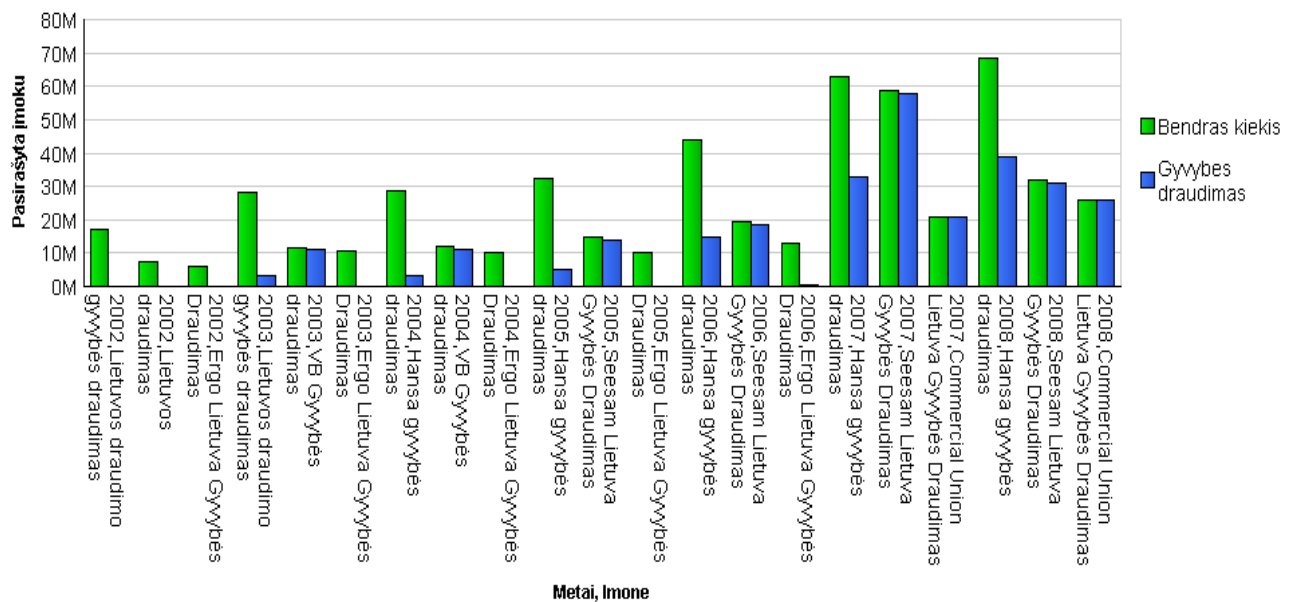
Šią užklausą praktiškai pritaikyti galima taip pat panaudojant Oracle Discoverer įrankį. Suvedus užklausą ir įvykdžius Oracle Discoverer įrankyje, gaunamas rezultatas, pavaizduotas lentelė paveiksle „Paveikslas 49. Trijų daugiausiai įmokų pasirašiusių įmonių sąrašas pateiktas lentelė Oracle Discoverer įrankio aplinkoje“ (psl. 59).

► Metai	► Pavadinimas	► rangas	► Bendras kiekis(mln.)	► Metinė procentinė suma	► Metinės sumos rangas	► Gyvybės draudimas
2002	Lietuvos draudimo gyvybės draudimas	1.00	17321844.00	2.52	12.00	20703.00
2002	Lietuvos draudimas	2.00	7407870.00	1.08	25.00	0.00
2002	Ergo Lietuva Gyvybės Draudimas	3.00	6122351.00	0.89	28.00	0.00
2003	Lietuvos draudimo gyvybės draudimas	1.00	28263422.00	4.11	8.00	3142472.00
2003	VB Gyvybės draudimas	2.00	11596624.00	1.69	19.00	10976794.00
2003	Ergo Lietuva Gyvybės Draudimas	3.00	10743170.00	1.56	20.00	0.00
2004	Hansa gyvybės draudimas	1.00	28896626.00	4.21	7.00	3392316.00
2004	VB Gyvybės draudimas	2.00	11832023.00	1.72	18.00	11226135.00
2004	Ergo Lietuva Gyvybės Draudimas	3.00	10112905.00	1.47	22.00	0.00
2005	Hansa gyvybės draudimas	1.00	32440155.00	4.72	5.00	4972976.00
2005	Seesam Lietuva Gyvybės Draudimas	2.00	14637113.00	2.13	15.00	13851741.00
2005	Ergo Lietuva Gyvybės Draudimas	3.00	10253912.00	1.49	21.00	167665.00
2006	Hansa gyvybės draudimas	1.00	44026395.00	6.41	4.00	14748578.00
2006	Seesam Lietuva Gyvybės Draudimas	2.00	19344272.00	2.82	11.00	18532618.00
2006	Ergo Lietuva Gyvybės Draudimas	3.00	13165849.00	1.92	16.00	569157.00
2007	Hansa gyvybės draudimas	1.00	63058810.00	9.18	2.00	32830389.00
2007	Seesam Lietuva Gyvybės Draudimas	2.00	58719457.00	8.55	3.00	57936969.00
2007	Commercial Union Lietuva Gyvybės Dra	3.00	20745228.00	3.02	10.00	20745228.00
2008	Hansa gyvybės draudimas	1.00	68266494.00	9.94	1.00	38682387.00
2008	Seesam Lietuva Gyvybės Draudimas	2.00	31858248.00	4.64	6.00	31069321.00
2008	Commercial Union Lietuva Gyvybės Dra	3.00	25850822.00	3.76	9.00	25850822.00

Paveikslas 49. Trijų daugiausių įmokų pasirašiusių įmonių sąrašas pateiktas lentelė Oracle Discoverer įrankio aplinkoje

Iš pateiktos lentelės matome, kad rangas parodo vietą, kurioje tais metais yra įmonė. Bendras kiekis nurodo tais metais pasirašytą kiekį įmokų. Metinė procentinė suma nurodo metinį procentą pagal pasirašytų įmokų kiekį. Metinės sumos rangas parodo metinį rangą pagal pasirašytų įmokų sumą. Gyvybės draudimas parodo konkretų kiekį pasirašytų įmokų būtent gyvybės draudimo srityje tais metais.

Šio uždavinio grafinis atvaizdavimas pateikiamas grafiku, pavaizduotu paveiksle „Paveikslas 50. Trijų daugiausių įmokų pasirašiusių įmonių sąrašas pateiktas diagrama Oracle Discoverer įrankio aplinkoje“ (psl. 60).



Paveikslas 50. Trijų daugiausiai įmokų pasirašiusių įmonių sąrašas pateiktas diagrama Oracle Discoverer įrankio aplinkoje

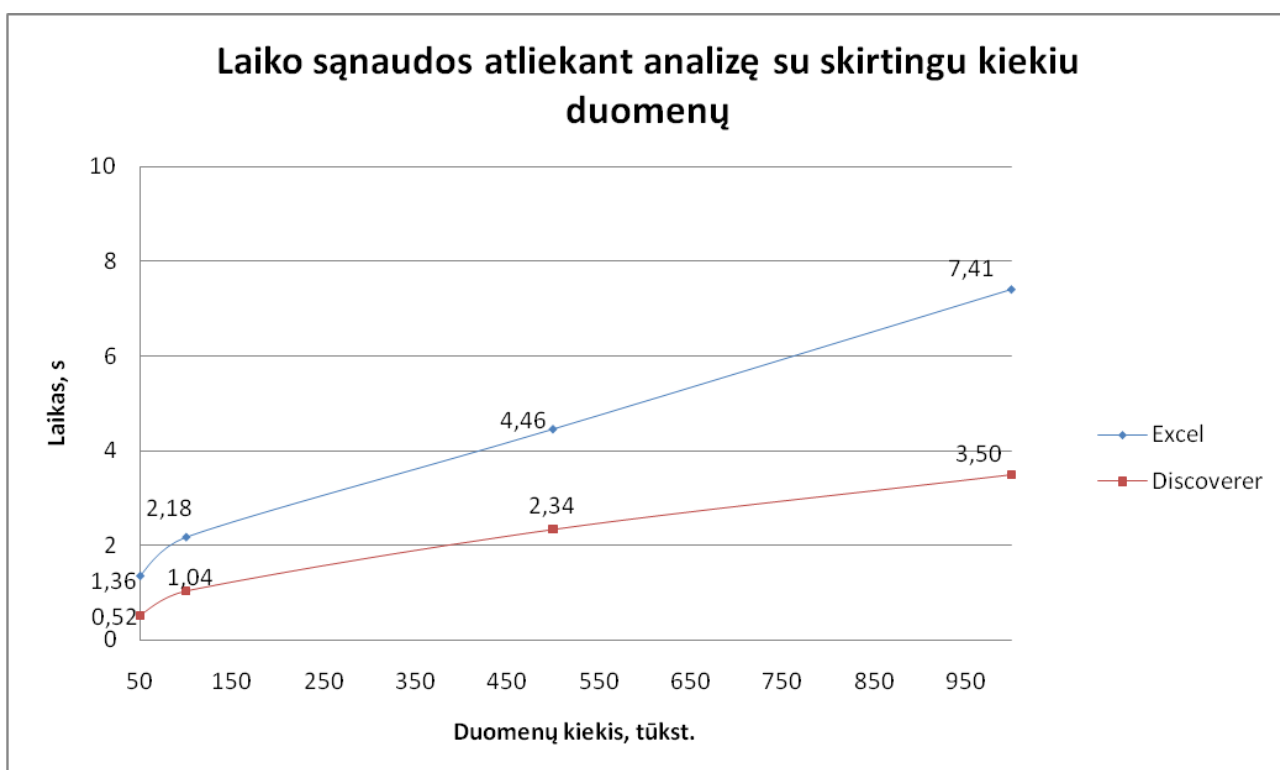
Kaip matome diagramoje, pagal rangą vaizduojama po 3 geriausias įmones, pasirašančias daugiausiai draudimo įmokų kasmet. Remiantis 2002 metų duomenimis, bendras pasirašytų draudimo įmokų kiekis yra apie 17 milijonų ir nuo 2002 metų kasmet auga po 10 milijonų pasirašytų įmokų, tai rodo tendenciją draudimo poreikiui. Gyvybės draudimas turėjęs labai mažą poreikį 2002 metais, pamažu ėgavo poreikį draudimo srityje ir jau 2008 metais sudarė pusę, o kai kuriais atvejais ir 90% bendro pasirašytų įmokų kiekio draudimo įmonėse.

5.2. Analizės atlikimo palyginimas naudojant DBVS priemones su Oracle Discoverer ir Microsoft Excel 2007

Eksperimentui atlikti bus naudojamos aprašytos užklausos skyriuje „Eksperimento uždavinių sprendimas“ (psl. 54). Microsoft Excel gali analizuoti duomenis naudojant Microsoft Query arba integruotomis funkcijomis, kurios naudojamos kai reikalingi duomenys yra parsiončiami į Excel iš DBVS sistemos. Analizė integruotomis funkcijomis yra neefektyvi, kadangi analizuojamų duomenų kiekis gali būti ir keletas milijonų įrašų, o kaip buvo minėta analizės dalyje, Microsoft Excel įrankis dirbdamas su dideliais kiekiais duomenų veikia neefektyviai, be to saugoti tokį didelį kiekį duomenų fiziniame atmintyje Microsoft Excel nesugeba. Naudojant Microsoft Query, Excel įrankis naudodamas ODBC arba JDBC tvarkykles, prisijungia prie duomenų bazės ir pateikia užklausa, kurią apdoroja DBVS sistema ir gražina rezultatą į Excel įrankį. Eksperimento metu, naudojant DBVS priemones tiek Oracle Discoverer tiek Microsoft Excel įrankiu, bus lyginamos resursų ir laiko sąnaudos

analizuojant skirtingą kiekį duomenų, analizės atlikimas keliems vartotojams tuo pačiu metu ir duomenų saugumo užtikrinimo galimybės.

Atliekant laiko sąnaudų analizę su skirtingu kiekiu duomenų, buvo naudojama užklausa ekstremumo paieškai atlikti (psl. 59). Šios užklauso rezultatas buvo nustatomas sekundėmis, per kiek laiko atliekama pasirinktos užklauso duomenų statistinė analizė. Užklausa pakartojama 20 kartų su tuo pačiu duomenų kiekiu ir nustatomas laiko vidurkis reikalingas analizei atlikti. Laiko sąnaudos, analizuojant skirtingą kiekį duomenų, pavaizduotos paveiksle „Paveikslas 51. Laiko sąnaudos atliekant analizę su skirtingu kiekiu duomenų panaudojant Oracle Discoverer ir Microsoft Excel 2007 įrankius“.

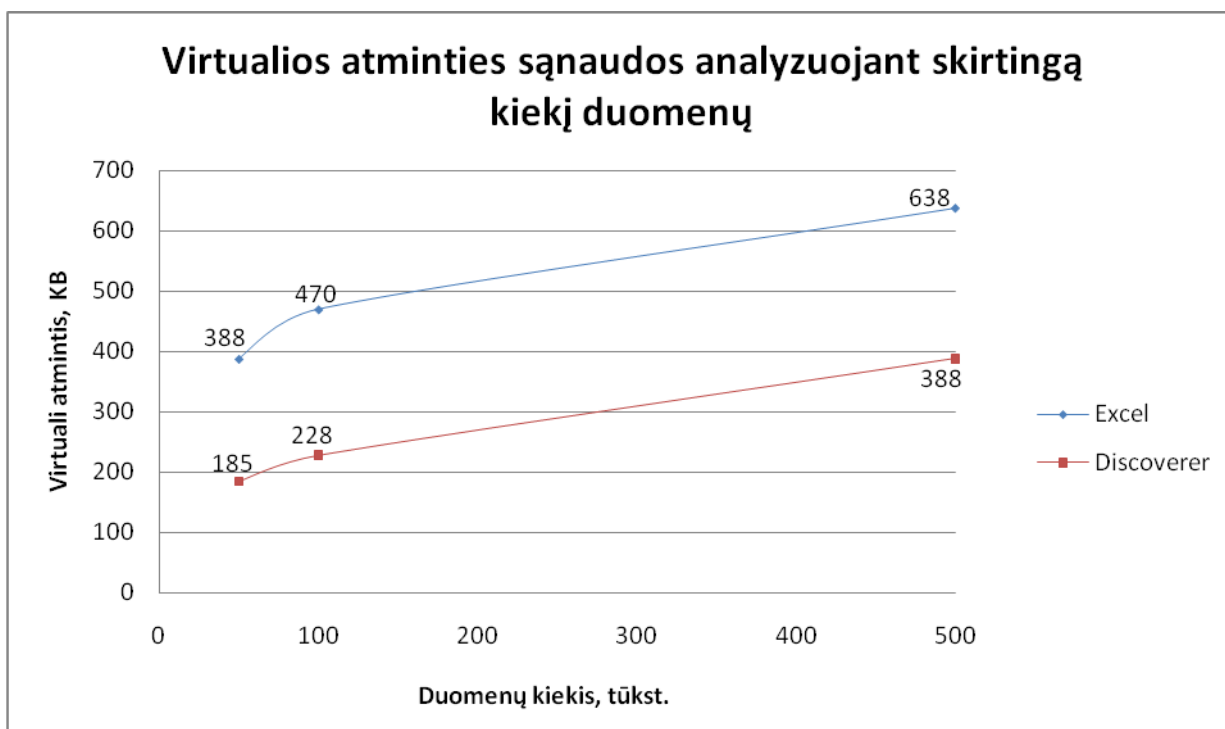


Paveikslas 51. Laiko sąnaudos atliekant analizę su skirtingu kiekiu duomenų panaudojant Oracle Discoverer ir Microsoft Excel 2007 įrankius

Iš grafiko matyti, kad buvo parinkti tokie duomenų kiekiai: 50000, 100000, 500000 ir 1000000. Analizuojant duomenis su 50000 įrašų kiekiu, rezultatų pateikimo laikas nėra žymus, tačiau didėjant duomenų kiekiui, didėja ir laiko skirtumas analizuojant skirtingais įrankiais. Kadangi Microsoft Excel atliekant analizę, užklausa yra siunčiama į Oracle DBVS, tuomet DBVS apdoroja šią užklausą ir rezultatą vėl siunčia į Microsoft Excel, todėl didėjant duomenų kiekiui, laiko sąnaudos taip pat didėja duomenų apsikeitimams tarp DBVS ir Microsoft Excel. Gavus užklauso rezultata ir norint pakartoti duomenų analizę, užklausa vėl perduodama apdoroti į DBVS ir vėl laukiamas rezultato gavimas, tuo atveju Oracle Discoverer įrankyje atlikta užklausa yra saugoma virtualioje atmintyje, todėl kartojant

užklausa duomenys yra imami iš virtualios atminties ir jokie rezultatų skaičiavimai nėra vykdomi, todėl nėra apkraunama duomenų bazė. Oracle Discoverer įrankis užklausa siunčia į Oracle DBVS ir rezultatas gražinamas į Oracle Discoverer įrankį, tačiau per trumpesnę laiką nei tai buvo daroma su Microsoft Excel įrankiu. Praktikoje analizuojamų duomenų kiekiai dažniausiai viršija 1000000, todėl analizė tampa neefektyvi naudojant Microsoft Excel įrankį.

Atliekant resursų sąnaudų analizę pasirinktas duomenų kiekis užklauskos pagalba yra paimamas iš Oracle DBVS ir parsiuočiomas į Excel ir Discoverer įrankius, ir nustatoma kiek sunaudojama sistemos resursų, kuomet šie duomenys yra paimami. Užklauskos atlikimo metu bus nustatomas užimamas atminties kiekis tiek fizinėje tiek virtualioje atmintyje. Kadangi Oracle Discoverer duomenis analizuoja ir saugo virtualioje atmintyje, todėl fizinėje atmintyje saugomas duomenų kiekis bus tiriamas tik su Microsoft Excel įrankiu. Užklausa kartojama 10 kartų su tuo pačiu duomenų kiekiu ir nustatomas virtualios ir fizinės atminties sąnaudos vidurkis. Virtualios atminties sąnaudos pavaizduotos paveiksle „Paveikslas 52. Virtualios atminties sąnaudos analizuojant skirtingą kiekį duomenų panaudojant Oracle Discoverer ir Microsoft Excel 2007 įrankius“.

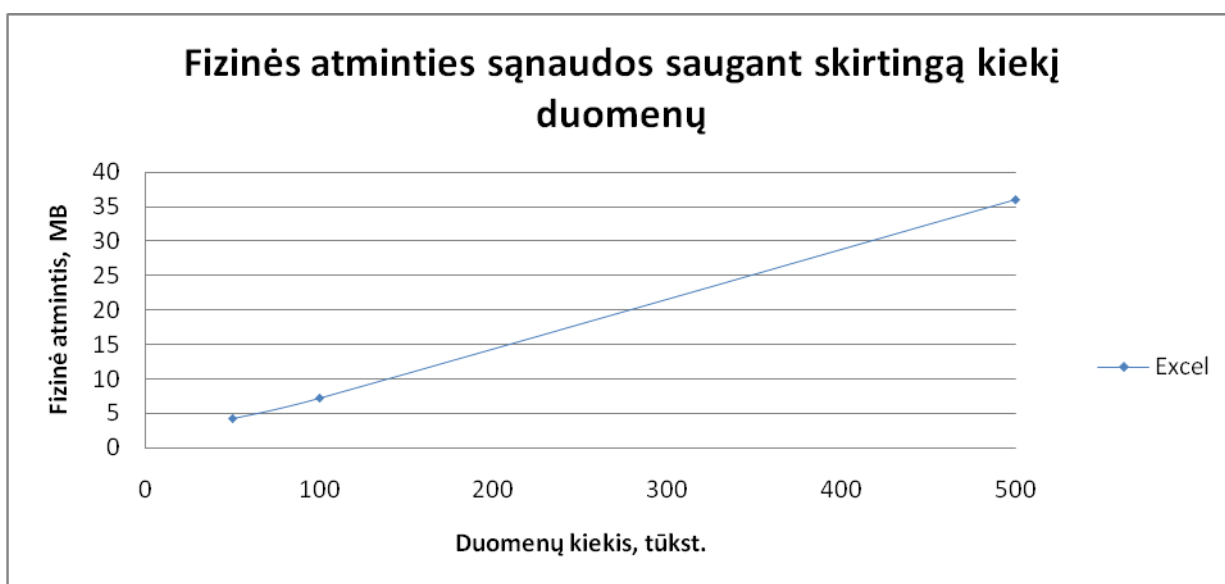


Paveikslas 52. Virtualios atminties sąnaudos analizuojant skirtingą kiekį duomenų panaudojant Oracle Discoverer ir Microsoft Excel 2007 įrankius

Virtualios atminties sąnaudų grafike matomas didelis skirtumas tarp Microsoft Excel ir Oracle Discoverer įrankių. Atliekant eksperimentą 10 kartų su tuo pačiu duomenų kiekiu buvo nustatomas virtualios atminties sunaudojimo vidurkis, kuris Microsoft Excel įrankio atžvilgiu buvo nuo 388KB iki 638KB, kadangi kiekvieną kartą kartojant virtualios atminties kiekis

neišsivalydavo ir vis didėdavo, nes duomenys vėl pakartotinai buvo imami iš DBVS ir gražinami į Excel darbalaukį. Tuo tarpu atliekant analizę su Discoverer įrankiu analizuojant tą patį duomenų kiekį, rezultatas buvo išsaugomas virtualioje atmintyje ir kartojant eksperimentą, tie patys duomenys buvo imami iš virtualios atminties.

Fizinės atminties sąnaudų paveiksle „Paveikslas 53. Fizinės atminties sąnaudos saugant analizei skirtingą kiekį duomenų Microsoft Excel 2007 įrankyje“ matomas įrankio Microsoft Excel bylos užimamas dydis išsaugojus nurodytą kiekį duomenų. Oracle Discoverer įrankis analizuojamus duomenis kaupia virtualioje atmintyje ir duomenų fiziškai nesaugo, kadangi duomenys yra saugomi duomenų bazėje.



Paveikslas 53. Fizinės atminties sąnaudos saugant analizei skirtingą kiekį duomenų Microsoft Excel 2007 įrankyje

Atliekant eksperimentą virtualioje ir fizinėje atmintyje su Microsoft Excel įrankiu buvo pastebėtas gana didelis sistemos resursų naudojimas, palyginus su Discoverer įrankiu. Kartojant eksperimentą keletą kartų, virtualios atminties resursai neišsivalydavo ir didėdavo, taip tolesnį analizės darbą vis apsunkinant. Dirbant su 500000 kiekiu duomenų Discoverer aplinkoje, analizė vyko greitai, neišėikvojant daug virtualios atminties ir neužimant vietos fizinėje atmintyje.

Eksperimento metu buvo nustatomas duomenų prieinamumas dirbant tiek su Microsoft Excel tiek su Oracle Discoverer įrankiu. Vienu metu darbą pradeda trys analitikai, kurie bando paimti iš DBVS duomenis ir juos analizuoti. Dirbant su Microsoft Excel, atsidarius pirmam analitikui byla, kurioje yra duomenų paėmimo šablonas, kiti du analitikai šios bylos negali redaguoti, taip suvaržomos analitikų teisės atlikti analizę vienu metu. Praktikoje analizė atliekama įvairių sričių pjūviais (pardavimai, gamyba, finansinės operacijos), todėl kiekvienas analitikas turi savo sritį kurioje analizuoja duomenis. Todėl Excel įrankio atveju tik vienas

vartotojas gali prieiti prie šablono ir su juo dirbti, o norint ir kitiems redaguoti Excel failę esančią užklausa, reikia susikurti naują analogišką Excel bylos šabloną. Tuo tarpu Oracle Discoverer prisijungimas vyksta naršyklės pagalba ir yra prisijungiama prie duomenų bazės su kuria vienu metu gali dirbti iki 100 organizacijos darbuotojų. Eksperimento metu visi trys analitikai prisijungę prie Discoverer aplinkos galėjo prieiti prie duomenų ir juos analizuoti pagal numatytus poreikius. Tokiu principu yra sumažinamas analizės atlikimo laikas, kadangi analizė yra atliekama keleto darbuotojų ir duomenys gali būti analizuojami įvairiais pjūviais tuo pačiu metu.

Eksperimento metu buvo nustatomas duomenų saugumas dirbant su Microsoft Excel ir Oracle Discoverer įrankiu. Saugumas yra labai svarbus faktorius analizuojant svarbius įmonės duomenis, kadangi iš duomenų galima nustatyti įmonės veiklos rodiklius, kurie atsidūrę konkurentų rankose gali padaryti labai didelę žalą tolimesnei įmonės veiklai. Eksperimento metu bus nustatomas duomenų saugumas analizės metu.

Duomenų priėjimas naudojant Oracle Discoverer įrankį vyksta prisijungiant prie duomenų bazės, duomenų bazės vartotojo vardu. Prisijungus ir atliekant analizę duomenys yra analizuojami prisijungusio asmens vardu ir niekur papildomai nėra saugomi.

Analizuojant duomenis Microsoft Excel įrankiu prisijungimas atliekamas taip pat prie duomenų bazės, tačiau rezultatas užklauso yra gaunamas atsiunčiant į Microsoft Excel darbalaukį. Išsaugojus Excel bylą, parsųsti analizės duomenys yra išsaugomi ir išsaugota byla gali būti pernešta arba persųsta į kitą vietą ir taip pažeidžiamas svarbių duomenų saugumas. Eksperimento metu atlikus duomenų parsųsiuntimą iš duomenų bazės ir rezultatų išsaugojimą Excel byloje, duomenys taip pat buvo išsaugomi. Šią bylą nusiuntus kitiems vartotojams, kurie turi įsidiegti Microsoft Excel paketą, byla su testiniais duomenimis buvo atidaryti.

Eksperimentų metu atlikti bandymai parodo, kad duomenų saugumas naudojant Microsoft Excel nėra užtikrinamas, kadangi duomenys gali būti išsaugomi Excel byloje ir pernešti arba persųsti trečiosioms šalims. Oracle Discoverer įrankių analizės metu yra prisijungiama prie duomenų bazės ir duomenų analizės rezultatas yra pateikiamas užklauso pagalba ir niekur kitur nėra saugomas.

5.3. Taikymo rekomendacijos

Pateikta statistinės analizės metodika gali būti taikoma ne vien su Oracle duomenų bazių valdymo sistema, bet ir su kitomis duomenų bazių valdymo sistemomis, kurios turi integruotas statistinės analizės funkcijas. Jei tokių funkcijų DBVS neturi, vartotojai gali susikurti patys reikalingas funkcijas ir jas naudoti analizės atlikimui arba naudoti grafinio įrankio integruotomis funkcijomis. Oracle Discoverer įrankis turi tas pačias integruotas statistines funkcijas, kurios yra integruotos ir duomenų bazėje. Microsoft SQL Server neturi tokio statistinių funkcijų rinkinio kaip Oracle, tačiau Microsoft SQL Server Reporting Services grafinis įrankis turi praplėstą statistinių funkcijų paketą, todėl pateiktą metodiką galima taikyti naudojant Microsoft SQL Server ir Microsoft SQL Server Reporting Services įrankius. Analizei atlikti rekomenduojama naudoti to paties prekinio ženklo grafinį įrankį ir DBVS.

6. Darbo rezultatų apibendrinimas

Atlikus Oracle Discoverer ir sudarytų užklausų su DBVS statistinėmis funkcijomis eksperimentinį tyrimą, reikalinga nustatyti pasirinktų kriterijų išpildymą. Sudaroma sprendimų palyginimo lentelė „Lentelė 8. Sprendimų palyginimas“, kurioje yra lyginami, SAS, Microsoft Excel ir Oracle Discoverer su DBVS integruotomis funkcijomis, įrankiai. Prie pateikto kriterijaus pateikiamas kiekvieno įrankio nustatyto kriterijaus atitikimo įvertinimas.

Lentelė 8. Sprendimų palyginimas

	SAS	Microsoft Excel	Oracle Discoverer ir DBVS integruotos funkcijos
Kaina	Didelė	Maža	Papildomai tai nieko nekainuoja
Sudėtingumas ir panaudojimas	Įrankis sudėtingas atlikti net nesudėtingą statistinę analizę. Norint efektyviai dirbti, reikia mokėti SAS kalbą. Sukurtos papildomos funkcijos veiks tik SAS įrankio aplinkoje.	Įrankiu lengva naudotis. Tačiau norint sukurti sudėtingesnes funkcijas, reikia mokėti naudoti Microsoft Visual Basic kalbą. Papildomos sukurtos funkcijos bus panaudojamos Microsoft Excel aplinkoje.	Naudoti DBVS funkcijas SQL kalboje nėra sudėtinga, reikia tik žinoti bendrus SQL kalbos standartus. Aprašytas užklausas ir funkcijas galima naudoti visose DBVS, kuriose yra naudojama SQL kalba.
Funkcijų pasirinkimas uždaviniams spręsti	Funkcijų kiekis didelis statistiniams uždaviniams spręsti.	Statistinės funkcijos yra standartinės.	Įmonių statistinei analizei integruotų funkcijų visiškai pakanka.
Darbas su dideliu kiekiu duomenų	Gali analizuoti terabaitinius duomenų kiekius.	Gali efektyviai dirbti su nedaugiau nei 1000000 eilučių duomenų.	Gali analizuoti terabaitinius duomenų kiekius.
Resursų sunaudojimas	Analizei paimti duomenys talpinami SAS laikinojoje srityje, taip užimdami vietą fizinėje atmintyje. Skaičiavimams atlikti naudojama operatyvi atmintis.	Paimti duomenys užima daug vietos fizinėje atmintyje, o vykdant ir operatyvioje atmintyje.	Duomenys analizuojami operatyvioje atmintyje.

Duomenų prieinamumas tuo pačiu metu keliems vartotojams	Tuo pačiu metu gali dirbti iki 20 vartotojų. Norint kad analizę atliktų daugiau nei 20 žmonių, reikia papildomai įrankį plėsti.	Analizuoti duomenis gali tik vienas analitikas. Tuo pačiu metu norintis analizuoti duomenis kitas analitikas redaguoti sukurtos bylos negali.	Tuo pačiu metu prie duomenų bazės gali prieiti iki 100 vartotojų.
Saugumas	Naudojimasis įrankiu yra saugus, kadangi analizuojami duomenys yra šifruojami, tačiau šifruoti duomenys saugomi laikinoje fiziniėje atmintyje.	Duomenys gali būti išsaugomi byloje ir pernešami arba persiunčiami.	Duomenys yra prieinami tik analizės metu ir niekur papildomai nėra saugomi, taip išvengiant svarbių duomenų pasisavinimo.
Duomenų analizavimas naudojant skirtingas duomenų saugyklas	Gali paimti duomenis tiek iš Oracle, DB2 ar Microsoft SQL Server duomenų saugyklos.	Duomenis analizei gali paimti iš Microsoft SQL Server, Oracle, DB2 duomenų saugyklos.	Duomenis analizei gali būti paimti iš Microsoft SQL Server, Oracle, DB2 duomenų saugyklų.

Pateiktoje lentelėje matome, kad pasiūlytas sprendimas duomenų analizei atlikti, panaudojus Oracle Discoverer ir DBVS integruotas funkcijas, atitinka pasirinktus kriterijus ir jis gali būti pranašesnis už esamus statistinei duomenų analizei skirtus sprendimus. Siūlomas sprendimas nereikalauja papildomų investicijų, analizės užklausų sudarymui pakanka turėti SQL kalbos žinių, o norimas analizės rezultatas gali būti gaunamas greičiau ir saugiau.

Išvados

1. Darbe sprendžiama mažoms ir vidutinėms įmonėms aktuali duomenų statistinės analizės problema. Esami duomenų analizės sprendimai netenkina mažų ir vidutinių įmonių poreikių ir lūkesčių, nes yra brangūs, sudėtingi arba per silpni, reikalaujantys papildomų techninių resursų ir žmogiškųjų išteklių.
2. Statistiniam duomenų analizavimui patobulinti pateiktas procesas su aprašyta darbo metodika, kuris leidžia statistiškai analizuoti esamus duomenis panaudojant duomenų bazių valdymo sistemų integruotas statistines funkcijas ir DBVS lydinčias grafines priemones arba trečiųjų šalių grafinius įrankius.
3. Pasiūlytas sprendimas buvo pritaikytas ir realizuotas draudimo įmonės duomenų statistinei analizei. Tačiau pasiūlytas procesas yra universalus, tinkantis įvairiose verslo srityse.
4. Duomenų statistinei analizei atlikti pasirinkta Oracle DBVS, kurią naudoja draudimo bendrovė. Oracle turi didelį kiekį integruotų statistinių funkcijų, todėl tai užtikrina įvairesnę analizę ir greitą rezultatų gavimą. Grafiniam vaizdavimui pasirinktas Oracle Discoverer įrankis, kuris optimaliai išnaudoja Oracle DBVS galimybes analizės metu.
5. Pateiktas sprendimo įgyvendinimas nereikalauja įmonės papildomų finansinių resursų, kadangi naudojami integruoti DBVS įrankiai, o įmonės personalo kvalifikacijos pakaks šio sprendimo pritaikymui ir naudojimui - pakanka suprasti bendrus SQL kalbos standartus ir analizuojamos srities ypatumus.
6. Pasiūlytas sprendimas gali būti taikomas su kitomis DBVS, kurios turi integruotas analitines funkcijas ir grafinius įrankius rezultatams atvaizduoti.

Santrumpų ir terminų žodynėlis

SQL – kalba, suteikianti teisę prieiti prie reliacinių duomenų bazių valdymo sistemų.

DBVS – duomenų bazių valdymo sistema.

ODBC – standartizuota sąsaja, leidžianti taikomosioms programoms prisijungti prie reliacinių duomenų bazių.

JDBC – Java kalbos sąsaja, leidžianti Java kalba parašytoms taikomosioms programoms prisijungti prie reliacinių duomenų bazių.

ETL - parsuntimas, transformavimas, pakrovimas. Terminas naudojamas, kai norima nusakyti duomenų migravimą parsunčiant juos iš vienos saugomos vietos, vėliau duomenis transformuojant ir pakraunant į kitą saugomą vietą, dažniausiai duomenų saugyklą.

Literatūra

1. David Loshin. Business Intelligence: The Savvy Manager's Guide (The Savvy Manager's Guides), 2003 m.
2. Gareth Herschel. Magic Quadrant for Customer Data Mining, 1Q06. Prieiga per internetą: <http://sameer1prakash.googlepages.com/CRM-DataMiningVendors2006.pdf>
3. Jeff Kelly. Rankings of customer data mining market show shift to user-friendly tools. Prieiga per internetą: http://searchdatamanagement.techtarget.com/news/article/0,289142,sid91_gci1321221_00.html
4. Colin Beast. Gartner Crowns SAS and SPSS Data Mining Kings. Prieiga per internetą: <http://www.destinationcrm.com/Articles/ReadArticle.aspx?ArticleID=42362>
5. Rick Sherman. Microsoft Excel: The king of BI. Prieiga per internetą: http://searchcrm.techtarget.com/news/column/0,294698,sid11_gci1081869_00.html
6. Gordon S. Linoff. Data Analysis Using SQL and Excel, 2007 m.
7. Eva Goldwater. Using Excel for Statistical Data Analysis – Caveats . Prieiga per internetą: <http://www-unix.oit.umass.edu/~evagold/excel.html>
8. Thomas Kyte. Expert One-on-One Oracle, 2001 m.
9. Carlos Ordonez . Building statistical models and scoring with UDF, Proceedings of the 2007 ACM SIGMOD international conference on Management of data[Beijing, China], 1005 – 1016 p.
10. Zhibo Chen; Carlos Ordonez .Efficient OLAP with UDF. Proceeding of the ACM 11th international workshop on Data warehousing and OLAP,Napa Valley, California, USA, 2008,p. 41 – 48.
11. Marcos M. Campos. Analytics in the Oracle Database. Prieiga per internetą: Internetinė nuoroda: http://oracledmt.blogspot.com/2006/01/analytics-in-oracle-database_05.html
12. Richard Earp, Sikha Baqui. Advanced SQL functions in Oracle 10g(Wordware Applications Library), 2006 m.
13. Gartner releases 2007 magic quadrant for data warehouse database management systems. Prieiga per internetą: <http://www.biblogs.com/2008/01/27/gartner-releases-2007-magic-quadrant-for-data-warehouse-database-management-systems/>
14. Liutauras Ulevičius. Oracle Lietuvoje: IT banginis įvertino rinką. Prieiga per internetą: <http://verslas.banga.lt/lt/leidinys.full/3cf4bf7ccb53b>

15. Paolo Giudici. Applied Data Mining: Statistical Methods for Business and Industry (Statistics in Practice), 2003m.
16. Michael Böhnlein, Achim Ulbrich-vom Ende. Business Process Oriented Development of Data Warehouse Structures, 2000 m.
17. Mingguang Xu; Denise Gardner. Oracle Analytic Functions for IR Analysis and Reporting. Prieiga per internetą:
<http://www.uga.edu/oir/reports/OracleAnalyticFunction-SAIR-2006.ppt>

**1. PRIEDAS. KONFERENCIJOJE PRISTATYTAS IR IŠSPAUSDINTAS
STRAIPSNIS**

VILNIAUS
UNIVERSITETAS

VYTAUTO DIDŽIOJO
UNIVERSITETAS

KAUNO TECHNOLOGIJOS
UNIVERSITETAS

14

-OSIOS

TARPUNIVERSITETINĖS

MAGISTRANTŲ *ir*

DOKTORANTŲ

MOKSLINĖS

KONFERENCIJOS

„INFORMACINĖS
TECHNOLOGIJOS“

pranešimų medžiaga

2009

14-osios tarpuniversitetinės magistrantų ir doktorantų mokslinės konferencijos „Informacinės technologijos“ pranešimų medžiaga 2009.05.08

REMIGIJUS KIMINASI.....	168
ELEKTRONINIŲ LAIŠKŲ PAPILDYMAS SIUNTĖJO VIETOS DUOMENIMIS	173
JULIUS KRIUKASI, DANGIS RIMKUS2	173
VEIKLOS TASYKLIŲ, GRINDŽIAMŲ ŠABLONAIŠ, VALDYMAS INFORMACINĖSE SISTEMOSE	177
AGNĖ RUČINSKAITĖ, LINA TUTKUTĖ	177
DUOMENŲ ANALIZAVIMAS NAUDOJANT STATISTINES DBVS PRIEMONES.....	182
ŽYGIMANTAS VASILIAUSKAS	182
TRANSFORMACIJA IŠ UML Į SBVR NAUDOJANT ATLAS KALBĄ ECLIPSE APLINKOJE.....	187
EDVINAS ŠINKEVIČIUS, RIMANTAS BUTLERIS	187
V SEKCIJA MOBILIOS TECHNOLOGIJOS	191
MOBILIŲJŲ SISTEMŲ VARTOTOJO SĄSAJOS TESTAVIMO PANAUDOJANT UML MODELIOUS KARKASO SUDARYMAS IR TYRIMAS.....	192
ROBERTAS JASAITIS	192
EFFICIENT USER TRACKING IN LOCATION-BASED FRIEND-FINDER	197
DAUGILAS KAKARASI, EDGARAS STAŠKAUSKAS2, CHRISTIAN VOGEL3	197
UNIVERSALIOS SĄSAJOS APLIKACIJŲ KŪRIMAS MOBILIEMS ĮRENGINIAMS	202
TADAS LINGĖ	202
AUTO-LOCATION SYSTEM FOR WIRELESS SENSOR NETWORK.....	206
VITALIJUS MARTUSEVIČIUS1, EGIDIJUS KAZANAVIČIUS2	206
VI SEKCIJA INFORMACINĖS TECHNOLOGIJOS MOKYME.....	211
SUBJEKTYVIŲ PARAMETRŲ ĮVERTINIMAS VIDURINĖS MOKYKLOS TVARKARAŠTYJE.....	212
KRISTINA BESPALOVA.....	212
E-GUARDIAN – THE NEW CERTIFICATION PROGRAMME ON SKILLS OF ENSURING CHILDREN SAFETY USING INTERNET.....	217
MANTAS LUKOŠEVIČIUS.....	217
ECDL PRODUKTŲ DIEGIMO LIETUVOS RINKOJE GALIMYBIŲ ĮVERTINIMAS.....	220
BERNADETA MIŠEIKAITĖ.....	220
SINGLE-PLATFORM USE FOR E-LEARNING	225
PETRAS TAMOŠIŪNAS.....	225
REIKALAVIMŲ INŽINERIJOS METODIKOS ĮGYVENDINIMAS MODERNIZUOJANT INFORMATIKOS MAGISTRANTŪROS STUDIJŲ PROGRAMAS	229
SAULIUS GUDASI, VITALIJUS DENISOVAS2, 5, JURIJ TEKUTOV1, 2, 4, 5, ILONA BRAUKLYTĖ2, 3, 6	229
ĮVAIRŪS BŪDAI ECDL TESTAVIMO EFEKTYVUMUI LIETUVOJE PAGERINTI	235
MARIJA ŽARENKOVA	235
ECDL TESTAVIMO KOKYBĖS UŽTIKRINIMAS NAUDOJANT GEROSIOS PRAKTIKOS PATARIMUS	240
IEVA ŽEKAITĖI	240

DUOMENŲ ANALIZAVIMAS NAUDOJANT STATISTINES DBVS PRIEMONES

Žygimantas Vasiliauskas

Informatikos fakultetas, Kauno Technologijos Universitetas, Kaunas, Lietuva

Santrauka. Veiklos analizavimas bei sprendimų priėmimas labai svarbus šiandieninėje įmonių veikloje. Norėdamos atlikti veiklos statistinę analizę, įmonės perka brangius ir sudėtingus produktus nesigilindamos apie tų produktų nauda įmonės veikloje. Vienas iš sprendimo būdų kaip efektyviai atlikti statistinę duomenų analizę neinvestuojant didelių resursų – naudoti standartines DBVS statistines priemones. Šiomis priemonėmis įmonės gali atlikti įvairią statistinę analizę kasdienėje veikloje naudojant statistinės analizės metodus: linijinę regresiją, koreliacijos analizę, nuspėjamąją duomenų analitiką, Pareto analitiką, Chi kvadrato analitiką, ANOVA.

Raktiniai žodžiai: veiklos analizė, DBVS statistinės analizės priemonės, statistinės analizės metodai

I. ĮVADAS

Veiklos analizė tai sąvoka, kuri glaudžiai susijusi su verslo valdymu ir apima reikalingų duomenų surinkimą iš pradinių šaltinių, norimos informacijos parengimą (pagal surinktus duomenis) ir jos teikimą vartotojams, kurie priima svarbius sprendimus, verslo efektyvumui didinti[1].

Veiklos analizavimas yra labai svarbus dabartinėje įmonių sferoje. Daugelis įmonių kaupia ir naudoja informacija verslo plėtrai ir naujų sprendimų priėmimui, tačiau informacija tampa tik duomenimis, kuriuos labai sunku panaudoti dėl jų vis didėjančio kiekio. Duomenys kurie nėra apdoroti, parengti darbui - labai sunku analizuoti ir priimti svarbius verslui sprendimus[2].

Prieš keletą dešimtmečių buvo sukurti pirmieji veiklos analizės įrankiai, kurie padėjo analizuoti organizacijų sukaupiamus ir saugomus duomenis, tai buvo pirmieji žingsniai, kai kompiuterizuotos priemonės padėdavo valdyti organizacijų vykdomas veiklas.

Šiandien veiklos duomenų analizavimo įrankiai tapo viena iš svarbiausių priemonių kokybiškai ir efektyviai verslo analizei, tačiau su kiekvienais metais didėjantis įrankių funkcionalumas daro juos vis sudėtingesnius ir reikalauja vis daugiau specifinių žinių ir nemažų finansinių investicijų.

Įmonės norėdamos analizuoti konkrečios veiklos duomenis, tam kad priimti sprendimus, kurių pagalba būtų įmanoma optimizuoti veiklą, naudoja brangiai kainuojančius įrankius, tokius kaip SAS ar SPSS, kurie savo funkcionalumu yra labai sudėtingi[2]. Dažnai įmonės išsigijusios šiuos produktus pajaučia kompetencijos stoka, be to investuodamos didelius resursus į žmonių apmokymus bei šių produktų palaikymą patiria nemažas sąnaudas. Įmonės poreikiai yra mažesni nei jiems gali duoti įrankis, todėl įmonėms neracionalu pirkti didelius, savo funkcionalumu ir kaina, įrankius. Ypač tai aktualu mažoms ir vidutinėms verslo įmonėms.

Alternatyvus problemos sprendimo būdas - naudoti integruotas DBVS statistines - analitines priemones, kurių pagalba galima atlikti statistinę veiklos analizę ir priimti reikalingus sprendimus, tam neišleidus didelių resursų ir pasinaudojant turimomis organizacijų specialistų žiniomis. Šios funkcijos naudojamos SQL kalbos užklausoje ir gali būti integruotos į duomenų grafinio vaizdavimo įrankius.

Šio straipsnio tikslas pateikti alternatyvų įmonės veiklos analizės procesą, kuriame būtų naudojamos integruotos DBVS statistinės priemonės, tuo siekiant minimizuoti statistinės veiklos analizės kaštus įmonėse bei panaudoti jau turimas organizacijos specialistų žinias.

II. DUOMENŲ STATISTINĖS ANALIZĖS PROCESAS

Veiklos analizei naudojami duomenys dažniausia saugomi reliacinėse duomenų bazėse. Jose yra saugomi finansiniai, gamybiniai ar kitokie įmonės duomenys, galintys įtakoti įmonės veiklą. Duomenys, reikalingi priimti sprendimus, būna iš keleto šaltinių, kurie gali būti ir išoriniai, gaunami iš verslo partnerių ar kitų informacijos teikėjų. Tuomet atlikti tikslią veiklos analizę tampa labai sudėtinga, kai duomenys nėra lokalizuoti ir yra nesutvarkyti. Šiai problemai spręsti naudojamos duomenų saugyklos, kuriose saugomi apdoroti duomenys paimti iš įvairių šaltinių, pavaizduotos veiklos analizės proceso diagramoje „Paveikslas 1. Veiklos analizės proceso diagrama“, kuriose saugomi integruoti, susieti duomenys.

Trys pagrindinės veiklos analizės technologijų dalys [2]:

4. Duomenų surinkimas ir parengimas

Reikalingų informacijai parengti duomenų surinkimas iš pradinių šaltinių, jos apdorojimas ir patalpinimas į duomenų saugyklą.

Informacijos naudotojams skirtos aktualios informacijos, esančios operacinėse sistemose ar kituose išoriniuose šaltiniuose, parengimas apjungiant (integruojant) surinktus duomenis, reikiamai juos transformuojant ir grupuojant, papildant skaičiavimų būdu gautais išvestiniais rodikliais bei pateikiant informacijos ir jos parengimo algoritmų paaiškinimus. Ši veiklos analizės dalis dažnai vadinama ETL

procedūra (Extraction, Transformation, Loading). ETL procedūra paprastai yra atliekama atskiroje tarnybinėje stotyje.

Paruoštos informacijos sukėlimas į analitinėms duomenų užklausoms vykdyti skirtą duomenų bazę, kuri yra vadinama duomenų saugykla.

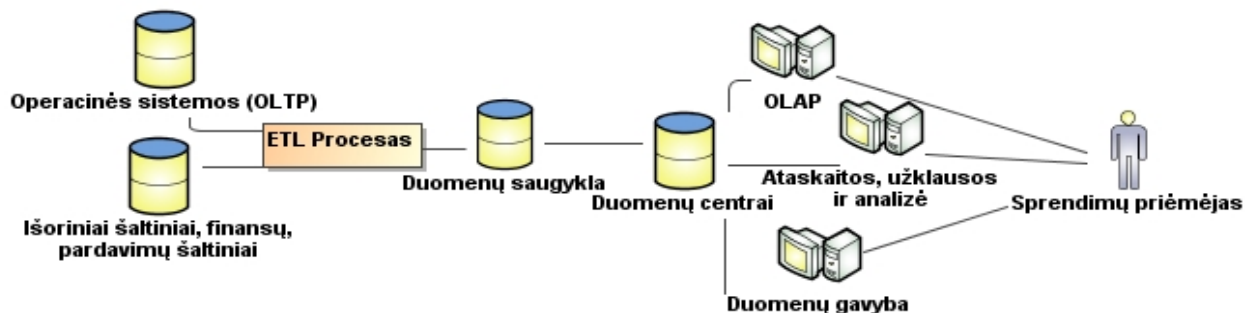
5. Duomenų saugykla

Svarbi veiklos analizės technologijų dalis yra duomenų saugyklą. Kaip jau minėta anksčiau, saugykloje saugomi sutvarkyti iš pradinių šaltinių paimti duomenys už tam tikrą aktualų laikotarpį. Informacija duomenų saugyklose parengta taip, kad ją būtų galima greitai ir patogiai nuskaityti.

Duomenų saugyklos duomenys reguliariai atnaujinami ETL procedūrų pagalba. Galutinis vartotojas arba tyrėjas duomenų saugyklos duomenų negali keisti ar taisyti, juos gali tik analizuoti.

6. Tyrėjo sąsaja

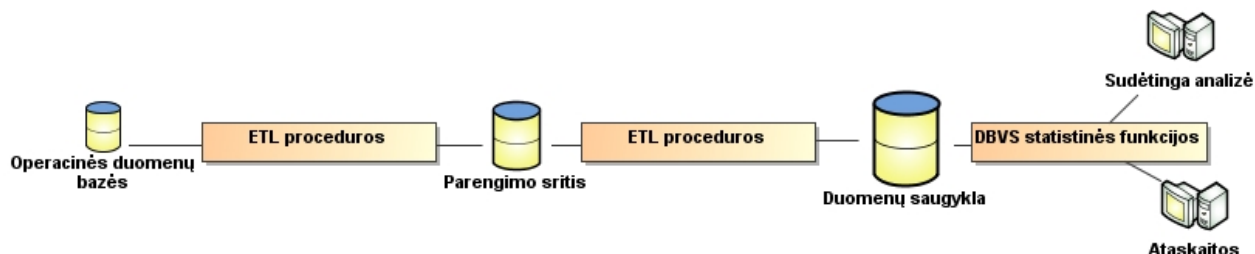
Tai kompiuterinė ir programinė įranga, kuria naudodamiesi tyrėjai iš savo darbo vietų prisijungia prie duomenų saugyklos, nusiskaito iš jos duomenys, juos analizuoja, formatuoja, rengia savo ataskaitas ir dokumentus, atlieka stebėjimo, planavimo, prognozavimo bei sprendimų priėmimo veiksmus.



Paveikslas 1. Veiklos analizės proceso diagrama

Daugelis DBVS turi integruotas analitines - statistines funkcijas, kurios gali būti naudojamos dažniausiai duomenų saugyklose, kuriose yra saugoma terabaitai įmonės duomenų, kurie jau yra paruošti analizei.

Statistines funkcijas galima panaudoti duomenų ar veiklos analizei naudojant statistinės analizės metodus: linijinę regresiją, koreliacijos analizę, nuspėjamąją duomenų analitiką, Pareto analitiką, Chi kvadrato analitiką, ANOVA[4]. Šios funkcijos gali būti panaudotos analizės užklausų rezultatų atvaizdavimo įrankiuose, pateikus SQL užklausa su šiomis statistinėmis - analitinėmis funkcijomis, kaip pateikiama 2 paveiksle.



Paveikslas 2. Duomenų saugyklos ir statistinių - analitinių funkcijų panaudojimo schema

Pagrindinis vartotojų tikslas – naudojant statistinės analizės įrankius iš sukaupytų duomenų gauti dalykinės srities informacija prasmingai apdorojus duomenis, o įvertinus informacija - priimti svarbius sprendimus, padėsiančius įmonei ateityje gauti kuo didesnę naudą. Vartotojų problemos, išskylančios naudojantis esamus sprendimus pateikiamos žemiau:

- Analizės atlikimas laiko atžvilgiu yra lėtas, nes vartotojai analizuodami duomenis statistinėmis priemonėmis sugaišta daug laiko nurodydami rodiklius bei kitus svarbius parametrus, tam kad gautų reikiamą rezultatą.
- Analizės atlikimas yra sudėtingas procesas, nes vartotojai analizuodami dalykinę sritį turi puikiai žinoti statistinius rodiklius, jų parametrus ir norimo rezultato vaizdą, todėl parinktas netikslus parametras gali sukelti visiškai priešingą analizės rezultatą nei buvo tikėtasi. Statistinės priemonės labai sudėtingos, todėl nėra tiek daug analitikų ar statistikų kurie efektyviai galėtų atlikti veiklos analizę[5].
- Dirbant su statistinės analizės sprendimais reikia turėti naudojimosi žinių, nes vartotojams reikia papildomų kursų norint turėti pagrindus dirbant su esamais sprendimais.

III. STATISTINĖS ANALIZĖS PRIEMONIŲ ĮVAIROVĖS APŽVALGA

SAS(Statistical Analysis System)

SAS yra viena iš populiariausių statistinių uždavinių analizei spręsti priemonių, atliekantis vartotojo sudėtingas ir kritiškas duomenų gavybos užduotis [2,4].

Šios priemonės pagrindinis tikslas - leisti vartotojui analizuoti įmonių veiklą ir priimti reikalingus sprendimus. Šis sprendimas naudojamas tada, kai įmonėms reikia atlikti veiklos analizę pagal tam tikrus statistinius rodiklius, atlikti sudėtingą duomenų gavybą numatant tolimesnę įmonės veiklą. Toliau pateikiami SAS privalumai ir trūkumai.

Privalumai: Gali atlikti sudėtingas duomenų gavybos užduotis;
SAS įrankių gausa leidžia efektyviai panaudoti sprendžiant įvairius statistinius uždavinius;
Vienu metu analizuoti duomenys gali daug vartotojų;
Veikia tiek Windows tiek UNIX operacinėse sistemose.

Trūkumai: Didelė kaina;
Labai sudėtingas panaudojimas. Įmonės naudoja tik 20% įrankio funkcionalumo;
Norint efektyviai panaudoti įrankį, reikia gerai mokėti programuoti SAS unikalia kalba;
Reikia keletu metų norint efektyviai išmokti naudotis priemone.

SAS daugiausia orientuotas į veiklos analizės sprendimus ir kasmet produktas funkcionalumu vis didėja, tuo pačiu kyla ir produkto kaina. Šiuo metu produktas gali pasiūlyti ne tik minėtas veiklos analizės priemones, bet ir duomenų integracijos sprendimus bei statistinės analizės sprendimus.

Microsoft Excel

Daugelis vartotojų naudojami šiuo įrankiu kasdienėje veikloje, sprendžiant matematinius ir statistinius uždavinius [5]. Daugelyje įmonių naudojamas įrankis, kuris gali pagelbėti daugumai vartotojų priimančiam sprendimus, atlikus duomenų analizę. Šis įrankis buvo skirtas tik atlikti nedidelius statistinius analizavimus, kadangi stokoja našumo, nesugeba apdoroti labai didelių duomenų kiekius (apie 100000 eilučių).

Įrankis puikiai tinka atlikti paprastus, kasdienes duomenų analizės uždavinius mažuose padaliniuose atliekant dalykinę veiklos analizę, kurios rezultatai gali būti ir nelabai tikslūs analizuojant didesnę kiekį duomenų. Įrankis yra nesudėtingas naudoti, tačiau norint atlikti analizę pagal sudėtingesnius algoritmus, reikia atlikti makro komandų programavimą Microsoft Visual Basic kalba [6]. Žemiau pateikiami Microsoft Excel privalumai bei trūkumai.

Privalumai: Atlieka nesudėtingų statistinių uždavinių analizę;
Labai lengvas išmokti;
Daug sukurtų funkcijų, kurių kasdienėje įmonės veikloje pakanka nesudėtingam uždaviniui spręsti;
Paprastas naudoti ir nėra brangus.

Trūkumai: Lėtai atlieka statistinę analizę;
Duomenis gali analizuoti tuo pačiu metu tik vienas vartotojas;
Veikia labai nestabiliai, naudojant net ir paprasčiausias funkcijas;
Papildomų funkcijų kūrimas reikalauja gylių Microsoft Visual Basic žinių;
Gali efektyviai dirbti tik su nedideliu kiekiu duomenų (iki 10000 eilučių);
Įrankis skirtas tik Windows operacinei sistemai. .

Sprendimas, kuris gali tapti puikia alternatyva leidžiančia išvengti paminėtų dalies trūkumų ir išlaikant paminėtų įrankių privalumus yra DBVS funkcijos, kurios gali būti vykdomos duomenų saugykloje, ir rezultatas gali būti grafiškai atvaizduojamas įvairiuose integruotuose įrankiuose. Kaip pavyzdinės DBVS galinčios dirbti su didžiausiais duomenų kiekiais ir paplitusios tarp verslo kompanijų yra Microsoft SQL Server, Oracle ir DB2.

DBVS statistines funkcijas naudinga naudoti, nes [7]:

- Saugumas, kadangi statistinė analizė atliekama pačioje DBVS.
- SQL statistinių funkcijų naudojimas duomenų saugyklos viduje, todėl galima sukurti tokias užklausas, kokios reikalingos veiklos analizei pagal vartotojo poreikius[9].
- Realus laiko skaičiavimai.
- Sudėtingų užklausų sudarymas ir panaudojimas kuriant dar sudėtingesnes užklausas.

DBVS statistinių funkcijų integracija versle:

- Užklausa integruoti į Oracle Discoverer, Oracle XML Publisher, Microsoft SQL Server Analysis Services, IBM Cognos.
- Užklausa galima integruoti į veiklos analizės įrankius, kurie gali bendrauti su DBVS sistemomis – Business Objects, QlickView.

Duomenų saugyklose galima sukurti vartotojo aprašytas funkcijas kartu naudojant ir SQL užklausas, kurių pagalba analizavimas pačioje duomenų saugykloje tampa efektyvesnis nei naudojant vien tik SQL kalbą[1].

IV. EKSPERIMENTINĖS DBVS IR DALYKINĖS SRITIES PASIRINKIMAS

Renkantis duomenų bazės sistemą, reikia atsižvelgti į kliento ir naudotojo poreikius, kurie gali iškilti analizuojant sukauptus duomenis. Pagal 2007 metų duomenų bazių valdymo sistemų analizę[12], geriausiai įvertintos DBVS yra Teradata, Oracle, IBM ir Microsoft.

Visos paminėtos DBVS leidžia įgyvendinti siūloma koncepcija, tačiau pagal ekspertų pateiktus duomenis, ir išanalizavus pagrindinių kūrėjų duomenų saugyklų analitines - statistines funkcijas[11, 14, 12,7], eksperimentui bus naudojamas Oracle DBVS. Oracle turintis didelį kiekį analitinių funkcijų ir priemonių, padės efektyviai analizuoti duomenis.

Atliekant veiklos analizę, galima pasirinkti įvairią sritį ir joje pritaikyti statistines funkcijas įvairiems uždaviniams spręsti. Analitinių funkcijų įvairovė ir panaudojimas labai platus, todėl efektyviai panaudoti galima įvairiose veiklose tik iškeliant vis kitokius statistinius uždavinius ir juos sprendžiant, panaudojus SQL užklausas ir analitines funkcijas.

Eksperimentui bus naudojama draudimo įmonių surinkti realūs duomenys, publikuojami nuo 2000m. apie įvykdytas draudimo sutartis. Duomenys bus apdoroti ir patalpinti Oracle DBVS. Duomenų bazės dydis yra apie 100MB, lentelių kiekis duomenų saugykloje yra 10.

Statistinio uždavinio pavyzdys gali būti pateikiamas toks:

Ekstremumo paieška pagal dimensija - kokia dimensija daro didžiausią įtaką rodikliui (koreliacija).

Šio uždavinio užklausa Oracle aplinkoje naudojant analitines funkcijas atrodys taip:

```
SELECT pasirasyta_imoka,RANK() OVER (ORDER BY pasirasyta_imoka desc) rangas,  
DENSE_RANK() OVER (ORDER BY pasirasyta_imoka desc) dens_rangas FROM pasirasytos_imokos  
WHERE RANK() OVER (ORDER BY pasirasyta_imoka desc) < 10.
```

Užklausos rezultatas bus gražinamas Oracle Discoverer įrankiu, kuris vartotojui pateiks grafinį užklausos vaizdinį.

V. IŠVADOS

- Šiame darbe išanalizuotos SAS ir Microsoft Excel įrankių bendrosios funkcinės galimybės norint atlikti statistinę duomenų analizę. Nustatyta, kad abu įrankiai padeda analizuoti įmonės duomenis ir priimti sprendimus, tačiau abu sprendimai turi ir trūkumų: SAS brangiai kainuoja ir reikalauja papildomų žinių, SAS ir Excel duomenų užkrovimas iš duomenų saugyklos užima daug vietos fiziniame atmintyje, o vykdant analizę ir operatyvioje atmintyje, SAS sudėtingas įrankių panaudojimas dalykinės srities vartotojams, Excel efektyviai gali dirbti tik su nedideliu kiekiu duomenų, kurie ypač aktualūs mažoms ir vidutinėms verslo įmonėms.
- Darbe pasiūlyta koncepcija, kai statistinei duomenų analizei naudojamos DBVS turimos vidinės funkcijos. Toks sprendimas leidžia atlikti saugų duomenų analizavimą DBVS aplinkoje, užklausos duomenų analizei sukuriama naudojant SQL kalbą ir DBVS turimas statistines funkcijas, galima atlikti realaus laiko skaičiavimus.
- Atlikus DBVS turimų statistinės analizės funkcijų aibę, nustatyta, kad didžiausią aibę tokių funkcijų turi Oracle DBVS, kuri bus naudojama ir tolimesniame eksperimentiniame tyrime.
- Atlikta SQL užklausų integravimo analizė SQL kalba palaikančiuose produktuose ir buvo nustatyta jog SQL užklausa galima integruoti į Oracle Discoverer, Oracle XML Publisher, Microsoft SQL Server Analysis Services, IBM Cognos, Business Objects, QlickView priemones.

LITERATŪRA

1. Jonathan G. Geiger; Abdul Rahman; Bruce Barth. *Reuse It or Lose It: Achieving Business Intelligence Efficiency*. Prieiga per internetą: http://findarticles.com/p/articles/mi_qa5525/is_200801/ai_n31108933
2. *Business Intelligence: The Savvy Manager's Guide (The Savvy Manager's Guides)*. Sudarytojas David Loshin; 2003 m., 250 p.
3. Gareth Herschel. *Magic Quadrant for Customer Data Mining, IQ06*. Prieiga per internetą: <http://sameer1prakash.googlepages.com/CRM-DataMiningVendors2006.pdf>
4. *Applied Data Mining: Statistical Methods for Business and Industry (Statistics in Practice)*. Sudarytojas Paolo Giudici; 2003m., 376 p.
5. Jeff Kelly. *Rankings of customer data mining market show shift to user-friendly tools*. Prieiga per internetą: http://searchdatamanagement.techtarget.com/news/article/0,289142,sid91_gci1321221,00.html
6. Colin Beast. *Gartner Crowns SAS and SPSS Data Mining Kings*. Prieiga per internetą: <http://www.destinationcrm.com/Articles/ReadArticle.aspx?ArticleID=42362>
7. Rick Sherman. *Microsoft Excel: The king of BI*. Prieiga per internetą: http://searchcrm.techtarget.com/news/column/0,294698,sid11_gci1081869,00.html
8. *Data Analysis Using SQL and Excel*. Sudarytojas Gordon S. Linoff; 2007 m., 645 p.
9. *Expert One-on-One Oracle*. Sudarytojas Thomas Kyte; 2001 m., 1297 p.
10. Carlos Ordonez. Building statistical models and scoring with UDF. *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, Beijing, China, 2007, p. 1005 – 1016.
11. Zhibo Chen; Carlos Ordonez. Efficient OLAP with UDF. *Proceeding of the ACM 11th international workshop on Data warehousing and OLAP*, Napa Valley, California, USA, 2008, p. 41 – 48.
12. Donald Feinberg; Mark A. Beyer. *Magic Quadrant for Data Warehouse Database Management Systems*. Prieiga per internetą: <http://mediaproducts.gartner.com/reprints/microsoft/vol3/article7/article7.html>

13. Marcos M. Campos. *Analytics in the Oracle Databases*. Prieiga per internetą: http://oracledmt.blogspot.com/2006/01/analytics-in-oracle-database_05.html
14. Mingguang Xu; Denise Gardner. *Oracle Analytic Functions for IR Analysis and Reporting*. Prieiga per internetą: <http://www.uga.edu/oir/reports/OracleAnalyticFunction-SAIR-2006.ppt>