

ხათუნა აბულაშვილი

საგადასახადო სისტემისათვის ინფორმაციული
უზრუნველყოფის სისრულის და წინააღმდეგობითობის
მოდელების აგების პრინციპების დამუშავება და კვლევა

წარმოდგენილია დოქტორის აკადემიური ხარისხის
მოსაპოვებლად

საქართველოს ტექნიკური უნივერსიტეტი
თბილისი, საქართველო
ოქტომბერი, 2008

საქართველოს ტექნიკური უნივერსიტეტი

ეკონომიკური ინფორმატიკა

ჩვენ, ქვემოთ ხელისმომწერნი ვადასტურებთ, რომ გავაცანით აბულაშვილი ხათუნას მიერ შესრულებულ სადისერტაციო ნაშრომს დასახელებით: საგადასახადო სისტემისათვის ინფორმაციული უზრუნველყოფის სისრულის და წინააღმდეგობითობის მოდელების აგების პრინციპების დამუშავება და კვლევა და ვაძლევთ რეკომენდაციას საქართველოს ტექნიკური უნივერსიტეტის ეკონომიკური ინფორმატიკის სადისერტაციო საბჭოში მის განხილვას დოქტორის აკადემიური ხარისხის მოსაპოვებლად.

ხელმძღვანელი: პროფ., ტ.მ.დ. ზ. ბოსიკაშვილი

რეცენზენტი: პროფ., ტ.მ.დ. ზ. გასიტაშვილი

რეცენზენტი: ტ.მ.კ. ა. ქევზიშვილი

რეცენზენტი:

საქართველოს ტექნიკური უნივერსიტეტი

2008

ავტორი: აბულაშვილი ხათუნა

დასახელება: საგადასახადო სისტემისათვის ინფორმაციული
უზრუნველყოფის სისრულის და
წინააღმდეგობითობის მოდელების აგების
პრინციპების დამუშავება და კვლევა

ფაკულტეტი : ეკონომიკური ინფორმატიკა

ხარისხი: დოქტორი

სხდომა ჩატარდა: თარიღი

ინდივიდუალური პიროვნებების ან ინსტიტუტების მიერ
ზემომოყვანილი დასახელების დისერტაციის გაცნობის მიზნით მოთხოვნის
შემთხვევაში მისი არაკომერციული მიზნებით კოპირებისა და გავრცელების
უფლება მინიჭებული აქვს საქართველოს ტექნიკურ უნივერსიტეტს.

ავტორის ხელმოწერა

ავტორი ინარჩუნებს დანარჩენ საგამომცემლო უფლებებს და არც
მთლიანი ნაშრომის და არც მისი ცალკეული კომპონენტების გადაბეჭდვა ან
სხვა რაიმე მეთოდით რეპროდუქცია დაუშვებელია ავტორის წერილობითი
ნებართვის გარეშე.

ავტორი ირწმუნება, რომ ნაშრომში გამოყენებული საავტორო
უფლებებით დაცული მასალებზე მიღებულია შესაბამისი ნებართვა (გარდა
ის მცირე ზომის ციტატებისა, რომლებიც მოითხოვენ მხოლოდ სპეციფიურ
მიმართებას ლიტერატურის ციტირებაში, როგორც ეს მიღებულია
სამეცნიერო ნაშრომების შესრულებისას) და ყველა მათგანზე იღებს
პასუხისმგებლობას.

მიძღვნა:

საღისერტაციო ნაშრომი ეძღვნება პირველ რიგში ჩემს
ახალშობილ შვილს მირაკოვ ბორისს და ბაბუას ვახტანგიშვილ
შალვას. აგრეთვე ყველა ჩემ დანარჩენ ოჯახის წევრებს,
რომლებმაც წვლილი შეიტანეს ამ საღისერტაციო ნაშრომის
შესრულებაში.

რეზიუმე

თემის აქტუალურობა. თანამედროვე ბიზნეს სამყაროს საქმიანობის ნებისმიერ სფეროში ინფორმაციის მოცულობა ორგანიზაციებში საკმაოდ დიდია და ორგანიზაციის წარმატებული მუშაობა დამოკიდებულია მასში არსებული კუთვნილი ინფორმაციიდან მაქსიმუმის გამოყენებასთან. მონაცემების მხოლოდ არსებობა არ არის საკმარისი მუშაობის მაჩვენებლების გასაუმჯობესებლად. საჭიროა არსებული მონაცემების ტრანსფორმირების ცოდნა, ამისათვის არის წარმოდგენილი ტექნოლოგია **Data Mining**.

Data Mining- ის გამოყენების სფერო არ არის განსაზღვრული. Data Mining-ის აპლიკაციები ვრცლად არის გავრცელებული: საბითუმო ვაჭრობაში, მარკეტინგში, ფინანსებში, მედიცინაში, წარმოებაში და სხვა სფეროებში.

საქართველოში საბაზრო ეკონომიკის დამყარებასა და განვითარებასთან ერთად რა თქმა უნდა იზრდება მცირე თუ მსხვილი საწარმოების რაოდენობა, ანუ გადამხდელების რაოდენობა. საგადასახადო სტრუქტურაში დასამუშავებელი ინფორმაციის მოცულობა ყოველდღიურად იზრდება, რაც მოითხოვს ახალი თანამედროვე და ეფექტური ინსტრუმენტებისა და გადაწყვეტილებების დანერგვას. გადამხდელისაგან მიღებული ინფორმაციის დამუშავებასთან არსებული პრობლემის გადაწყვეტას და მისგან ცოდნის მოპოვების საშუალებას იძლევა Data Mining ტექნოლოგიის გამოყენება. ამ ტექნოლოგიის გამოყენება ჩემი თვალსაზრისით მოგვცემს საშუალებას შევამციროდ დროის და შრომითი დანახარჯები გადამხდელთა პირადი აღრიცხვის ბარათის ინფორმაციის დამუშავებისას და მისი მიხედვით პროგნოზირების ჩატარების, აგრეთვე კრიტერიუმების შერჩევის საფუძველზე, შეგვიძლია გამოვავლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი, რითაც შესაძლებელი გახდება ყურადღების გამახვილება იმ დარგებზე, რომლებსაც ბიუჯეტისთვის ნაკლები შემოსავალი მოაქვთ.

თემის დამუშავების ხარისხი. ამ დროისათვის Data Mining – თან დაკავშირებული საკითხები, სამამულო ლიტერატურაში არ არის გაშუქებული. არსებობს ცალკეული მცდელობები არსებული ტექნოლოგიის გამოყენების. ამ ტექნოლოგიის ასპექტები სხვადასხვა სფეროებში განიხილება. გაცილებით ფართოდ ხდება ამ ტექნოლოგიის გამოყენების და მისი არსის აღწერა უცხოურ ლიტერატურაში. მაგრამ აქაც კომპლექსური ანალიზი და ოპერაციები, რომლებიც ტარდება ამ ტექნოლოგიის გამოყენებისას, სამეცნიერო ლიტერატურაში ნაჩვენები არ არის.

უნდა აღინიშნოს, რომ არსებული ტექნოლოგია როგორც წესი წარმოდგენილია და აღწერილია შემდეგი სფეროებისათვის: საბანკო საქმე, საბითუმო ვაჭრობა, ტელეკომუნიკაცია, დაზღვევა, მედიცინა.

ამ დროისათვის DATA MINING ტექნოლოგიის გამოყენება საგადასახადო სისტემისათვის არ არის გავრცელებული. ნამუშევარში აღწერილი მაქვს, თუ როგორ გამოვიყენეთ ეს ტექნოლოგია საგადასახადო სტრუქტურაში და როგორ მოხდა მისი დანერგვა და ინტეგრირება სხვა პროდუქტებთან. თემის კვლევასთან მიახლოებული

არის შემდეგი ავტორების ნამუშევრები: Дюк В., Большаков П.С., Пройдаков Э., Бергер Ч., Киселев М., Соломатин Е., Han J., Newquist H. P., Parsaye K.

კვლევის საგანი და ობიექტი. კვლევის ობიექტს წარმოადგენს ტექნოლოგია Data Mining. ხოლო კვლევის საგანს წარმოადგენს გადამხდელთა პირადი აღრიცხვის ბარათი საქართველოს რესპუბლიკის შემოსავლების სამსახურის ინფორმაციული ტექნოლოგიების სამმართველოში არსებული მონაცემთა საცავების საფუძველზე.

კვლევის ამოცანები და მიზნები. სადისერტაციო ნამუშევრის მიზანია წარმოვადგინოთ, თუ როგორ შეიძლება Data Mining ტექნოლოგიის გამოყენება საგადასახადო სტრუქტურაში. კვლევის საერთო მიზნის წარმოსადგენად დისერტაციაში დასმულია შემდეგი ამოცანები:

- კომპლექსური შინაარსის ანალიზი, ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურა და მისი რეალიზაციისათვის არსებული პროდუქტები. ერთ-ერთი ძირითადი Data Mining-ია;
- Data Mining-ის არსის და მასთან არსებული სხვა საკვანძო ტერმინოლოგიის განსაზღვრა;
- Data Mining ტექნოლოგიის გამოყენების სფეროები;
- Data Mining ტექნოლოგიის გამოყენების პერსპექტიულობის აღწერა საგადასახადო სისტემისათვის;
- Data Mining-ის მეშვეობით წინააღმდეგობითობის მოდელების შექმნა და მისი გამოყენება;

კვლევის მეთოდოლოგია და თეორიული საფუძველი. არსებული დისერტაცია ეფუძვნება საგნის შესწავლის სისტემურ და დიალექტურ მიდგომებს და გამომდინარეობს იქიდან, რომ Data Mininig ერთ-ერთი წამყვანია ინტელექტუალურ ანალიზში.

კვლევისას და მოდელის აგებისას გამოყენებულია შემდეგი მეთოდები და მიდგომები: ანალიზი, სინთეზი, მოდელირება, ისტორიული მეთოდები.

დისერტაციის მეცნიერულ სიახლეს წარმოადგენს, შემუშავებული ალგორითმის მიხედვით, პროგნოზირების ჩატარება. კრიტერიუმების შერჩევის საფუძველზე ვლდებულობით პროგნოზს, რომლის შედეგადაც შეგვიძლია გამოვავლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი.

დიდი ყურადღება უნდა მიექცეს მონაცემების მომზადებას. უნდა მოხდეს მათი გადარჩევა და გამოიყოს არასწორი მნიშვნელობები, ანუ მოხდეს ინფორმაციის “გაცრა”.

აგრეთვე სიახლეს წარმოადგენს ის, რომ Data Mininig-ი პირველად გამოვიყენეთ ისეთ სპეციფიურ სფეროში, როგორიც საგადასახადო სისტემაა.

კვლევის რეზულტატების დანერგვა

სადისერტაციო ნაშრომში წარმოდგინილი კვლევის რეზულტატები დანერგილია საქართველოს შემოსავლების სამსახურის საგადასახადო სისტემაში.

მოდელის შექმნისას წარმოიქმნა შემდეგი სიძნელე: მონაცემთა ბაზა Oracle9-ზე გართულდა ODM10-ის დაყენება, რის გამოც ჩვენ დაგვჭირდა სპეციალური აპლიკაციების დაყენება.

დისერტაციის სტრუქტურა განსაზღვრულია კვლევის ამოცანებისა და მიზნების, აგრეთვე საგნობრივი განსაკუთრებულობისა და ობიექტის დამუშავების გათვალისწინებით. ნამუშევარი შედგება შესავლის, სამი თავის და ბიბლიოგრაფიული სიისაგან.

პირველ თავში განხილულია თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის (იას) არქიტექტურა. სადისერტაციო ნაშრომში განხილულია იას არქიტექტურის დონეები და წარმოდგენილია ტიპური ინსტრუმენტების მაგალითები. აღწერილია შერეული გადაწყვეტილებების მიღების მეთოდის გამოყენება. მაგალითში შერეული გადაწყვეტილებების რეალიზაციისათვის ვიყენებთ ფირმა Cognos-ის მონაცემთა ანალიზის ინსტრუმენტებს და ფირმა Oracle – ის მონაცემთა ბაზების მართვის სისტემების ინსტრუმენტებს.

ამგვარად, პირველ თავში ჩვენ გავცნობით თუ რა პროდუქტებზე და პლატფორმებზე არის დამუშავებული კვლევის ძირითადი საკითხები ჩვენს მიერ წარმოდგენილ სადისერტაციო ნაშრომში.

მეორე თავში წარმოდგენილია ინტელექტუალური ანალიზისათვის გამოყენებული პროდუქტი Data Mining. აღწერილია, რას წარმოადგენს ეს ტექნოლოგია და მისი გამოყენების სფეროები. ნაჩვენებია ტიპების კანონზომიერება და ინტელექტუალური ანალიზის კლასების სისტემები. გამოიყოფა ხუთი კანონზომიერების სტანდარტული ტიპი, რომლებიც საშუალებას იძლევიან გამოვააშკაროთ Data Mining-ის მეთოდები. მოცემულ თავში ნაჩვენებია, თუ რა უპირატესობა გააჩნია Data Mining-ს სხვა ტექნოლოგიებთან.

მესამე თავში წარმოდგენილია სადისერტაციო ნამუშევრის კვლევითი ნაწილი და ჩვენს მიერ კვლევის რეზულტატების დანერგვა საგადასახადო სტრუქტურაში. ნაჩვენებია, თუ როგორ ხდება Data Mining ტექნოლოგიის გამოყენება ამ სფეროში.

მოცემულ თავში აღწერილია, თუ რას წარმოადგენს გადამხდელის პირადი აღრიცხვის ბარათი და როგორ ხდება მისი წარმოება. შემდგომ წარმოდგენილია საგადასახადო სისტ. ინფ. უზრუნველყოფის სისტემის და წინააღმდეგობითობის მოდელების აგების პრინციპები. და ბოლოს, დამუშავება და კვლევა. ჩვენს შემთხვევაში ვიყენებთ Data Mining –ის კლასიფიკაციის მეთოდს, რომელიც იყენებს ჩვენს მიერ შექმნილ ალგორითმს. ხდება ინფორმაციის მომდევნო მომზადება ჩვენს მიერ შექმნილი ალგორითმისათვის. პროგნოზირებისათვის ვიყენებთ სხვადასხვა ცხრილებს, რომლიდანაც ვღებულობთ ინფორმაციას ჩვენს მიერ დაწერილი SELECT-ით.

მიღებული მონაცემები წარმოდგენილია აგრეთვე chart-ების სახით. და ბოლოს კრიტერიუმების შერჩევის საფუძველზე ვღებულობთ პროგნოზს, რომლის შედეგადაც შეგვიძლია გამოვავლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი, რის საფუძველზეც შეგვიძლია დავადგინოთ, თუ რომელ დარგს შეუძლია მოუტანოს ბიუჯეტისთვის მეტი შემოსავალი და რომელ დარგს უნდა მიექციოს ყურადღება.

SUMMARY

Actuality of this work. In any sphere of the activity modern business world the volume of information in the organization is significantly large and successful activities of the organizations depend on how they manage to use the maximum of information available. The only fact that information exists is not sufficient. It is necessary to know how to transform the data available. This is what the Data Mining technology is used for. We may use this new knowledge for the improvement of marketing efficiency of the database.

The fields where Data Mining is applicable are not defined.

First Data Mining methods became subject to interest of commercial enterprises which carry out their projects on the basis of Data Warehousing.

Application of Data Mining is widely spread in wholesale trading, marketing, finances, healthcare, production and other fields.

Together with the process of establishment and development of market economy in Georgia of course there is an increase in the number of private small or large enterprises e.g. taxpayers. The volume of information in the taxation structure is increasing daily and this in turn requires implementation of new modern and effective instruments and decisions. One of the most significant aspects of the development of main state organ is development of information technologies. This gives possibility of fast and exact processing of large flow of information. The implication of Data Mining technology gives possibility to solve the problem of receiving and processing information from taxpayers and gathering all the necessary knowledge. Use of this technology, in my opinion, will give us possibility to decrease time and work expenditure during the process of identification of taxpayers and processing of all the relevant information. Also it will give a possibility of forecasting and thus revealing the fields, which are insufficiently developed and as a result giving emphasis to fields, which bring less income to budget.

Quality of the topic processing. At this moment issues, connected to Data Mining are not available in local literature. There are several trial applications of this technology. We can find some discussions on the aspects of this technology in foreign literature. But complex analysis of the use this technology in scientific literature is not available.

It should be mentioned, that existing technology mainly is represented and described for the following fields: banking, wholesales trading, telecommunications, insurance and healthcare.

For this moment the use of Data Mining technology for taxation system is not spread. The work presents possibilities of implication of this system in Taxation System and the ways of integration with other products. The most closest works related to this theme are presented by: Duke B, Bolshakov P.S., E. Proidakov, C. Berger, M. Kiselev., E. Solomatine, Han J., Newquist H.P., Parsaye K.

Subject and objective of research. The subject of research is Data Mining technology. And the objective is personal identification of taxpayers in Georgia by information and technology departments of country's state tax incomes service on the basis of existing databases.

The aim and goal of the research. The aim of this work is to imagine how it is possible to use Data Mining technology in taxation structure. In order to present the aim of the dissertation provides following tasks:

- Complex content analysis, architecture of information analytical system and products available for its realization. One of the main is Data Mining;
- Main point of Data Mining and definition of terminology;
- Fields where Data Mining technology is used.
- Prospective for the use of Data Mining technology in taxation system;
- Creation of opposite models through Data Mining and the use of it.

Methodology of research and theoretical basis. Existing dissertation is based on the system and dialect approach to the topic research and outcomes from the fact that Data Mining is one of the leading technologies in intellectual analysis. Following methods and models are used during the research and creation of the model: analysis, synthesis, modeling, and historical methods.

The scientific innovation of the dissertation is prospective analysis in accordance with worked out algorithms. With the choice of criteria we get prognosis revealing fields, which are badly developed.

Special attention should be paid to the data preparation. Selection and separation of incorrect denominates should be made, in particular, the information should be well sifted.

As an innovation Data Mining is becoming to be applied in such specific fields as taxation system.

Implication of the results of research.

The results, presented in dissertation work are introduced in taxation system of the ministry of state incomes of Georgia.

During the creation of the model following complications appeared: It was hard to install ODM10 on the Oracle9 database, so we had to install additional special supporting applications.

First chapter of the dissertation describes the architecture of modern information-analytical system. The dissertation describes levels of violet architecture and examples of typical instruments. The use of mixed decision making methods is described. For the realization of mixed decisions we use Cogno data analysis and Oracle database management instruments.

Second Chapter presents Data Mining intellectual analysis product. There is an explanation on what exactly this product represents and what the fields of use are. The types and classes of intellectual analysis of classes are also shown. Five standard natural types revealing the Data Mining technology are presented. The chapter also provides some explanations on advantages of Data Mining technology over other existing technologies.

Third Chapter represents research part of dissertation work as well as the implication of research results in the taxation structure. It is showed how Data Mining Technology is implemented in that field.

This chapter represents the meaning of personal identification of taxpayer and how it should be performed.

Then there are some explanations on the information provision of taxation system and opposite modeling principals.

And finally there is some processing and research. In this particular case we take Data Mining classification method which uses algorithm created by us. We make further preparation of the information for the algorithm created by us. For the prognosis we use different tables where we take information through the written select.

The data is also represented by charts. And finally on the basis of choosing criteria we get prognosis which reveal insufficiently developed fields. And this in turn enables us to define which field might provide more income to the budget and which field is subject to more attention.

შინაარსი:

შესავალი	15
I თავი. თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის მიმოხილვა	21
1.1 თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურა	21
1.2. ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურის რეალიზაცია შერეული გადაწყვეტილებების საფუძველზე	42
1.3. ინფორმაციულ-ტექნოლოგიური სისტემის არქიტექტურის რეალიზაციის პლატფორმულ-ბაზირებული გადაწყვეტილებების საფუძველზე	52
II თავი. DATA MINING-ტექნოლოგიის არსი და მისი გამოყენების სფეროები	77
2.1. Data Mining-ის შინაარსი	77
2.2. ტიპების კანონზომიერება და ინტელექტუალური ანალიზის კლასების სისტემები	85
2.3. Data Mining-ის უპირატესობები	93
III თავი. შედეგების პრაქტიკული რეალიზაცია და ანალიზი ..	98
3.1. გადამხდელის პირადი ბარათი	98
3.2. საგადასახადო სისტემისათვის ინფორმაციული უზრუნველყოფის სისტემის და წინააღმდეგობითობის მოდელების აგების პრინციპი	104
3.3. დამუშავება და კვლევა	106
დასკვნა	111
გამოყენებული ლიტერატურა	112

ცხრილების ნუსხა:

№	დასახელება	გვ.
1	ინფორმაციულ-ანალიტიკური სისტემის შერეული გადაწყვეტილებების რეალიზაციისათვის საჭირო პროდუქტები	43-44
2	ინფორმაციულ-ანალიტიკური სისტემის რეალიზაცია Oracle ინსტრუმენტების მეშვეობით. პროდუქტების შესაბამისობის სია	53-54
3	Oracle Data Mining –ში რეალიზებული ალგორითმები	75
4	OLAP და Data Mining – ში ამოცანების ფორმულირების და მეთოდების გამოყენების მაგალითები	78
5	ცხრილი T_PROB	108

ნახაზების ნუსხა:

№	დასახელება	გვ.
1	თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურა. მონაცემთა წყაროების მრავალფეროვნება და მათი გამოყენება ყოველ კონკრეტულ შემთხვევაში აიხსნება ინფორმაციის შენახვის სხვადასხვა საჭიროებით, ანუ ორგანიზაციის წინაშე მდგომი ამოცანების მიხედვით	22
2	მოთხოვნათა წარმოდგენა	34
3	COGNOS საქმიანი ანალიზის პაკეტი	45
4	Data Warehouse Builder-ი გვაძლევს საცავეების მოდელირების სქემის შექმნის საშუალებას	56
5	მონაცემთა ბაზების მართვის სისტემა	61
6	Workflow Builder გრაფიკული არე	67
7	ანალიტიკური სისტემების შექმნა, რომლებიც შექმნილნი არიან მრავალგანზომილებიანი ანალიზისა და OLAP ტექნოლოგიის საფუძველზე	69
8	თავისუფალი ფორმის რეპორტების ფორმირება Oracle Discoverer-ში	73
9	Oracle Data Mining ინსტრუმენტალური არე	74
10	DATA MINING-ის პოპულარული პროდუქტები	92
11	გადამხდელის პირადი ბარათი	101
12	კონკრეტული ბარათის წარმოების პროცესი	105
13	წარმოდგენილია კლასიფიკაციის მოდელის შექმნის პროცესის ფრაგმენტი. ინფორმაციის მომდევნო მომზადება ჩვენს მიერ შექმნილი ალგორითმისათვის	106
14	ჩვენს მიერ შექმნილი ალგორითმი	107
15	ინფორმაცია ყოველი საქმიანობის რაოდენობრივი მაჩვენებლების შესახებ. კერძო შემთხვევა	109
16	დიაგრამა, მონაცემების განხილვა გადასახადების მიხედვით	109
17	დიაგრამა, მონაცემების განხილვა საქმიანობის სახეების მიხედვით	109
18	დიაგრამა, მონაცემების განხილვა საგადასახადო ინსპექციების მიხედვით	110
19	ალგორითმის კრიტერიუმების შერჩევის საფუძველზე მიღებული პროგნოზი	110

მაღლიერება:

მაღლობა მინდა გადავუხადო ჩემს ხელმძღვანელს, ბატონ ზურაბ ბოსიკაშვილს წლების განმავლობაში გაწეული დახმარებისათვის. ბატონ ზურაბ ბოსიკაშვილის მიერ კვლევისათვის საჭირო ინფორმაციის მოწოდებით და კონსულტაციებით მოგახერხე საღისერტაციო ნაშრომის შესრულება.

ნაშრომის შესრულებისას ზოგად საკითხებში დახმარებისათვის მაღლიერება მინდა გამოვხატო ბატონ არჩილ ფრანბიშვილის მიმართ. მისი ხელშეწყობით შესაძლებელი გახდა ნაშრომთან დაკავშირებული პრობლემების გადაწყვეტა.

მაღლობას ვუხდი ქალბატონ ლილი იოსელიანს ასპირანტურაში სწავლების პერიოდში საორბანიზაციო საკითხებში გაწეული დახმარებისათვის.

ასევე, მაღლობას ვუხდი ეკატერინე მაღრაძეს საღისერტაციო ნაშრომში კვლევის ჩატარებისას დახმარების გაწევისათვის.

შესავალი

თემის აქტუალურობა. თანამედროვე ბიზნეს სამყაროს საქმიანობის ნებისმიერ სფეროში ინფორმაციის მოცულობა ორგანიზაციებში საკმაოდ დიდია და ორგანიზაციის წარმატებული მუშაობა დამოკიდებულია მასში არსებული კუთვნილი ინფორმაციიდან მაქსიმუმის გამოყენებასთან. საკმაოდ რთულია მონაცემთა მნიშვნელოვნების შეფასება, რომლებიც უწყვეტად გროვდება ადამიანების საქმიანობის პროცესში, რომლებიც მონაწილეობენ ბიზნესის მართვასა და წარმოებაში, საბანკო საქმიანობაში, სამეცნიერო, ინჟინერული თუ სამედიცინო ამოცანების გადაწყვეტაში. მძლავრი კომპიუტერული სისტემები, რომლებიც მოიცავენ, ინახავენ და მართავენ უზარმაზარ მონაცემთა ბაზებს წარმოადგენენ განუყოფელ ატრიბუტს მსხვილი კორპორაციების საქმიანობაში და არა მარტო მსხვილი კორპორაციებისათვის, მაგრამ მონაცემების მხოლოდ არსებობა არ არის საკმარისი მუშაობის მაჩვენებლების გასაუმჯობესებლად. საჭიროა არსებული მონაცემების ტრანსფორმირების ცოდნა, მათგან სარგებლიანობის მისაღწევად და გადაწყვეტილებების მისაღებად. ამისათვის არის წარმოდგენილი ტექნოლოგია **Data Mining**. ახალი ცოდნის მოპოვება შესაძლოა გამოვიყენოთ მონაცემთა ბაზების მარკეტინგული ეფექტიანობის ასამაღლებლად. მომხმარებლის სათანადო მონაცემების ანალიზის მეშვეობით შესაძლოა ვიწინასწარმეტყველოთ კლიენტის მოქმედება და აგრეთვე მოვახდინოთ მასზე ზეგავლენაც კი.

Data Mining-ის გამოყენების სფერო არ არის განსაღვრული – ის ყველგან არის სადაც კი არსებობს მონაცემები. მაგრამ პირველ რიგში Data Mining –ის მეთოდებმა დააინტრიგა კომერციული საწარმოები, რომლებიც თავიანთ პროექტებს ახორციელებენ ინფ. მონაცემთა საცავის (Data Warehousing) საფუძველზე ასეთი საწარმოების მაგალითი გვჩვენებს, რომ Data Mining-ის გამოყენების ეფექტიანობა აღწევს 100%.

Data Mining-ი წარმოადგენს დიდ ფასეულობას ანალიტიკოსებისათვის მათ ყოველდღიურ საქმიანობაში. საქმიანმა ხალხმა კარგად გაითავისა, რომ Data Mining-ის მეთოდების მეშვეობით

მათ შეუძლიათ მიიღონ საგრძნობი უპირატესობა კონკურენტულ ბრძოლაში.

Data Mining- ის აპლიკაციები ვრცლად არის გავრცელებული: საბითუმო ვაჭრობაში, მარკეტინგში, ფინანსებში, მედიცინაში, წარმოებაში და სხვა სფეროებში.

საქართველოში საბაზრო ეკონომიკის დამყარებასა და განვითარებასთან ერთად რა თქმა უნდა იზრდება მფლობელი სუბიექტების (მცირე თუ მსხვილი საწარმოების და ა.შ.), ანუ გადამხდელების რაოდენობა. საგადასახადო სტრუქტურაში დასამუშავებელი ინფორმაციის მოცულობა ყოველდღიურად იზრდება, რაც მოითხოვს ახალი თანამედროვე და ეფექტური ინსტრუმენტებისა და გადაწყვეტილებების დანერგვას. ერთ-ერთი ყველაზე ძირითადი სახელმწიფო ორგანოს ფუნქციონირების ასპექტს, რომელიც არეგულირებს გადასახადების მოგროვების პროცესს, წარმოადგენს ინფორმაციული ტექნოლოგიების განვითარება, რომელიც იძლევა საშუალებას უფრო ოპერატიულად და ზუსტად დავამუშავოთ ინფორმაციის დიდი ნაკადები. გადამხდელისაგან მიღებული ინფორმაციის დამუშავების, პრობლემის გადაწყვეტის და მისგან ცოდნის მოპოვების საშუალებას იძლევა Data Mining ტექნოლოგიის გამოყენება. ცოდნის მოპოვება ამ ტექნოლოგიის და კერძოდ, პროდუქტ Oracle Data Miner-ის გამოყენება, ჩემი თვალსაზრისით, საშუალებას მოგვცემს შევამციროთ დროისა და შრომითი დანახარჯები გადამხდელთა პირადი აღრიცხვის ბარათის ინფორმაციის დამუშავებისას და მის მიხედვით პროგნოზირების ჩატარებისას; აგრეთვე კრიტერიუმების შერჩევის საფუძველზე შეგვიძლია გამოვავლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი რითაც შესაძლებელი გახდება ყურადღების გამახვილება იმ დარგებზე რომლებსაც ბიუჯეტში ნაკლები შემოსავალი მოაქვთ. Data Mining ტექნოლოგიის საშუალებები გამოიწვევს სამუშაოს ოპტიმიზაციას და საგადასახადო სტრუქტურის მუშაობის ეფექტიანობას. საბოლოოდ შეიძლება ითქვას, რომ კვლევის აქტუალურობა გამოიხატება, ერთის მხრივ საქართველოს საგადასახადო სისტემის ინფორმაციის დამუშავების პროცესების ოპტიმიზაციაში და მონაცემების მოცულობის

პერმანენტული ზრდის ფონზე მუშაობის ეფექტურობის ამაღლებაში, ხოლო მეორეს მხრივ ქვეყნის ინფორმაციული ტექნოლოგიების თანამედროვე განვითარების პირობებში პროგრამული პროდუქტების და ინსტრუმენტების ადაპტირების აუცილებლობაში.

თემის დამუშავების ხარისხი. ამ დროისათვის Data Mining – თან დაკავშირებული საკითხები, სამამულო ლიტერატურაში არ არის გაშუქებული. არსებობს ცალკეული მცდელობები არსებული ტექნოლოგიის გამოყენების, განიხილება ამ ტექნოლოგიის ასპექტები სხვადასხვა სფეროებში. გაცილებით ფართოდ ხდება ამ ტექნოლოგიის გამოყენების და მისი არსის აღწერა უცხოურ ლიტერატურაში. მაგრამ აქაც კომპლექსური ანალიზი და ოპერაციები, რომლებიც ტარდება ამ ტექნოლოგიის გამოყენებისას, სამეცნიერო ლიტერატურაში ნაჩვენები არ არის.

უნდა აღინიშნოს, რომ არსებული ტექნოლოგია როგორც წესი წარმოდგენილია და აღწერილია შემდეგი სფეროებისათვის: საბანკო საქმე, საბითუმო ვაჭრობა, ტელეკომუნიკაცია, დაზღვევა, მედიცინა. კერძოდ, გამოვიკვლიე რუსეთის ფედერაციაში ტექნოლოგია DATA MINING-ის გამოყენება კერძო სტრუქტურაში, ისეთში როგორიც არის საბანკო საქმიანობა. არსებობს კომპანია BaseGroup Labs, რომელიც მუშაობს მონაცემების ანალიზის სისტემების შექმნაზე. გავრცელებულია კომპლექსური სკორინგული გადაწყვეტილებების მიღება სამომხმარებლო კრედიტების გასაცემად DATA MINING ტექნოლოგიის გამოყენების საფუძველზე. ამ კომპანიას შექმნილი აქვს პროგრამა, რომელსაც დაარქვეს Deductor:Loans. მას იყენებენ კრედიტების გაცემისას მათი დაუბრუნებლობის რისკების შესამცირებლად. იყენებენ DATA MINING ტექნოლოგიის გადაწყვეტილებების ხესა და ნეირონულ ქსელებს, რომლებიც აღწერილი იქნება ჩვენს ნამუშევარში. Deductor:Loans ინტეგრირებულია სხვადასხვა პროდუქტებთან.

ამ დროისათვის DATA MINING ტექნოლოგიის გამოყენება საგადასახადო სისტემისათვის არ არის გავრცელებული. ნამუშევარში აღწერილი მაქვს თუ როგორ გამოვიყენეთ ეს ტექნოლოგია საგადასახადო სტრუქტურაში და როგორ მოხდა მისი დანერგვა და

ინტეგრირება სხვა პროდუქტებთან. თემის კვლევასთან მიახლოებული არის შემდეგი ავტორების ნამუშევრები: Дюк В., Большаков П.С., Пройдаков Э., Смирнов Н., Бергер Ч. , Самойленко А., Киселев М., Соломатин Е., Han J., Newquist H. P., Parsaye K.

ამ პროდუქტის დამუშავებაზე მუშაობენ ისეთი დიდი კომპანიები, როგორებიც არიან ORACLE, MICROSOFT, IBM.

კვლევის საგანი და ობიექტი. კვლევის ობიექტს წარმოადგენს ტექნოლოგია Data Mining. ხოლო კვლევის საგანს წარმოადგენს გადამხდელთა პირადი აღრიცხვის ბარათი საქართველოს რესპუბლიკის შემოსავლების სამსახურის ინფორმაციული ტექნოლოგიების სამმართველოში არსებული მონაცემთა საცავების საფუძველზე.

კვლევის ამოცანები და მიზნები. სადისერტაციო ნამუშევრის მიზანია წარმოვადგინოთ, თუ როგორ შეიძლება Data Mining ტექნოლოგიის გამოყენება საგადასახადო სტრუქტურაში. კვლევის საერთო მიზნის წარმოსადგენად დისერტაციაში დასმულია შემდეგი ამოცანები:

- კომპლექსური შინაარსის ანალიზი, ინფორმაციულ ანალიტიკური სისტემის არქიტექტურა და მისი რეალიზაციისათვის არსებული პროდუქტები, ინსტრუმენტები და საფეხური რომელიც Data Mining ტექნოლოგიას უჭირავს ანალიტიკური სისტემების არქიტექტურაში;
- Data Mining ტექნოლოგიის განვითარების ისტორიული აღმოცენებისა და განვითარების პროცესი;
- Data Mining ტექნოლოგიის არსის და მასთან არსებული სხვა საკვანძო ტერმინოლოგიის განსაზღვრა;
- განხილვა საკითხის, თუ რა ადგილი უჭირავს ამ ტექნოლოგიის გამოყენებას ამა თუ იმ სფეროებში;
- შესწავლა, თუ რა სარგებელი მოუტანა ამ ტექნოლოგიის გამოყენებამ საბაზრო ეკონომიკას;
- სტატისტიკური მასალის გაერთიანება, Data Mining ტექნოლოგიის გამოყენებისას, კანონზომიერების გამოვლენა, ამ ტექნოლოგიის პერსპექტივების აღწერა;

- Data Mining ტექნოლოგიის გამოყენების პერსპექტიულობის აღწერა საქართველოს რესპუბლიკის საგადასახადო სისტემისათვის;
- Data Mining ტექნოლოგიის მეშვეობით წინააღმდეგობითობის მოდელების შექმნა და მისი გამოყენება პრაქტიკაში;

კვლევის მეთოდოლოგია და თეორიული საფუძველი. არსებული დისერტაცია ეფუძნება საგნის შესწავლის სისტემურ და დიალექტურ მიდგომებს და გამომდინარეობს იქიდან, რომ Data Mining ტექნოლოგია ერთ-ერთი წამყვანია ინტელექტუალურ ანალიზში, რის გარეშეც თითქმის შეუძლებელი გახდა მუშაობა დიდ კორპორაციებში. კვლევის ჩატარებისა და მოდელის აგებისას გამოყენებულია შემდეგი მეთოდები და მიდგომები: ანალიზი, სინთეზი, მოდელირება, ისტორიული მეთოდები.

სადისერტაციო კვლევაში აგრეთვე გამოყენებულია წამყვანი კომპანიების ფუნდამენტალური ანალიტიკური დამუშავებები და პროდუქტები. მათ წარმოგვიდგენს კომპანია: Oracle, Microsoft, IBM, SAS Institute, Ward Systems Group, Acknosoft, Integral Solutions, Megaputer და სხვა.

დისერტაციის მეცნიერულ სიახლეს წარმოადგენს ის, რომ ჩვენს მიერ შემუშავებული ალგორითმის მიხედვით, შესაძლებელი გახდება პროგნოზირების ჩატარება. გვეძლევა საშუალება ვივარაუდოთ, თუ რომელ საქმიანობის სახეს უნდა მიექცეს მეტი ყურადღება, თუ როგორ განაწილდება მონაცემები საგადასახადო ინსპექციებს შორის, ანუ თუ სად უფრო მოსალოდნელია საგადასახადო ჩარიცხვების ზრდა და ამის შედეგად, თუ რომელ გადასახადზე უნდა გამახვილდეს ყურადღება.

აგრეთვე კრიტერიუმების შერჩევის საფუძველზე ვღებულობთ პროგნოზს, რომლის შედეგადაც შეგვიძლია გამოვაკლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი.

დიდი ყურადღება უნდა მიექცეს მონაცემების მომზადებას. უნდა მოხდეს მათი გადარჩევა და გამოიყოს არასწორი მნიშვნელობები, ანუ შეიძლება ითქვას უნდა მოვახდინოთ ინფორმაციის “გაცრა”.

აგრეთვე სიახლეს წარმოადგენს ის, რომ Data Mining ტექნოლოგია პირველად გამოვიყენეთ ისეთ სპეციფიურ სფეროში, როგორიც საგადასახადო სისტემაა.

კვლევის რეზულტატების დანერგვა

სადისერტაციო ნაშრომში წარმოდგენილი კვლევის რეზულტატები დანერგილია საქართველოს შემოსავლების სამინისტროს საგადასახადო სისტემაში.

მოდელის შექმნისას წარმოიქმნა შემდეგი სიძნელე: მონაცემთა ბაზა Oracle9-ზე გართულდა ODM10-ის დაყენება, რის გამოც ჩვენ დაგვჭირდა სპეციალური აპლიკაციების დაყენება.

დისერტაციის სტრუქტურა განსაზღვრულია კვლევის ამოცანებისა და მიზნების, აგრეთვე საგნობრივი განსაკუთრებულობისა და ობიექტის დამუშავების გათვალისწინებით. ნამუშევარი შედგება შესავალის, სამი თავის, რეზიუმეს და ბიბლიოგრაფიული სიისაგან.

I თავი. თანამედროვე ინფორმაციულ-ანალიტიკური

სისტემის მიმოხილვა

თანამედროვე კომპანიები წლების განმავლობაში აგროვებენ საკმაოდ დიდი მოცულობის ინფორმაციას. წარმატებული მუშაობა დამოკიდებულია მასში არსებული, კუთვნილი ინფორმაციიდან მაქსიმუმის გამოყენებასთან. ეფექტური ინფორმაციულ-ანალიტიკური სისტემის შექმნა კი წარმატების საფუძველია.

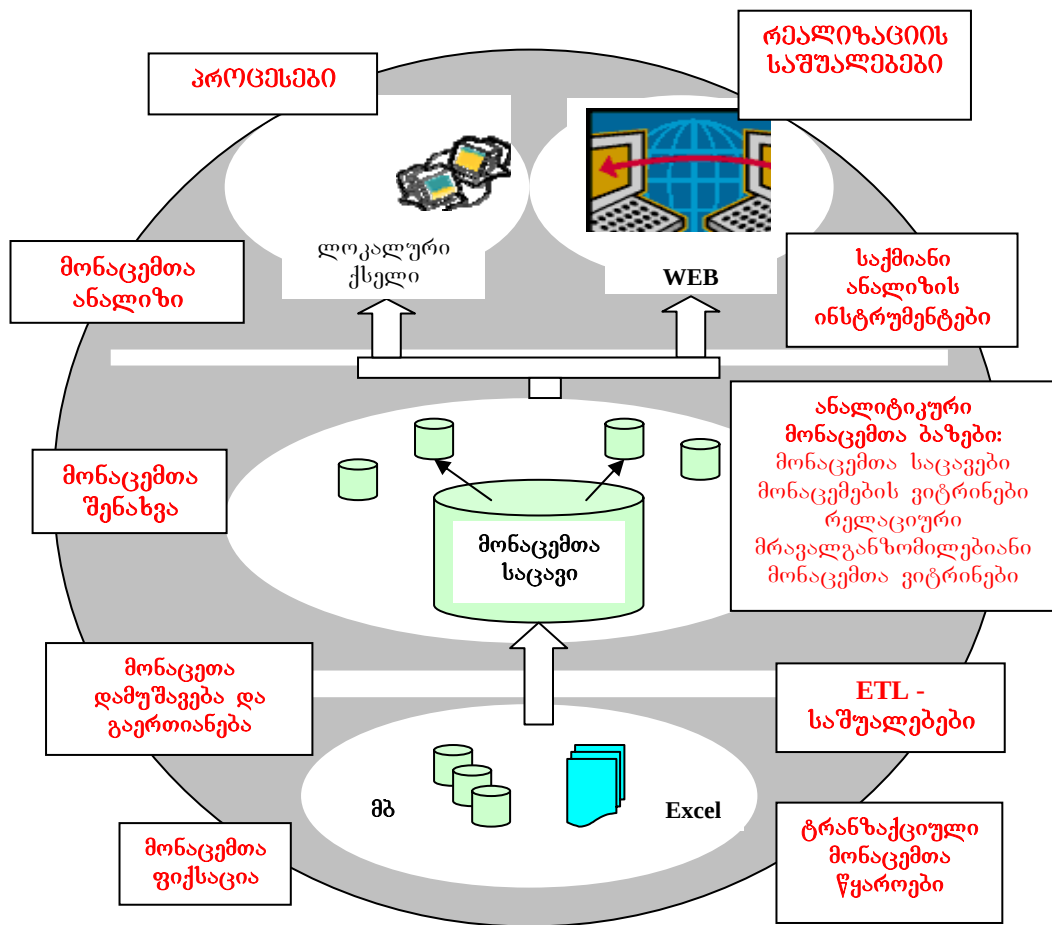
ინფორმაციულ-ანალიტიკური სისტემის გამოყენება წარმოებაში განპირობებულია შემდეგი ფაქტორებითა და მიზეზებით: ბიზნესპროცესების საერთო რეორგანიზაციის სწრაფვით, საქმიანი ინფორმაციის ხარისხის ამაღლების სურვილით, სტრატეგიული დაგეგმარების მხარდაჭერის საჭიროებით და მაღალეფექტური გადაწყვეტილებების მიღწევით.

1.1 თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურა

ნებისმიერი ინფორმაციულ-ანალიტიკური სისტემის მუშაობისათვის წარმოებს შემდეგი ამოცანების რეალიზაცია: მონაცემების ეფექტური შენახვა, დამუშავება და ანალიზი. ამ დროისათვის, ამ სფეროში დაგროვებულია საკმაოდ ცოდნა. ინფორმაციის შენახვა მიიღწევა ინფორმაციულ-ანალიტიკურ სისტემაში არსებული მთელი რიგი მონაცემების წყაროებიდან. ინფორმაციის დამუშავება და გაერთიანება მიიღწევა აღმოფხვრის ინსტრუმენტების გამოყენებით, და მონაცემების გარდაქმნითა და ჩატვირთვით. მონაცემების ანალიზი ხორციელდება თანამედროვე მონაცემების საქმიანი ანალიზის ინსტრუმენტების მეშვეობით.

ორგანიზაციის თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურა მთლიანობაში წარმოდგენილია ნახ.1 – ზე ¹.

¹ Галахов И. В., Волков И. Ю. Архитектура современной информационно-аналитической системы // Директор ИС №3, 2002год.



ნახ.1. თანამედროვე ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურა. მონაცემთა წყაროების მრავალფეროვნება და მათი გამოყენება ყოველ კონკრეტულ შემთხვევაში აიხსნება ინფორმაციის შენახვის სხვადასხვა საჭიროებით ანუ ორგანიზაციის წინაშე მდგომი ამოცანების მიხედვით (ამ ნახატის საფუძვლად არის აღებული შემდეგი სტატია: Галахов И. В., Волков И. Ю. Архитектура современной информационно-аналитической системы // Директор ИС №3, 2002год.)

მოყვანილი არქიტექტურა დემონსტრირებას უკეთებს იმ გრძელ გზას, რომელიც უნდა გაიაროს მონაცემებმა ანალიტიკოსთან მაგიდაზე მოხვედრამდე. მონაცემების წყაროს მრავალფეროვნება და მათი გამოყენების საჭიროება ყოველ კონკრეტულ შემთხვევაში აიხსნება ინფორმაციის სხვადასხვაგვარად შენახვის მოთხოვნით, რომელიც დამოკიდებულია ორგანიზაციის წინაშე მდგომი ამოცანებისაგან. თუ ვეცდებით მოვახდინოთ მონაცემების წყაროების კლასიფიცირება ტიპებისა და მნიშვნელობების მიხედვით, მაშინ ყოველივე მათგანი პირობითად შეგვიძლია მივაკუთვნოთ სამიდან ერთ-ერთ ჯგუფს: ტრანზაქციული მონაცემთა წყაროები, მონაცემთა საცავები, მონაცემთა ვიტრინები.

მონაცემების სისტემაში შეტანა შესაძლებელია, როგორც ხელით, ასევე ავტომატურადაც. პირველადი ფიქსაციის ეტაპზე მონაცემები მოდის აკრეფის სისტემისა და ინფორმაციის დამუშავებიდან ეგრეთწოდებულ ტრანზაქციულ მონაცემთა ბაზაში. ტრანზაქციული მონაცემთა ბაზები ორგანიზაციაში შესაძლოა იყოს რამოდენიმე.

რადგან მონაცემების ტრანზაქციული წყაროები, როგორც წესი არ არიან ერთმანეთთან შეთანხმებულნი, საჭიროა ასეთი მონაცემების ანალიზისათვის მათი გაერთიანება და გარდაქმნა. ამიტომ შემდგომ ეტაპზე წყდება მონაცემების კონსოლიდაციის ამოცანა, მათი გარდაქმნა და გაწმენდა, რის შედეგადაც მონაცემები ხდება ეგრეთწოდებულ მონაცემთა ანალიტიკურ ბაზებში. ანალიტიკური მონაცემთა ბაზები, საცავები, თუ ვიტრინები, წარმოადგენს იმ ძირითად წყაროებს, საიდანაც ანალიტიკოსი იღებს ინფორმაციას სათანადო საქმიანი ანალიზის ინსტრუმენტების მეშვეობით.

ინფორმაციულ-ანალიტიკურმა სისტემამ საშუალო და დიდი წარმოებისათვის თუ ორგანიზაციებისათვის უნდა უზრუნველყოს მომხმარებლებისათვის წვდომა ანალიტიკურ ინფორმაციასთან, რომელიც დაცულია არასანქცირირებული მოხმარებისაგან როგორც გახსნილ ორგანიზაციის შიდა ქსელში, ასევე ინტრანეტ და ინტერნეტ ქსელის მომხმარებლებისათვის. ამის შედეგად თანამედროვე ინფორმაციულ-ანალიტიკური სისტემა მოიცავს შემდეგ დონეებს:

- 1) მოგროვება და მონაცემების პირველადი დამუშავება;
- 2) აღმოფხვრა, გარდაქმნა და მონაცემების ჩატვირთვა;
- 3) მონაცემთა დაგროვება;
- 4) მონაცემთა წარმოდგენა მონაცემთა ვიტრინებში;
- 5) მონაცემთა ანალიზი;
- 6) web – პორტალი.

განვიხილოთ ჩამოთვლილი არქიტექტურის დონეები და გაგჩერდეთ ტიპური ინსტრუმენტების მაგალითებზე, რომლებიც შესაძლოა ყოველივე მათგანის შექმნისათვის.

იას არქიტექტურის პირველ დონეს მიეკუთვნება უკვე ზემოთ ხსენებული მონაცემთა წყაროები, ტრანზაქციულ და ოპერაციულ

მონაცემთა (ბაზებად) წყაროებად წოდებული, რომლებიც ეგრეთწოდებულ OLTP – სისტემის (online transactional processing²) ნაწილს წარმოადგენენ. ტრანზაქციული მონაცემთა ბაზები მოიცავენ მონაცემთა წყაროებს, რომლებიც ორიენტირებულნი არიან ორგანიზაციის ყოველდღიური საქმიანობის რეზულტატების ფიქსაციაზე. ტრანზაქციულ მონაცემთა ბაზებთან წარმოდგენილმა მოთხოვნებმა განაპირობეს შემდეგი განმასხვავებელი განსაკუთრებულობა: მონაცემთა სწრაფი დამუშავება და მათი ცვალებადობის მაღალი სიხშირის მხარდაჭერა, აგრეთვე ორიენტირება, როგორც წესი ერთ-ერთი პროცესის მომსახურებაზე და არა ორგანიზაციის საქმიანობაზე მთლიანობაში.³

აქ მაგალითად შეგვიძლია გამოვიყენოთ მონაცემთა ბაზები, რომლებიც გამოიყენება ფიჭური ქსელის ოპერატორებით ბილინგის სისტემაში, კომერციულ და სახელმწიფო ბანკების ავტომატიზირებულ სისტემებში, ინტერნეტ-მაღაზიებში.

ტრანზაქციული ბაზები შესანიშნავად ერევიან ყოველდღიური ინფორმაციის ზეგებას, რომლებიც როგორც წესი ყოველდღე უნდა გადამუშავდეს, მაგრამ არ იძლევიან საშუალებას ვნახოთ ორგანიზაციაში მიმდინარე საქმიანობის სურათი მთლიანობაში და იშვიათად გამოიყენებიან კომპლექსური ანალიზის ჩატარების წყაროებად.

ტრანზაქციული მონაცემების წყაროების ერთიანობა ქმნის ნებისმიერი ორგანიზაციის ინფრომაციულ-ანალიტიკური არქიტექტურის ქვედა რგოლს. შემდგომში ვიტყვით, რომ წარმოების იას იქმნება უკვე არსებული მოგროვების სისტემებისა და პირველადი მონაცემების დამუშავების საფუძველზე, რომლებიც შეიცავენ ტრანზაქციულ მონაცემთა წყაროებს.

მონაცემების აღმოფხვრა, გარდაქმნა და ჩატვირთვა ხორციელდება ეგრეთწოდებული ETL (extraction, transformation, loading

² Глоссарий // www.rdtex.ru/docs/glossary/O54162.html

³ OLAP, OLTP Приложения // www.datahousecorp.com/rus/technology/olap_oltp.htm

«извлечение, преобразование, загрузка»)⁴- ინსტრუმენტებით, რომლებიც წარმოდგენილია სხვადასხვა ტრანზაქციული ქვედა დონის წყაროების მონაცემთა აღმოფხვრისათვის, მათი გარდაქმნისათვის და კონსოლიდაციისათვის, და აგრეთვე მიზნობრივ მონაცემთა ანალიტიკურ ბაზებში ჩასატვირთად – მონაცემთა საცავებში და მონაცემთა ვიტრინებში. გარდაქმნის ეტაპზე ხდება მონაცემების გახშირება, წარმოებს აუცილებელი გამოთვლები და აგრეგირება. სამ საფეხურიანი აღმოფხვრის, გარდაქმნის და ჩატვირთვის პროცესი უნდა ხორციელდებოდეს დამტკიცებული რეგლამენტის საფუძველზე.

იას არქიტექტურის მესამე დონეს მიეკუთვნებიან მონაცემთა წყაროები, რომლებსაც ეუწოდებთ მონაცემთა საცავებს (ინგლ. Data Warehouse). მონაცემთა საცავები შეიცავენ მონაცემთა წყაროებს, რომლებიც ორიენტირებულნი არიან ინფორმაციის შენახვაზე და ანალიზზე. ასეთ წყაროებს შეუძლიათ გააერთიანონ რამოდენიმე ტრანზაქციული სისტემის ინფორმაცია და იძლევიან საშუალებას ჩავატაროთ მონაცემთა ანალიზი თანამედროვე საქმიანი ანალიზის ინსტრუმენტებთან გამოყენების კომპლექსში.

დღესდღეობით, მონაცემთა დამუშავების სისტემები რომლებიც რეალიზებულნი არიან სხვადასხვა საფუძველზე საკმაოდ გამართულად მუშაობენ და საკმაოდ სრულად ქმნიან და ავსებენ უზარმაზარ მრავალტომეულ ელექტრონულ არქივებს. ასეთი სისტემების ძირითადი დანიშნულებაა ოპერატიული გადამუშავება და მათ არ აქვთ საშუალება შეინახონ მონაცემები რამოდენიმე თვეზე მეტი დროით. როდესაც მონაცემები მოძველდება, ისინი გამოიზიდება და გაიწმინდება ოპერაციული მონაცემთა ბაზიდან.

რადგან ნებისმიერ ორგანიზაციაში ფუნქციონირებს რამოდენიმე განსხვავებული ან მერთალად დაკავშირებული ომბი, მათგან გამოტანილ მონაცემებს, როგორც წესი გააჩნია სხვადასხვა სტუქტურა, ფორმატი, თარიღების და ფულადი ერთეულების წარმოდგენის

⁴ С. Чаудхури, У. Дайал, В.Ганти Технология баз данных в системах поддержки принятия решений 22 января 2002г. //статья с сайта www.citforum.ru

სტანდარტები. ერთი და იგივე ობიექტების შესაფასებლად გამოიყენება სხვადასხვა კოდირებები.

შედგად, უზარმაზარი არქივული მასივები, რომლებიც მოგროვებულია წლების განმავლობაში მონაცემთა დამუშავების სისტემებით და რომლებიც შეიცავენ მრავალფეროვან და საჭირო ინფორმაციას ორგანიზაციისათვის, რჩებიან ხელშეუხებელნი. წინასწარ დამუშავებისა და შეთანხმების გარეშე არქივული მონაცემები არასრულფასოვანია და შეუძლებელია მათი გამოყენება ანალიზის ამოცანებში.

მონაცემები, რომლებიც არის აღმოჩენილი მონაცემთა დამუშავების სისტემების მიერ – წარმოადგენს სრულფასოვანი ბიზნეს-გადაწყვეტილებების ინფორმაციის მხოლოდ ნაწილს. ორგანიზაცია არსებობს და ფუნქციონირებს რეალურ სამყაროში. ანალიტიკურ მონაცემთა სისტემაში რეგიონების ეკონომიკის, საკანონმდებლო ბაზების ჩართვა სხვადასხვა ელექტრონულ სტატიკური კრებულებიდან (როგორც ყველასათვის ხელმისაწვდომი, ისევე კომერციული) იძლევა საშუალებას ახლებურად შევხედოთ მრავალ კანონზომიერებებს რომლებიც აღმოჩენილია შიდა მონაცემების ანალიზის პროცესში. როგორც გვიჩვენებს პრაქტიკა, ნებისმიერი გადაწყვეტილება, მიღებული შიდა მონაცემების საფუძველზე, აღმოჩნდება არაკორექტული.⁵

DW კონცეფცია წარმოდგენილ იქნა 1990 წელს ბ. ინმონის მიერ და გახდა ერთ-ერთი მოწინავე ინფორმაციულ ტექნოლოგიებში 90-იან წლებში, კერძოდ მონაცემთა დამუშავებაში. ამ კონცეფციის წარმოდგენა იყო იმ ფაქტის არასრული გააზრება, რომ არსებობს ორი ძირითადი ფუნქციონალურად განსხვავებული ინფორმაციის დამუშავების სისტემათა კლასი. პირველის არსებობის ბაზას წარმოადგენს ტრანზაქციების ნაკადის დამუშავება და წარმოადგენს ინფორმაციის ნაწყვეტს, რომელიც მოიცავს მიმდინარე ან რაღაც მცირე დროის პერიოდის ნაწყვეტს. მეორე დაფუძნებულია საჭირო ინფორმაციის შეგროვებაზე დროის დიდი პერიოდის განმავლობაში (5 წლიდან), რომელიც განკუთვნილია მონაცემთა ანალიზისათვის. DW კონცეფციის

⁵ Inmon W. H. Building the Data Warehouse. New York: John Wiley & Sons, Inc.

განვითარებამ ამ ორი ტიპის სისტემების საზღვრებს შორის ხაზის გასმის საშუალება მოგვცა. ავტორმა შემდეგნაირად ახსნა მონაცემთა საცაგები: “საგნობრივ-ორიენტირებული, ინტეგრირებული, გარდაუქმნელი, მონაცემთა ნაკრები, რომლებიც მოიცავენ ქრონოლოგიას და რომლებიც არიან ორგანიზებულნი მართვის მიზნებისათვის” და მოწოდებულნი არიან “ერთიანი და მხოლოდ ერთადერთი ჭეშმარიტების წყაროდ”, რომლებიც იძლევიან ანალიტიკოსებისათვის ჭეშმარიტ ინფორმაციას ოპერეტიული ანალიზისა და გადაწყვეტილებების მისაღებად.⁶

მონაცემთა საცაგების კონცეფციის საფუძვლად დევს ორი ძირითადი იდეა:

- წინასწარ დეტალიზირებული და განაწილებული მონაცემების ინტეგრაცია;
 - o ისტორიული არქივები;
 - o მონაცემები ტრადიციული მონაცემთა დამუშავების სისტემებიდან;
 - o მონაცემები გარე წყაროებიდან ერთიან მონაცემთა საცაგებში, მათი შეთანხმება და შესაძლო აგრეგაცია.
- მონაცემთა ნაკრების დაყოფა ოპერაციულ დამუშავებისათვის და მონაცემთა ნაკრებები, რომლებიც გამოიყენება ანალიზის ამოცანების გადასაწყვეტად; არსებობს ძირითადი მოთხოვნები მონაცემთა საცაგებისათვის⁷:

- მაღალი სიჩქარის მხარდაჭერა საცავიდან მონაცემების მისაღებად;

- მონაცემების შიდა ერთმანეთთან შეთანხმებულობის მხარდაჭერა;

⁶ Сахаров А.А Концепции построения и реализации информационных систем, ориентированных на анализ данных // статья опубликована на www.olap.ru

⁷ А. Федоров, Н. Елманова Введение в OLAP: часть 1. Основы OLAP// КомпьютерПресс 4' 2001

- ეგრეთწოდებული მონაცემების ჩამონატრების (slice and dice) მიღებისა და შედარების საშუალება;
- მოსახერხებელი დათვალიერების საშუალებების არსებობა მონაცემთა საცავებში;
- შენახული მონაცემების სრულყოფა და სისრულე;
- მონაცემთა შევსების პროცესის სისრულის მხარდაჭერა.

ყველა ზემოთ აღნიშნული მოთხოვნების შესრულება ერთი და იგივე პროდუქტის საშუალებით ხშირად შეუძლებელია. ამრიგად, მონაცემთა საცავების რეალიზაციისათვის ხშირად გამოიყენება რამოდენიმე პროდუქტი, საიდანაც ზოგიერთი მათ შორის წარმოადგენს მონაცემთა შენახვის საშუალებებს, სხვები მათი აღმოღვრისა და დათვალიერების, აგრეთვე საშუალებებს მათი შემატებისათვის და ა.შ.

ტიპიური მონაცემთა საცავები, როგორც წესი განსხვავებულია ჩვეულებრივი რელაციური მონაცემთა ბაზისაგან. ჩვეულებრივი მონაცემთა ბაზები განკუთვნილია, იმისათვის, რომ გაუადვილოს მომხმარებელს ყოველდღიური მუშაობა, მონაცემთა საცავები კი წარმოდგენილია გადაწყვეტილებების მისაღებად. მაგალითად, საქონლის გაყიდვა და ანგარიშის გამოწერა ხდება მონაცემთა ბაზების მეშვეობით, რომელიც განკუთვნილია ტრანზაქციების დასამუშავებლად, ხოლო გაყიდვების დინამიკის ანალიზი რამოდენიმე წლის განმავლობაში, რომელიც იძლევა საშუალებას დავგეგმოთ მუშაობა მომწოდებლებთან – მონაცემთა საცავების მეშვეობით. ჩვეულებრივი მონაცემთა ბაზები მომხმარებლების მუშაობის პროცესში ექვემდებარება ხშირ ცვლილებებს, ხოლო მონაცემთა საცავები გაცილებით სტაბილურია. მონაცემების განახლება მათში ხდება განრიგის მიხედვით (მაგალითად, ყოველ კვირა, ყოველდღიურად ანდა საათში ერთხელ – მოთხოვნების მიხედვით). იდეალის მიხედვით პროცესი წარმოადგენს ახალი მონაცემების დამატებას რაღაც პერიოდის განმავლობაში, უკვე არსებულ საცავში ინფორმაციის ცვლილების გარეშე. ამას გარდა, საცავი შესაძლოა შეივსოს გარე წყაროების მეშვეობით, მაგალითად სტატისტიკური რეპორტები. გადაწყვეტილებების მიღების სისტემებს ხშირად გააჩნია მომხმარებლისათვის აგრეგატული მონაცემების წარმოდგენის

საშუალებები არსებული ნაკრებიდან არჩევისათვის, იოლი აღქმისათვის და ანალიზისათვის. როგორც წესი, ასეთი აგრეგატული ფუნქციები ქმნიან მრავალ მრავალგანზომილებიან ანუ არარელაციურ მონაცემთა ნაკრებს (მას ხშირად უწოდებენ ჰიპერკუბს ან/და მეტაკუბს), რომლის დერძები შეიცავენ პარამეტრებს, ხოლო უჯრედები – მათგან დამოკიდებულ აგრეგატულ მონაცემებს და თანაც ასეთი მონაცემები შესაძლოა ინახებოდნენ რელაციურ ცხრილებში, მაგრამ ამ შემთხვევაში ჩვენ ესაუბრობთ მონაცემების ლოგიკურ ორგანიზაციაზე, არათუ მათი შენახვის ფიზიკურ რეალიზაციაზე). ყოველი დერძის გასწვრის მონაცემები შესაძლოა ორგანიზებულნი იქნენ იერარქიულად, რომლებიც წარმოადგენენ სხვადასხვა დეტალიზაციის დონეებს. ასეთი მონაცემთა მოდელის მეშვეობით მომხმარებლებს შეუძლიათ რთული წვდომის ფორმულირება, რეპორტების გენერირება, მონაცემთა ქვესიმრავლეების მიღება.

რადგან მონაცემთა საცავები წარმოადგენს ინფორმაციულ-ანალიტიკური სისტემის ერთ-ერთ ძირითად რგოლს ნებისმიერი საშუალო თუ დიდი ორგანიზაციისათვის, ამის გამო იგი გამოდის როგორც ძირითადი მონაცემთა წყარო მრავალფეროვანი არსებული ორგანიზაციაში ანალიზისათვის.

ინფორმაციულ-ანალიტიკური სისტემის მეოთხე დონეს წარმოადგენს მონაცემთა წყაროები, რომელთაც უწოდებენ მონაცემთა ვიტრინებს (data marts⁸) და რომლებიც წარმოადგენილნი არიან გამიზნული საქმიანი ანალიზისათვის. მონაცემთა ვიტრინები როგორც წესი იქმნებიან მონაცემთა საცავის ინფორმაციის საფუძველზე, მაგრამ მათი ფორმირება ხდება ტრანზაქციული სისტემებიდან აღებული მონაცემების საფუძველზე, როდესაც ორგანიზაციაში რაღაც მიზეზების გამო არ არის რეალიზებული მონაცემთა საცავები.

კონცეფცია მონაცემთა ვიტრინები (Data Mart) იყო შემოთავაზებული Forrester Research – ის მიერ 1991 წელს. ავტორების მიერ, მონაცემთა ვიტრინები წარმოადგენს თემატიკურ მონაცემთა ბაზების სიმრავლეს, რომლებიც ეხება ორგანიზაციის საქმიანობის

⁸ Глоссарий от BaseGroup Labs [//www.glossary.basegroup.ru/v/datamart.htm](http://www.glossary.basegroup.ru/v/datamart.htm)

კერძო ასპექტებს და რომლებიც უნდა ყოფილიყვნენ IBM ფირმის ინფორმაციული საცავების (Information Warehouse) ალტერნატივები.

მონაცემთა ვიტრინების კონცეფციას გააჩნია რიგი ღირსებები:

- ანალიტიკოსები მუშაობენ მხოლოდ იმ მონაცემებთან, რომლებიც მათ რეალურად სჭირდება;
- მონაცემთა ვიტრინების მიზნობრივი მონაცემთა ბაზა მაქსიმალურად არის მიახლოებული საბოლოო მომხმარებელთან;
- მონაცემთა ვიტრინები ხშირად მოიცავენ ადრე აგრეგირებული მონაცემების თემატიკურ ქვესიმრავლეებს, მათი პროექტირება და დამუშავება უფრო ადვილია.

მონაცემთა ვიტრინები არ საჭიროებს მძლავრ გამოთვლით ტექნიკას. ამ კონცეფციას აგრეთვე გააჩნია ძალიან სერიოზული ჩავარდნები. აქ წარმოიდგინება ტერიტორიულად განაწილებული ინფორმაციული სისტემის რეალიზაცია არც თუ ისე მძლავრ კონტროლირებადი სიჭარბით, მაგრამ არ წარმოიდგინება წესები, თუ როგორ უზრუნველყოთ მასში შენახული მონაცემების ერთიანობა და წინააღმდეგობა.

მონაცემთა ვიტრინებისა და მონაცემთა საცავების კონცეფციის გაერთიანების იდეა ეკუთვნის მ. დემარესტს (M.Demarest), რომელმაც 1994 წელს თავის ნამუშევარში “მონაცემთა ვიტრინების შექმნა”, შემოიტანა ორი კონცეფციის გაერთიანების წინადადება და აგრეთვე მონაცემთა საცავებისათვის ერთიანი ინტეგრირებული მონაცემთა წყაროების გამოყენების წინადადება.⁹

დღესდღეობით არსებობს ზუსტად ასეთი მრავალდონიანი გადაწყვეტილებები:

- პირველი დონე – საერთო კორპორატიული მონაცემთა ბაზა რელაციური მონაცემთა სისტემის მართვის საფუძველზე ნორმალიზებული და მკრთლად დენორმალიზირებული სქემით (დეტალიზირებული მონაცემები);
- მეორე დონე – დაყოფის დონის მონაცემთა ბაზა მართვის სისტემის საფუძველზე რეალიზებული (აგრეგირებული მონაცემები);

⁹ Demarest M. Building the Data Mart // DBMS. – 1994. – July

- მესამე დონე – საბოლოო მომხმარებელთა სამუშაო ადგილები, რომლებზეც დაყენებულია ინსტრუმენტები ანალიზისათვის; დროთა განმავლობაში ეს სტანდარტი ხდება ძირითადი, რომელიც იძლევა საშუალებას მოვახდინოთ რეალიზაცია და გამოვიყენოთ ამ მიდგომებიდან ყოველივე:

- დეტალიზირებული მონაცემების კომპაქტური შენახვა და დიდი მონაცემთა ბაზების მხარდაჭერა, რომლებიც რეალიზებულნი არიან რელაციური მონაცემთა ბაზების მართვის სისტემით;
- დაყენების სიმარტივე და პასუხის გაცემის მოკლე ვადები, აგრეგირებულ მონაცემებთან მუშაობაში, რომლებიც უზრუნველყოფილნი არიან მბმს (მონაცემთა ბაზის მართვის სისტემები) - ით.

მონაცემთა რელაციური წარმოდგენის ფორმა, რომელიც გამოიყენება ცენტრალურ საერთო კორპორატიულ მონაცემთა ბაზაში უზრუნველყოფს მონაცემთა შენახვის უფრო კომპაქტურ წესს. ხოლო თანამედროვე რელაციურ მონაცემთა ბაზების მართვის სისტემა კი უკვე მუშაობს ტერაბაიტულ ბაზებთან. მიუხედავად იმისა, რომ ასეთ ცენტრალურ სისტემას არ შეუძლია უზრუნველყოს ანალიტიკური წვდომის ოპერატიული დამუშავების რეჟიმი მონაცემთა შენახვის ინდექსაციის ახლებური წესების გამოყენებით და აგრეთვე რიგი ცხრილების დენორმალიზაცია, რეგლამენტირებული წვდომების დამუშავების დრო მაინც დამაკმაყოფილებელია.

ამავე დროს მბმს-ის გამოყენება უზრუნველყოფს მომხმარებელს ქვედა დონის კვანძებში დამუშავებისა და პასუხის გაცემის მინიმალურ დროს არარეგლამენტირებულ წვდომებზე. ამის გარდა, ზოგიერთ მბმს - ში არსებობს მონაცემთა შენახვის საშუალება როგორც მუდმივ საფუძველზე, ასევე დინამიკურად მონაცემების ჩატვირთვაზე რელაციურ მონაცემთა ბაზიდან.

ასეთი სახით, გვეძლევა საშუალება მუდმივ საფუძველზე შევინახოთ მხოლოდ ის მონაცემები, რომლებიც ხშირად გამოიძახებიან მოცემულ კვანძში, ხოლო ყველა დანარჩენისათვის ინახება მხოლოდ ცენტრალური ბაზიდან აღმოფხვრის პროგრამების სტრუქტურა და აღწერა. ასეთი პირველადი მიმართვის პასუხის გაცემის დრო

ვირტუალურ მონაცემებთან შესაძლოა საკმაოდ ხანგრძლივი აღმოჩნდეს, სამაგიეროდ უზრუნველყოფს მაღალ მოქნილობას და იყენებს არც თუ ისე ძვირფას აპარატურულ საშუალებებს.

ინფორმაციის შენახვის ტიპების მიხედვით ვიტრინები იყოფიან რელაციურ და მრავალგანზომილებიანად. პირველი ტიპის ვიტრინები ორგანიზებულნი არიან რელაციური მონაცემთა ბაზის სახით “ვარსკლავი” სქემით, სადაც ცენტრალური ცხრილი, ფაქტების ცხრილი, რომელიც განკუთვნილია რაოდენობრივი ინფორმაციის შესანახად, დაკავშირებულია ცნობარ-ცხრილებთან.

კომპლექსური მრავალგანზომილებიან მონაცემთა ანალიზს უწოდებენ OLAP(Online Analytical Processing¹⁰). OLAP- მონაცემთა საცავის ძირითადი კონცეფციაა.

OLAP კონცეფცია 1993 წელს აღწერა ედვარ კოდმა. კოდმა შემოიტანა ეგრეთ წოდებული FASMI ტესტის ანალიზი, რომელიც მოიცავს მრავალგანზომილებიანი ანალიზისათვის შემდეგ მოთხოვნებს.

- მომხმარებლისათვის ანალიზის რეზულტატის მიწოდება მისაღებ დროში (ხშირად არაუმეტეს 5 წამის), მიუხედავად ნაკლებად დეტალური ანალიზისა;
- ნებისმიერი ლოგიკური თუ სტატიკური ანალიზის ჩატარების საშუალება და მისი შენახვა საბოლოო მომხმარებლისათვის გასაგებ ფორმაში;
- მრავალმომხმარებლივი წვდომა მონაცემებთან მასში არსებული ბლოკირების მექანიზმების გამოყენებით და წვდომის ავტორიზირებული საშუალებებით;
- მრავალგანზომილებიანი კონცეპტუალური მონაცემების წარმოდგენა, რომელიც მოიცავს იერარქიულ და იერარქიული ქვესიმრავლეების მხარდაჭერას (ეს OLAP კონცეფციის ძირითადი მოთხოვნაა);
- ნებისმიერ საჭირო ინფორმაციასთან დაკავშირება მისი მოცულობისა და შენახვის ადგილმდებარეობის მიუხედავად.

¹⁰ Глоссарий://www.rdtex.ru/docs/glossary/O54153.html

მრავალგანზომილებიანი ვიტრინები ორგანიზებულნი არიან როგორც OLAP მრავალგანზომილებიანი მონაცემთა ბაზები, სადაც ცნობარი ინფორმაცია წარმოიდგინება განზომილებების სახით. ინფორმაცია მრავალგანზომილებიან მონაცემთა ვიტრინებში ბიზნეს ტერმინებით წარმოიდგინება საბოლოო მომხმარებლისათვის მაქსიმალურად მისაღები სახით, რაც იძლევა დროის შემცირების საშუალებას ინფორმაციის გადაწყვეტილებების მისაღებად.

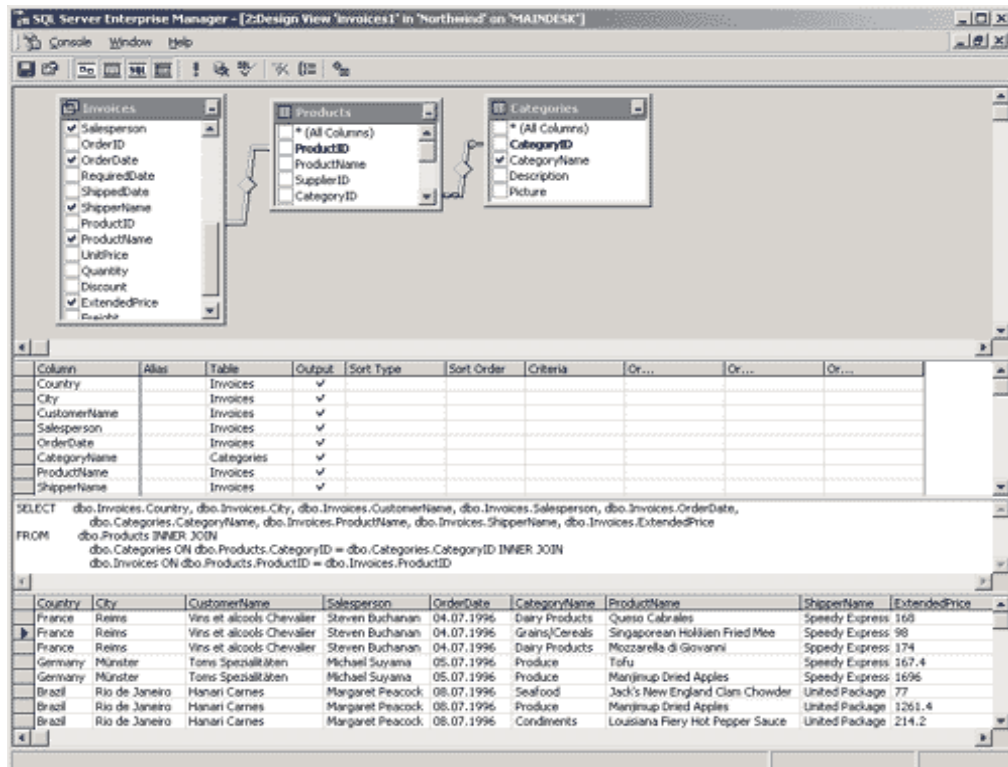
მომხმარებლის თვალსაზრისით მონაცემთა ვიტრინების განსხვავება მონაცემთა საცავებისგან მდგომარეობს იმაში, რომ მონაცემთა საცავები ეთანხმება მთელი ორგანიზაციის დონეს, ხოლო ყოველი ვიტრინა ემსახურება დონეს არაუმეტეს ერთი ქვედანაყოფისა და ზოგჯერ შესაძლოა შეიქმნას ინდივიდუალური მოხმარებისათვის, რომელიც განსხვავდება საკმაოდ ვიწრო მიზნობრივი სპეციალიზაციით.

მონაცემთა ვიტრინების განსხვავება ტრანზაქციული ბაზებისაგან მდგომარეობს შემდეგში: მონაცემთა ვიტრინები გამოიყენება საბოლოო მომხმარებლების მოთხოვნილებების დასაკმაყოფილებლად ანუ ანალიტიკოსების, სხვადასხვა დონის მენეჯერების მიერ, რომლებიც წყვეტენ სხვადასხვაგვარ ბიზნეს ამოცანებს. ტრანზაქციული მონაცემთა ბაზები კი გამოიყენება ძირითადად ოპერატორებით, რომლებიც პასუხს აგებენ ინფორმაციის შეტანაზე და პირველადი ინფორმაციის დამუშავებაზე, არა თუ მის ანალიზზე, რომელიც გამიზნულია გადაწყვეტილებების მიღებაზე.

მონაცემთა ვიტრინების გამოყენება, მრავალგანზომილებიანის თუ რელაციურის, თანამედროვე საქმიანი ანალიზის ინსტრუმენტებთან ერთად იძლევა საშუალებას მონაცემები გარდაქმნას საჭირო ინფორმაციად, რომლის საფუძველზეც შესაძლოა მიიღო ეფექტური გადაწყვეტილებები.

უნდა აღინიშნოს, რომ OLAP- ფუნქციონალურობა შესაძლოა რეალიზებულ იქნეს სხვადასხვა საშუალებებით; მარტივი მონაცემთა ანალიზის საშუალებებით, დაწყებული საოფისე აპლიკაციებით და საბოლოოდ განაწილებული ანალიტიკური სისტემებით სერვერულ პროდუქტებზე დაფუძნებით. მაგრამ ამ ფუნქციონალურობის სხვადასხვაგვარ რეალიზაციებზე საუბრამდე, მოდით განვიხილოთ რას

წარმოადგენს OLAP კუბები ლოგიკური თვალსაზრისით. შემდგომში ჩვენ განვიხილავთ კონცეფცია OLAP-ს და მრავალგანზომილებიან კუბებს უფრო ღრმად. რელაციური მონაცემთა ბაზის მაგალითად, რომელსაც ჩვენ გამოვიყენებთ OLAP პრინციპის ილუსტრირებისათვის, გამოვიყენოთ მონაცემთა ბაზა (ნახ. 2).



ნახ.2 მოთხოვნათა წარმოდგენა¹¹

მონაცემთა საცავეების კონცეფციის საფუძვლად დევს ორი იდეა – დეტალიზირებული დაყოფილი მონაცემების ინტეგრაცია (დეტალიზირებული, ანუ რომლებიც არ წერენ ზოგიერთ კონკრეტულ ფაქტებს, თვისებებს და ა.შ.) ერთიან საცავეებში და მონაცემთა ნაკრების და დანართების დაყოფა, რომლებიც გამოიყენება ოპერატიული დამუშავებისათვის და ამოცანების გადაწყვეტილების მიღების გამოსაყენებლად.

¹¹ SQL Server Enterprise Manager-ში ნაჩვენები რელაციური მონაცემთა ბაზა, სკრიპტი და ცხრილები გადმოტანილი პრინტსკრინით

მარეგისტრირებელი ინფორმაციული სისტემების ზრდის პერიოდში გამოვლინდა მათი შესაძლებლობების გამოყენების შეზღუდულობა მონაცემთა ანალიზის მიზნებისათვის და მათ საფუძველზე მხარდაჭერის სისტემების შექმნა და გადაწყვეტილებების მიღება. მარეგისტრირებელი სისტემები იქმნებოდა ბიზნესის წარმოების ოპერაციების ავტომატიზაციისათვის – ანგარიშების გამოწერა, შეთანხმებების გაფორმება, საცავის მდგომარეობის შემოწმება და ა.შ. ასეთი სისტემების მომხმარებლად ითვლება ხაზობრივი პერსონალი. ასეთი სისტემების მიმართ ძირითადი მოთხოვნები იყო ცვლილებების შეტანის ტრანზაქციულობის უზრუნველყოფა და შესრულების სიჩქარის მაქსიმიზაცია. ამ მოთხოვნებმა განსაზღვრა რელაციური მონაცემთა ბაზების დამუშავების სისტემების არჩევანი და მონაცემთა წარმოდგენის მოდელი “არსებობა-კავშირი” ძირითადი ტექნიკური გადაწყვეტილებების საფუძველზე მარეგისტრირებელი სისტემების შექმნისას. ხოლო მენეჯერებისათვის და ანალიტიკოსებისათვის თავის მხრივ საჭირო იყო სისტემები, რომლებიც მისცემდათ საშუალებას:

- ინფორმაციის ანალიზი დროის ასპექტში;
- სისტემისადმი მოთხოვნების ფორმირება;
- დიდი მოცულობის ინფორმაციის დამუშავება;
- მონაცემების ინტეგრირება სხვადასხვა მარეგისტრირებელი სისტემებისათვის.

ნათელია, რომ მარეგისტრირებელი სისტემები არ აკმაყოფილებდნენ ზემოთ აღნიშნულ არცერთ მოთხოვნას. მარეგისტრირებელ სისტემაში ინფორმაცია აქტუალურია მხოლოდ მონაცემთა ბაზებთან დაკავშირებისას, შემდგომ მომენტში იგივე მოთხოვნაზე თქვენ შეიძლება მიიღოთ სულ სხვა პასუხი. მარეგისტრირებელი სისტემების ინტერფეისი გათვლილია მკაცრად განსაზღვრული ოპერაციების ჩატარებაზე და რეზულტატების მიღება არარეგლამენტირებულ მოთხოვნებზე (ad-hoc) მკვეთრად შეზღუდულია. დიდი მონაცემთა მასივების დამუშავების შესაძლებლობა შეზღუდულია ასევე მონაცემთა ბაზების მართვის სისტემის მიმართვის გამო მოკლე ტრანზაქციებზე, რაც იწვევს სხვა მომხმარებელთა მუშაობის შენელებას.

წარმოქმნილი მოთხოვნების პასუხად გამოვლინდა ახალი ტექნოლოგია მონაცემთა ბაზების ორგანიზაციისათვის – ზემოთ ხსენებული მონაცემთა საცავების ტექნოლოგია. მონაცემთა საცავები წარმოადგენს შემდეგ არქიტექტურას: ვირტუალური საცავი, ორდონიანი, სამდონიანი.

მონაცემთა აღმოფხვრა საცავებიდან შესაძლოა SQL ენაზე დაწერილი მოთხოვნების მეშვეობით:

მაგალითად¹²:

SELECT

Invoices.ins, Invoices.said, Invoices.iden, Invoices.name,
Invoices.entry_Date,

Categories.oper_Id, Invoices.payed, Invoices.assessed, Invoices.returned

FROM Products INNER JOIN

Categories ON Products. iden = Categories.iden INNER JOIN

Invoices ON Products.said = Invoices.said¹³

ინფორმაციულ ანალიტიკური სისტემის შემდეგ არქიტექტურის დონეს მიეკუთვნება თანამედროვე პროგრამული საშუალებები, რომლებსაც ეწოდებათ ინტელექტუალური და საქმიანი მონაცემთა ანალიზის ინსტრუმენტები (Business Intelligence Tools), BI-ინსტრუმენტები.¹⁴

BI-ინსტრუმენტები ორგანიზაციის მართვის რგოლს აძლევს ყოველგვარი ინფორმაციის ანალიზის საშუალებას, აგრეთვე გვიადვილებს ორიენტაციის საშუალებას დიდი მოცულობის

¹² გამოყენებულია საქართველოს რესპუბლიკის საგადასახადო სტრუქტურის ბაზა

¹³ Абулашвили Х., Маградзе Е.: Построение и реализация Хранилищ Данных,-“INTELECT”/ - 2006| №1(24)

¹⁴ R. Kimball. The Data Warehouse Toolkit. Practical Techniques for Building Dimansional Data Warehouses John Wiley & Sons; Bk&CD Rom edition February 2, 1996.

მონაცემებში, ინფორმაციის ანალიზირების საშუალებას, პროგნოზების შექმნის საშუალებას, რომლის მეშვეობითაც ხდება არასწორი გადაწყვეტილებების რისკების შემცირება მინიმუმამდე.

მონაცემთა ინტელექტუალური ანალიზის ინსტრუმენტებს იყენებენ საბოლოო მომხმარებლები მონაცემთან წვდომის, მისი ვიზუალიზაციისა და მრავლგანზომილებიანი ანალიზისათვის და როგორც განსაზღვრული ფორმისა და შემცველობის, ასევე განუსაზღვრელი რეპორტების ფორმირებისათვის, რომლებიც იქმნება მმართველის ან ანალიტიკოსის მიერ (პროგრამისტის გარეშე). როგორც უკვე ვთქვით, შემაჯალ ინფორმაციას საქმიანი ანალიზისთვის წარმოადგენს დამუშავებული მონაცემები საცავეებიდან ან/და მონაცემთა ვიტრინებში წარმოდგენილი, არა თუ დამუშავებული მონაცემები.

კონცეფცია, მეთოდები და მონაცემთა საცავეების საშუალებები (Data warehousing) განსაზღვრავენ მიდგომებსა და ინტეგრაციას, გაწმენდას, რეტროსპექტიულ ინფორმაციის შენახვას, რომელიც წარმოდგენილია ანალიზისათვის და პასუხობს შემდეგ კითხვაზე: “როგორ მოვამზადოთ ინფორმაცია ანალიზისათვის?”. ბიზნეს-ინტელექტის ტექნოლოგია განსაზღვრავს მეთოდებს, ოპერატიული ინფორმაციის ანალიზსა და წვდომის საშუალებებს საგნობრივი ტერმინების სფეროში. BI-საშუალებები არ მოითხოვს მონაცემთა საცავეების ინფრასტრუქტურაში მუშაობის აუცილებლობას, მაგრამ ამ შემთხვევაში გაწმენდის და მონაცემების შეთანხმების პრობლემა მათ ევალებათ, და რაც მთავარია ამ ოპერაციების ჩატარება მათ მოუხდებათ მომენტალურად ან/და წინასწარ დამუშავების საშუალებით, მაგრამ განზოგადოებული რესურსისათვის. ამის გარდა, არსებობს ზეგავლენის ეფექტი ნაყოფიერებაზე და ტრანზაქციების ოპერატიული სისტემის დამუშავების სიმყარეზე. აი რატომ ითვლება საკმაოდ კარგ პრაქტიკად ტრანზაქციული და ანალიტიკური შემადგენლების გამოყოფა და მეორისათვის მონაცემთა საცავეების სხვადასხვა გადაწყვეტილებების გამოყენება. ძირითადი კვანძები მოდის არა მარტო ინფორმაციის დონეზე არამედ მეტამონაცემების დონეზეც. მონაცემთა საცავეების შემთხვევაში შესაძლოა ცენტრალიზებული მეტამონაცემების მართვა.

უნდა აღინიშნოს, რომ ტერმინი “მონაცემთა საცავი” ხშირად

გამოიყენება DSS გადაწყვეტილების მიღების მხარდამჭერი სისტემის ან/და ინფორმაციულ-ანალიტიკურ სისტემის აღსანიშნად, რომლებიც დაფუძნებული არიან მონაცემთა საცავებისა და ბიზნეს-ინტელექტის ტექნოლოგიებზე.¹⁵¹⁶

დღესდღეობით BI- პროდუქტების კატეგორიები მოიცავენ: BI- ინსტრუმენტებსა და BI- დანართებს. BI- ინსტრუმენტები იყოფიან: მოთხოვნისა და რეპორტების გენერატორებად; განვითარებულ BI- ინსტრუმენტებად, რომლებიც პირველ რიგში წარმოადგენენ ოპერატიული ანალიტიკური დამუშავების ინსტრუმენტებს (online analytical processing, OLAP); კორპორატიული BI- ნაკრებებად (enterprise BI suites, EBIS); BI-პლატფორმებად. BI-ინსტრუმენტების ძირითადი ნაწილი იყოფა კორპორატიულ BI-ინსტრუმენტებსა და BI-პლატფორმებზე. მოთხოვნებისა და რეპორტების გენერაციის საშუალებები ძირითადად ჩაინაცვლება კორპორატიული BI-ნაკრებებით. მრავალგანზომილებიანი OLAP-მექანიზმები და სერვერები, აგრეთვე რელაციური OLAP-მექანიზმები წარმოადგენენ BI-ინსტრუმენტებს (BI-პლატფორმის ინფრასტრუქტურას). BI-ინსტრუმენტების დიდი ნაწილი გამოიყენება საბოლოო მომხმარებლებით რეპორტებში მონაცემებთან წვდომის, ანალიზის და გენერაციისათვის. ეს მონაცემები ხშირად საცავებში და მონაცემთა ვიტრინებშია წარმოდგენილი. აპლიკაციების დამუშავებლები იყენებენ BI-პლატფორმებს BI-დანართების შექმნისა და დანერგვისათვის, რომლებიც განიხილებიან როგორც BI- ინსტრუმენტები. BI-დანართის მაგალითად შეგვიძლია მოვიყვანოთ EIS ხელმძღვანელის ინფორმაციული სისტემა.

მოთხოვნისა და რეპორტების გენერატორები – ტიპური “სამაგიდე” ინსტრუმენტებია, რომლებიც მომხმარებელს აძლევენ მონაცემთა ბაზებთან წვდომის საშუალებას, რომლებიც აწარმოებენ რაღაც დონის ანალიზს და რეპორტების ფორმირებას. მოთხოვნები

¹⁵ Э. Спирли Корпоративные хранилища данных. Планирование, разработка, реализация. Том.1: Пер. с англ. // М. Вильямс, 2001

¹⁶ С. Архипенков, Д. Голубев, О. Максименко Хранилища Данных. От концепции до внедрения/ Под общ. Ред. С.Я. Архипенкова // М.: ДИАЛОГ-МИФИ, 2002

შესაძლოა იყოს, როგორც ჩაუფიქრებელი (ad hoc), ასევე შესაძლოა ჰქონდეს რეგლამენტირებული ხასიათი. არსებობს რეპორტების გენერაციის სისტემა (როგორც წესი სერვერული), რომლებიც მხარს უჭერენ რეგლამენტურ მოთხოვნებსა და რეპორტებს. სამაგიდე მოთხოვნებისა და რეპორტების გენერატორები აგრეთვე გაფართოვებულნი არიან OLAP-ის ზოგიერთი გაადვილებული საშუალებებით. ამ კატეგორიის განვითარებული ინსტრუმენტები თავის თავში აერთიანებენ რეგლამენტური რეპორტების და სამაგიდე მოთხოვნების გენერატორებს, რეპორტების გადაგზავნის და ოპერატიული განახლების პაკეტური გენერაციის საშუალებებს და ქმნიან ეგრეთწოდებულ კორპორატიულ რეპორტების ერთობლიობას (corporate reporting)¹⁷. მის არსენალში შედის რეპორტების სერვერი, გადაგზავნის საშუალებები, მოვლენების შეტყობინების და გადახრების (alerts) რეპორტების Web-ზე პუბლიკაცია. მახასიათებელი წარმომადგენლები: Crystal Reports, Cognos Impromptu და Actuate e.Reporting Suite.

EBIS-ი წარმოადგენს BI-ინსტრუმენტებისათვის ბუნებრივ წარმოდგენის გზას, რომლებიც ადრე მიეწოდებოდნენ დანაწევრებული პროდუქტების სახით. ეს ნაკრებები ინტეგრირდება მოთხოვნების გენერაციის ინსტრუმენტების ნაკრებებად, რეპორტებად და OLAP. კორპორატიულ BI-ნაკრებებს უნდა გააჩნდეს მაშტაბირება და შესაძლებლობა გავრცელდნენ არა მარტო შიდა მომხმარებლებისათვის, არამედ აგრეთვე დამკვეთებისათვის, მომწოდებლებისათვის და ა.შ. BI-ნაკრებების პროდუქტები დანერგვისას და მისი მართვის დროს ადმინისტრატორებს უნდა ეხმარებოდნენ ახალი რესურსების დამატების გარეშე. კორპორატიული BI-ნაკრებებისა და Web-ის მჭიდრო “ნათესაური” კავშირის გამო ზოგიერთი მომწოდებელი BI-ნაკრებების - BI-პორტალებს უწოდებს. ამ პორტალურ წინადადებებს უზრუნველყოფენ EBIS საშუალებების ქვესიმრავლეები Web-ბრაუზერის

¹⁷ К. Комафорд. Корпоративная отчетность: Серверная архитектура для распределенного доступа к информации. // *Открытые системы*, 1999, 2.

მეშვეობით, მაგრამ მომწოდებლები ხშირად ზრდიან მათ ფუნქციონალურობას, და ამის მეშვეობით აახლოვებენ მათ მსხვილი კლიენტების ინსტრუმენტების საშუალებებთან. ტიპური EBIS-ის მომწოდებლები არიან Business Objects და Cognos.

ბიზნეს გადაწყვეტილებების მისაღებად საგნობრივი სფეროს ტერმინებში მომხმარებლის შესაძლებლობები, ინფორმაციის ოპერატიული და მრავალ ასპექტური ანალიზის წარმოებაში, სწრაფად იზრდება. პარალელური მოძრაობა ინფორმაციული ანარქიიდან და დიკტატურიდან ინფორმაციული დემოკრატიისაკენ აფარტოვებს business intelligence-ის მომხმარებლების კონტინენტს. პირველ ადგილზე გადის კორპორატიულ მონაცემებთან მოქნილი წვდომა, არა თუ უბრალოდ კონკრეტული ამოცანის ფუნქციონალური გადაწყვეტა. მცირდება ინფორმაციული ტექნოლოგიების ქვედანაყოფებთან პირდაპირი კავშირი, რომლებიც ამზადებენ მოთხოვნებსა და რეპორტებს შეკვეთით. შესაძლოა გადასვლა სტატიკური რეგლამენტური რეპორტებიდან “ცოცხალ რეპორტებზე” და ანალიტიკოსები ღებულობენ კროს-თემატიკური ანალიზის ჩატარების და აგრეგაციული რეპორტების ნულიდან შექმნას საშუალებას სემანტიკური ფენით, რომელიც აღწერს ყველა მაჩვენებელს და კორპორატიული ინფორმაციის ჭრილებს. იგივე საშუალებები შეუძლიათ გამოიყენონ პროგრამისტებმა სწრაფი რეგლამენტური და პარამეტრული რეპორტების შესაქმნელად.

BI-სადმი Web წვდომა (როგორც სტატიკურისადმი, ისევე დინამიკურისადმი) იძლევა საშუალებას უზრუნველყოს რეალური კორპორატიული ინფორმაციული სივრცე და თანამშრომლების კოლექტიური მუშაობა.

ძირითად რისკად ითვლება ძალიან სწრაფი ცვლილებები BI ტექნოლოგიებში და აგრეთვე შეუმოწმებელი გადაწყვეტილებებისა და საშუალებების გამოყენება. საჭიროა: მომწოდებლების გადამოწმება, მათი მდგრადობის შეფასება, განვითარების მიმართულება, რეგულარულად ახალი საშუალებების შემოწმება, BI-ის ტიპიზაცია და უნიფიკაცია. სხვა რისკი დაკავშირებულია მონაცემთა ხარისხზე - თუ ისინი არ არიან საჭირო დონეზე გარდაქმნილნი, გასუფთავებულნი ან კონსოლიდირებულნი, მაშინ BI- ინსტრუმენტების საშუალებები ან

აპლიკაციები ვერ შეძლებენ მონაცემების ჭეშმარიტად გაზრდას. რიგი პრობლემები შესაძლოა შეიქმნას მეტამონაცემების შეუთავსებლობის გამო. დიდი კორპორაციის ფარგლებში ეს საკითხები წყდება ინფრასტრუქტურულ დონეზე კორპორატიული მონაცემთა საცავების შექმნითა და მეტამონაცემების ცენტრალიზებული მართვით.

საცავის შექმნა იძლევა საშუალებას წესრიგი დავამყაროთ შემაგროვებელი მაჩვენებლების ნომენკლატურაში, მონაცემების შეგროვებაში, მათ გავრცელებასა და წვდომის სანქციონირებაში.

BI-ტექნოლოგიას არ გააჩნია ამ პრობლემების კომპლექსურად გადაწყვეტის საშუალება, ხოლო მათი გვერდის ავლა კი გვიბრუნებს ინფორმაციულ ანარქიას და მონაცემების დიდ “ზეაფებს”.

დასავლური კომპანიების მსგავსად ყველა კომპანია აქტიურად იწყებს ინტერნეტ-ტექნოლოგიების დანერგვას. დღესდღეობით სულ უფრო მეტი და მეტი სპეციალისტები, რომლებიც მუშაობენ არა მარტო ტექნოლოგიების სფეროში, ხვდებიან თუ რამდენად ეფექტურია ინტერნეტ ტექნოლოგიები მათი ბიზნესის წინსვლისათვის.

მონაცემთა ინტელექტუალური ანალიზი პროგრამული გადაწყვეტილებების გამოყენებით, არა მარტო ლოკალურ გარემოში, არამედ ინტრანეტ და ინტერნეტ გარემოშიც, ანალიტიკოსებს უზენაეს მონაცემთა მუშაობის ახალ საშუალებებს.

ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურის თანამედროვე განვითარების ტენდენციები ფუძნდება ინტერნეტ-ტექნოლოგიებზე. ტრადიციულ ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურას, არც თუ ისე დიდი ხნის წინ შეემატა Web-პორტალი, რომელიც სულ უფრო მთავარ ადგილს იკავებს ინფორმაციულ-ანალიტიკურ სისტემაში.

ინფორმაციის მიღება ჩვეული Web-ბრაუზერის მეშვეობით იძლევა საშუალებას ეკონომია გავუწიოთ დანახარჯებს, რომლებიც დაკავშირებულია შესყიდვებთან და სამაგიდე ანალიტიკური აპლიკაციების მხარდაჭერასთან კლიენტების დიდი რიცხვისათვის.

Web-პორტალის რეალიზაცია იძლევა საშუალებას მოვამარაგოთ ანალიტიკური ინფორმაციით მომხმარებლები როგორც ოფისის

შიგნით, ასევე მსოფლიოს ნებისმიერ წერტილში, რომლებიც არიან დაკავშირებულნი პორტალთან ინტერნეტის მეშვეობით.

1.2. ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურის რეალიზაცია შერეული გადაწყვეტილებების საფუძველზე

პირველ ქვეთავში ჩვენ აღწერეთ ინფორმაციულ-ანალიტიკური სისტემის კორპორატიული ამოცანების გაერთიანება, შენახვა და ინფორმაციის ანალიზი.

მძიმე პროცესის გვერდის ასავლელად გამოიყენება შერეული გადაწყვეტილებების მიღების მეთოდოლოგია, რისთვისაც გამოიყოფა სხვადასხვა პროგრამული პროდუქტების ჯგუფი, რომლებიც იძლევიან ინფორმაციულ-ანალიტიკური სისტემის რეალიზაციის საშუალებას დიდი ბლოკებით. ასეთი შერეული გადაწყვეტილებების ჯგუფების განსახილველად წარმოვადგენთ საქმიანი ანალიზისა და მონაცემთა ბაზების მართვის სისტემების ინსტრუმენტებს.

Gartner ტერმინოლოგიაში საქმიანი მონაცემთა ანალიზის ინსტრუმენტები მიეკუთვნებიან ეგრეთწოდებულ BI – პაკეტებს(Enterprise BI Suites), რომელიც ქმნის ინფორმაციული ტექნოლოგიების ბაზრის გარკვეულ სეგმენტს. ეს სეგმენტი წარმოდგენილია 14 კომპანიის მიერ: Actuate, AlphaBlox, Business Objects, Brio, CA, Cognos, Crystal, Hummingbird, Information Builders, Microsoft, MicroStrategy, Oracle, Sagent, Viador. ორი კომპანია - Cognos და Business Objects – ლიდირებს გარკვეული წლების განმავლობაში.

მაგალითში შერეული გადაწყვეტილებების რეალიზაციისათვის გამოვიყენოთ ფირმა Cognos მონაცემთა ანალიზის ინსტრუმენტები და ფირმა Oracle – ის მონაცემთა ბაზების მართვის სისტემების ინსტრუმენტები. ინფორმაციულ-ანალიტიკური სისტემის შერეული გადაწყვეტილებების რეალიზაციისათვის საჭირო პროდუქტების სია წარმოდგენილია ცხრილ 1-ში.

იფორმაციულ- ანალიტიკური სისტემის კომპონენტები			პროდუქტის დანიშნულება	პროდუქტის დასახელება	პლატ- ფორმა
მონაცემთა ანალიზი	ინტერნეტ და ინტრანეტ ქსელებში		Web-პორტალი	Upfront	Cognos Series 7
			რეგლამენტირებული ანგარიშგება Web-ში	Impromptu Web Reports	
			თავისუფალი მოთხოვნები Web-ში	Congos Query	
			მრავალგანზომი- ლებიანი ანალიზი (OLAP) Web-ში	PowerPlay Enterprise Server	
			პრიორიტეტების ვიზუალიზაცია და გამოვლენა Web-ში	VisualizerWeb	
	ლოკარულ ქსელში LAN		რეგლამენტირებული ანგარიშგება და თავისუფალი მოთხოვნები	Impromptu	Cognos Series 7
			მრავალგანზომილებია ნი ანალიზი (OLAP)	PowerPlay	
			ცოდნის მოპოვება (data mining)	4Thought, Soenario	
			პრიორიტეტების გამოვლენა და ვიზუალიზაცია	Visualizer	
			მეგამონაცემების ერთიანი მოდულის აღწერა	Architect	
აღმოფხვ- რა, გარდაქმნა, ჩატვირთვა და მონაცემთა შენახვა	მონაც ემთა ვიტრი ნები	მრავალ- გან- ზომი- ლები- ანი	მონაცემთა ვიტრინების მრავალგანზომილებია ნი სისტემის მართვა (OLAP)	PowerPlayTransformati on Server	Cognos Series 7
			შექმნა და პროექტირება	PowerPlayTransformer	
			ET კომპონენტი მრავალგანზომილებია ნი მონაცემთა ვიტრინებისათვის - (OLAP)	DecisionStream	

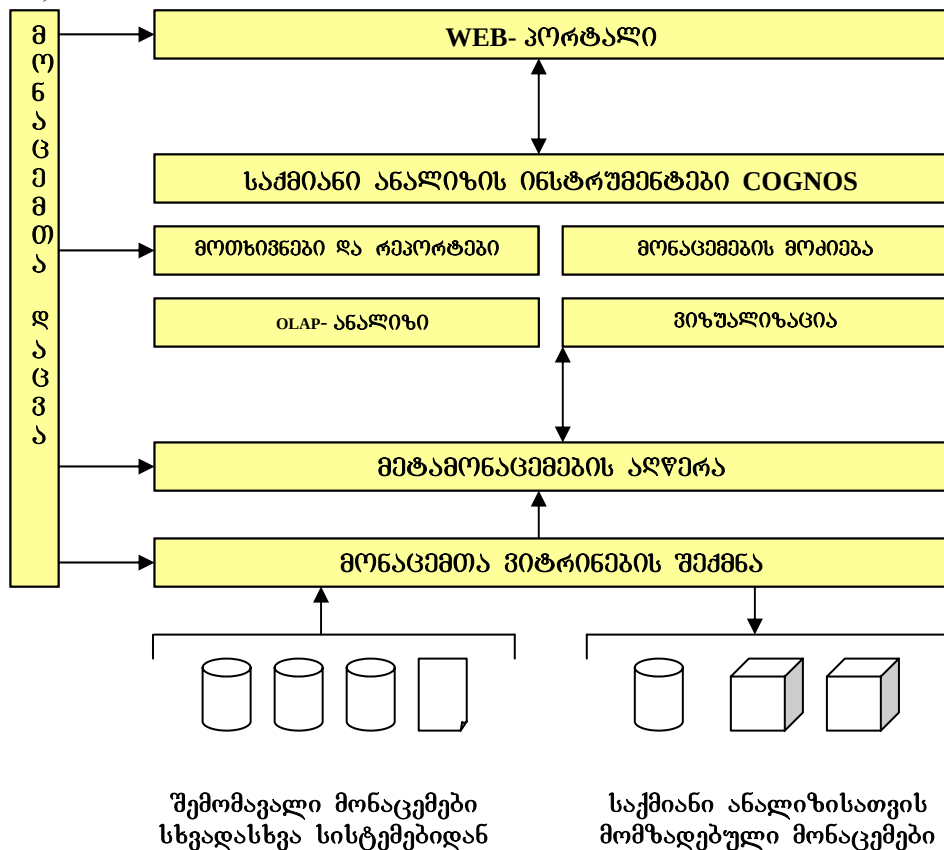
აღმოფხვრა, გარდაქმნა, ჩატვირთვა და მონაცემთა შენახვა	მონაცემთა ვიტრინები	მრავალ-განზომილებიანი	ET-კომპონენტი რელაციური მონაცემთა ვიტრინებისათვის	DecisionStream	
		რელაციური	მონაცემთა ბაზების მართვის სისტემა რელაციური მონაცემთა ვიტრინებისათვის	Oracle Database	Oracle9i
			რელაციური მონაცემთა ბაზების პროექტირება და შექმნა	Oracle Warehouse Builder	
	მონაცემთა საცავი		ETL – კომპონენტი მონაცემთა საცავებისათვის	Oracle Warehouse Builder	Oracle9i
			მონაცემთა ბაზების მართვის სისტემა მონაცემთა საცავებისათვის	Oracle DataBase	
			მონაცემთა საცავების პროექტირება და შექმნა	Oracle Warehouse Builder	

ცხრილი1. ინფორმაციულ-ანალიტიკური სისტემის შერეული გადაწყვეტილებების რეალიზაციისათვის საჭირო პროდუქტები

უნდა აღინიშნოს, რომ Cognos BI-პაკეტი მოიცავს ერთმანეთთან ინტეგრირებულ BI-ინსტრუმენტებს. პროგრამული კომპონენტების მჭიდრო ინტეგრაცია ერთიან პაკეტურ პლატფორმაში მიიღწევა მეტამონაცემების BI ინსტრუმენტების საერთო მოდულების აღწერის გაერთიანებით, წვდომის უფლებების განაწილებით და აგრეთვე Web-პორტალის მეშვეობით – ინტერნეტ-მომხმარებლებისათვის ერთიანი შესვლის წერტილით კორპორაციულ ინფორმაციასთან მუშაობის გარემოში. გადაწყვეტილების სისრულე, რომელიც იძლევა საშუალებას BI-პლატფორმისათვის ბაზარზე დაიკავოს მნიშვნელოვანი აგილი, წარმოადგენს ETL-ინსტრუმენტები მონაცემთა ვიტრინების შესაქმნელად (რელაციურისათვის და მრავალგანზომილებიანისათვის). მონაცემთა ვიტრინები, მრავალგანზომილებიანი თუ რელაციური, იქმნებიან DecisionStream-სა და PowerPlay Transformer-ის მეშვეობით. მონაცემთა

ინტელექტუალური ანალიზისათვის გამოიყენება რეპორტების გენერაციის ინსტრუმენტი Impromptu, მრავალგანზომილებიანი ანალიზისათვის ინსტრუმენტი – PowerPlay, მონაცემთა დაზვერვის ინსტრუმენტი - Scenario და 4Thought, მონაცემთა ვიზუალიზაციის ინსტრუმენტი Visualizer. მეტამონაცემების აღსაწერად გამოიყენება პროგრამული მოდული Architect – ყველა მონაცემთა ანალიზის ინსტრუმენტებისათვის საერთო. ინტერნეტ-პორტალი, რომელიც აძლევს საშუალებას მობილურ მომხმარებლებს ჩაერთონ კორპორატიული ინფორმაციის ანალიზში, ხორციელდება Upfront ინსტრუმენტის მეშვეობით.

განვიხილოთ უფრო ვრცლად Cognos-ის ინსტრუმენტები, რომლებიც პასუხს აგებენ აღმოფხვრის, გარდაქმნის, ჩატვირთვის, შენახვის, მონაცემთა საქმიანი ანალიზის პროცესების მხარდაჭერაში (ნახ.3).



ნახ. 3 COGNOS საქმიანი ანალიზის პაკეტი¹⁸

¹⁸ Галахов И. В., Волков И. Ю. Архитектура современной информационно-аналитической системы // Директор ИС №3, 2002год.

ანალიტიკოსები იშვიათად მუშაობენ “დამუშავებელ” მონაცემებთან ტრანზაქციული მონაცემთა ბაზებიდან. ისინი უპირატესობას ანიჭებენ მონაცემების მოპოვების გზას ვიტრინებიდან და მონაცემთა საცავებიდან. აქედან გამომდინარე Cognos-ის პაკეტური პლატფორმა აგრეთვე მოიცავს ინფორმაციის წინასწარ დამუშავების ინსტრუმენტს, რაც უზრუნველყოფს წარმოდგენილი გადაწყვეტილების მთლიანობასა და სისრულეს.

DecisionStream ინსტრუმენტის მეშვეობით წარმოებს მონაცემების წინასწარი მომზადება, ესე იგი აღმოფხვრა, დამუშავება და მონაცემების ჩატვირთვა მონაცემთა ანალიტიკურ წყაროებში (მრავალგანზომილებიანი და რელაციური) OLAP-ანალიზის ინსტრუმენტების, რეპორტების გენერატორების, მონაცემთა დაზვერვის ინსტრუმენტების, ვიზუალიზაციისა და პრიორიტეტების გამოვლენის ეფექტური მუშაობისათვის.

DecisionStream-ს გააჩნია რელაციური და არარელაციური ფორმატის ინფორმაციის წყაროებთან მიერთების საშუალება, გაერთიანება, მონაცემების გარდაქმნა და მათ საფუძველზე მონაცემთა ვიტრინების შექმნა რელაციური სქემა “ვარსკვლავის” სახით (მაგალითად, მონაცემთა ბაზა Oracle-ში); და დამატებით მონაცემთა წარმოდგენა მრავალგანზომილებიანი მოდელით, სადაც ინფორმაცია წარმოდგენილია მაჩვენებლებისა და განზომილებების სახით. შექმნის ავტომატიზაციის შესაძლებლობასთან ერთად რელაციურ ბაზასთან ერთად მისი მრავალგანზომილებიანი მოდელი გამოიყოფა, ეს არის Cognos DecisionStream-ი პროგრამული პროდუქტი, რომლის მეშვეობითაც ხდება ჩვეულებრივი აღმოფხვრის ინსტრუმენტების, გარდაქმნისა და მონაცემების ჩატვირთვის გამოყოფა.

დანაწევრებული მონაცემების გაერთიანების შესაძლებლობა ერთიან ინფორმაციის შენახვის წყაროში, მათი დამუშავება და ყველაფერ ამასთან დაკავშირებული, მონაცემებთან წვდომის სისწრაფე DecisionStream-ს ხდის შეუცვლელ საშუალებად ინფორმაციულ-ანალიტიკური სისტემებისათვის, სადაც საჭიროებს სისტემის სწრაფი რეაგირება მომხმარებლის მოთხოვნაზე მონაცემების მისაღებად, შესასწავლად და ანალიზისათვის.

მეტამონაცემების აღწერის ინსტრუმენტი Cognos Architect-ი პასუხობს მეტამონაცემების ცენტრალიზებული მართვის მოთხოვნებს, რაც უზრუნველყოფს Cognos ინსტრუმენტების ინტეგრაციას. Architect-ი წარმოადგენს ცენტრალურ ბირთვს მეტამონაცემების შექმნასა და მართვაში და აგრეთვე ორგანიზაციის ბიზნეს-წესებში.

ასეთი სახით, მეტამონაცემები, რომლებიც აღწერილია ერთ Architect მოდელში ხდება ხელმისაწვდომი ყოველივე Cognos BI ინსტრუმენტში. Architect-ის ინსტრუმენტალური მოდული გვაძლევს ინფორმაციის წყაროსთან წვდომის აღწერის საშუალებას, ლოგიკურ ბიზნეს-წარმოდგენებს, ყოველივე ანალიტიკური აპლიკაციების მაშტაბებს.

მთელი ორგანიზაციის მეტამონაცემების უნიფიცირებული აღწერის რეზულტატი წარმოადგენს სწორ და დადებით მონაცემების წარმოდგენას.

საქმიანი ანალიზის ინსტრუმენტები

Cognos კომპანიის მონაცემთა საქმიანი ანალიზის ინსტრუმენტები საშუალებას იძლევიან ჩატარდეს საქმიანი ანალიზის მთელი სპექტრი: შეიქმნას მოთხოვნები, რომლებიც იძლევა რეპორტების ფორმების შექმნის საშუალებას, ჩატარდეს მრავალგანზომილებიან მონაცემებზე OLAP-ანალიზი, ჩატარდეს მონაცემების მოძიება (data mining¹⁹) და აგრეთვე წარმოგიდგინოს ინფორმაციის ვიზუალიზაცია და გამოავლინოს პრიორიტეტები.

Cognos გადაწყვეტილებები მოთხოვნების შექმნის სფეროში და რეპორტებთან მუშაობა საშუალებას იძლევა მომხმარებლები უზრუნველყოთ რეპორტებით საჭირო ფორმატში მოთხოვნილ პლატფორმაზე (Windows, Excel, Web). ხელმძღვანელებისათვის, რომლებიც ღებულობენ გადაწყვეტილებებს, Cognos საშუალებები უზრუნველყოფს ინფორმაციასთან წვდომას მათთვის მოსახერხებელ ფორმაში (ცხრილები, დიაგრამები, გისტოგრამები, სამგანზომილებიანი გრაფიკები და ა.შ.), ანალიტიკოსებისათვის კი წარმოიდგინება

¹⁹ Глобальный// <http://www.rdtex.ru/docs/glossary/D57123.html>

საშუალება პრაქტიკულად ყველა დონის სირთულის ინფორმაციის გადათვალიერების, მათ შორის მოთხოვნების შექმნა, მონაცემთა კვლევა და ნავიგაცია Web-ში. ამგვარად, საბოლოო მომხმარებლის წინაშე მდგომი ამოცანებისგან დამოკიდებულია შესაბამისი ინსტრუმენტალური საშუალებების მოძიება. რეგლამენტირებული რეპორტებისათვის წარმოდგინება Cognos Impromptu, ხოლო არარეგლამენტირებული მოთხოვნებისათვის - Cognos Query.

დღევანდელ დღეს Cognos კომპანიას გააჩნია ფაქტიურად ერთ-ერთი ყველაზე საუკეთესო OLAP-გადაწყვეტილებები, არსებულ მონაცემთა ანალიტიკური დამუშავების ინსტრუმენტების ბაზარზე. OLAP-ინსტრუმენტების ბაზრის შესწავლის რეზულტატების მიხედვით (www.olapreport.com), კომპანია Cognos, PowerPlay ინსტრუმენტით იკავებს მეორე ადგილს მსოფლიო OLAP-სისტემის ბაზარზე Hyperion-ის²⁰ შემდეგ.

კომპანია Cognos-ის OLAP-ინსტრუმენტთა ძირითად განსაკუთრებულობას წარმოადგენს OLAP-აპლიკაციების ავტომატიზაციის პროცესის ეფექტური შექმნა, რაც იძლევა პრაქტიკულად ყველანაირი სირთულის სისტემისათვის დამუშავების ვადების შემცირების საშუალებას, მხოლოდ ორგანიზაციის შიდა ადამიანთა რესურსების გამოყენებით.

ამოცანის სირთულე და PowerPlay ინსტრუმენტის არჩეული კონფიგურაცია განსაზღვრავს OLAP-სისტემების მასშტაბს. PowerPlay ინსტრუმენტის მეშვეობით შესაძლებელია დამუშავდეს სისტემები ერთი მომხმარებლისათვის, მომხმარებელთა ჯგუფისათვის დეპარტამენტში, რამოდენიმე განყოფილებისათვის, აგრეთვე ორგანიზაციის მასშტაბის სისტემისათვის, რომელთა მომხმარებლები შესაძლებელია იყვნენ აგრეთვე გარე პირები და ორგანიზაციები, რომლებიც ავტორიზებულნი არიან მონაცემებთან მუშაობისათვის.

Cognos გადაწყვეტილებები OLAP –ტექნოლოგიის ბაზაზე Windows-, Excel- და Web- მომხმარებლებს უხსნის მოხერხებულ წვდომას მათ მიერ წარმოდგენილ საქმიან ენაზე (მრავალგანზომილებიან ინფორმაციას),

²⁰ www.olapreport.com

რომელიც წარმოდგენილია გადათვალიერებისათვის, რეპორტების ფორმირებისათვის და OLAP –ანალიზის ჩატარებისათვის. Cognos კომპანიის BI-ინსტრუმენტების ინტეგრაციის მეშვეობით მრავალგანზომილებიანი მონაცემები შესაძლოა გამოყენებულნი იყვნენ შემდგომი მონაცემთა მოძიების, ვიზუალიზაციისა და პრიორიტეტების გამოვლენის შესწავლისათვის.

Cognos გადაწყვეტილებები მონაცემთა მოძიების (data mining) სფეროში გვეხმარებიან ჩავატაროთ კვლევა მნიშვნელოვან კორელაციებზე მონაცემში, მოვიპოვოთ დაფარული ტენდენციები და შევქმნათ პროგნოზები. მონაცემთა მოძიების ინსტრუმენტები Scenario და 4Thought გვეხმარებიან ადვილად განვსაზღვროთ, რომელი ფაქტორები ზემოქმედებენ ბიზნესის საკვანძო მაჩვენებლებზე, მაგალითად მოგებაზე, გაყიდვების რაოდენობაზე, მიყიდველთა ჯგუფის მიზნობრივ ღირებულებაზე და ა.შ.

მონაცემთა მოძიების Scenario ინსტრუმენტი გვაძლევს საშუალებას შევაფასოთ სხვადასხვა ფაქტორების ზეგავლენა ანალიზის პროცესში არჩეულ მიზნობრივ მაჩვენებელზე. ანალიზის რეზულტატები წარმოდგინება კლასიფიკაციური და რეგრესიული ხეების მეშვეობით, რომლებსაც აგრეთვე უწოდებენ გადაწყვეტილებების ხეებს. ამავე დროს წარმოებს გამონაკლისების ძებნა, რომელიც მიმართულია ელემენტების გამოვლენაზე და სტატისტიკურად არ აკმაყოფილებენ გამოვლენილ დამახასიათებელ ურთიერთქმედებებს მონაცემებში.

მონაცემთა მოძიების 4Thought ინსტრუმენტი, რომელიც იყენებს ნეირონულ ქსელების ტექნოლოგიას, გვაძლევს რთული არახაზობრივი მოდელების შექმნის საშუალებას, რომლებიც მყარნი არიან არახუსტ და ვრცელი მონაცემების მიმართ. 4Thought ხელმისაწვდომია იმ მენეჯერებისთვისაც კი, რომლებიც არ არიან დეტალურად გაათვითცნობიერებულნი მონაცემების შეფასების სტატისტიკურ მეთოდებში და აძლევს მათ საშუალებას ჩაატარონ მონაცემთა ანალიზი არა თუ შექმნან პროგნოზები.

მონაცემთა მოძიების 4Thought ინსტრუმენტისა და Scenario –ს ღირსება მდგომარეობს შემდეგში: შემომაგალი ინფორმაციის

მონაცემთა მოძიებისათვის ამ ინსტრუმენტებს შეუძლიათ მოიპოვონ მონაცემები პრაქტიკულად ყველა ანალიტიკური მრავალგანზომილებიანი მონაცემთა ბაზებიდან: Cognos PowerCubes, Oracle Express, SAP BW, Hyperion Essbase, IBM DB2 OLAP Server, Microsoft SQL Server OLAP Services, NCR TeraCube.

მონაცემთა მოძიების ინსტრუმენტების გამოყენებით მოძიებული დაფარული ტენდენციები შესაძლებელია გამოყენებულ იქნას მმართველობითი კვალიფიცირებული გადაწყვეტილებების მისაღებად. მაშასადამე, იგი ერთ-ერთი წარმატებული საფეხურია სტრატეგიული დაგეგმვის ჩატარებისას.

ცნობილია, რომ გრაფიკულ ინფორმაციას, ტექსტურისაგან განსხვავებით, ადამიანი გაცილებით სწრაფად აღიქვავს. მონაცემთა წარმოდგენა ხელმძღვანელებისათვის ინტუიციურად გასაგებ ფორმაში აადვილებს დიდი მოცულობის ინფორმაციაში ორიენტაციას და იმავე დროს უსარგებლო ინფორმაციის აღმოფხვრას.

Cognos გადაწყვეტილებები მონაცემთა ვიზუალიზაციის სფეროში და პრიორიტეტების გამოვლენაში გადათვალთვალებისა და ანალიზის პროცესში საშუალებას იძლევა დიდი მოცულობის მონაცემებიდან გამოვეყოთ საკვანძო ინფორმაცია და გავაკეთოდ მასზე აქცენტი. Cognos-თვის მონაცემთა ვიზუალიზაციაში Windows და Web გარემოში პრიორიტეტების გამოვლენაში ძირითად ინსტრუმენტს წარმოადგენს Visualizer.

ინფორმაციის წარმოდგენის ვიზუალური შაბლონების გაფართოვებული კრებული (ორგანზომილებიან და სამგანზომილებიან დიაგრამებიდან და გრაფიკებიდან რაოდენობრივი მაჩვენებლებისათვის გეოგრაფიულ რუქებამდე რეგიონალური დამოკიდებულებების გამოსავლენად, აგრეთვე სპეციალური სემაფორები და ვიზუალური ელემენტები, რომლებიც გამოიყენება პრიორიტეტების გამოსავლენად) იძლევა დროის შემცირების საშუალებას საჭირო ინფორმაციის მოსაძებნად.

მონაცემთა ვიზუალიზაციისა და პრიორიტეტების გამოვლენის პროცესში მიღებული რეზულტატები წარმოადგენენ ეფექტური გადაწყვეტილებების მიღების საფუძველს.

Cognos ინსტრუმენტალური საშუალებები ინტერნეტისათვის მომხმარებლებს აძლევს საშუალებას მოამზადონ რეპორტები და მოახდინონ მონაცემთა ბაზებთან მოთხოვნები, შექმნან პროგნოზები, იმუშაონ არსებული რეპორტების ფორმების ინფორმაციასთან, ჩაატარონ სრული მრავალგანზომილებიანი მონაცემების OLAP-ანალიზი, მოახდინონ ინფორმაციის ვიზუალიზაცია და გამოავლინონ პრიორიტეტები (scorecarding). ყველა ზემოდ ჩამოთვლილი ინფორმაციასთან მუშაობის საშუალებები ხელმისაწვდომია მომხმარებლისათვის პროგრამისტების ჩარევის გარეშე.

Cognos პლატფორმისათვის ინტერნეტ-პორტალის შექმნისა და დამუშავების ინსტრუმენტს Upfront მოდული წარმოადგენს. Upfront-ის მიერ შექმნილი პორტალი შესაძლოა ინტეგრირებულ იქნას არსებულ ორგანიზაციაში პორტალში, ინტრანეტში ან/და ექსტრანეტ გადაწყვეტილებებში (B2B) და აგრეთვე იქნას ცალკე გამოყენებული. WEB-პორტალის შექმნა წარმოებს პუბლიკაციის სფეროში მუშაობით და არ მოითხოვს პროგრამირებას. ასეთი სახით, Upfront მოდული ადმინისტრატორებს წარმოუდგენს ცენტრალიზებული მუშაობის საშუალებას ადვილად ცვალებადი ინტერფეისის მეშვეობით, ხოლო საბოლოო მომხმარებელს, რომლებიც იმყოფებიან ეკრანის (firewall) გარეთ თუ შიგნით, საერთო შესვლის წერტილს, რომლიდანაც ხდება ინფორმაციასთან წვდომა, რომელსაც გააჩნია ინფორმაციის წარმოდგენის სხვადასხვა ფორმები. ამასთან ყოველ მომხმარებელს, რომელსაც გააჩნია ინტერნეტ-ბრაუზერი და კორპორატიულ ინფორმაციასთან წვდომის უფლება ეძლევა საშუალება იმუშაონ ადრე მომზადებულ რეპორტების ფორმებთან რელაციურ და მრავალგანზომილებიან მონაცემებზე, აგრეთვე დამოუკიდებლად მოახდინონ რეპორტების ფორმირება და პუბლიკაცია კერძო და საერთო მომხმარებლისათვის. რელაციური მონაცემების მანიპულირების საშუალებისა და OLAP-ანალიზის გარდა, მომხმარებლებს აქვს პრიორიტეტების გამოვლენისა და მონაცემთა ვიზუალიზაციის საშუალება. Upfront ინტეგრირებულია PowerPlay OLAP-ინსტრუმენტთან, Impromptu Web Reports რეპორტების გენერატორთან, Cognos Query

გენერატორთან და პრიორიტეტების გამოვლენისა და ვიზუალიზაციის ინსტრუმენტ Visualizer Web-თან. ამის გარდა, პორტალში დამატებითი კონფიგურაციების შეუცვლელად, მომხმარებელს საშუალება რჩება მისთვის უცვლელი ფორმატების Excel, Word, PDF დოკუმენტებთან მუშაობის.

ასეთი სახით ინფორმაციულ-ანალიტიკური სისტემის რეალიზაციისას შერეული გადაწყვეტილებების მიღების საფუძველზე, რომელიც ყველაზე ხშირად გამოიყენება პრაქტიკაში, შესაძლოა სხვადასხვა საფეხურზე გამოყენებული იყოს ერთმანეთისაგან დამოუკიდებელი ინსტრუმენტებით. მონაცემთა ანალიზის ინსტრუმენტების ჯგუფი შესაძლოა იყოს დამოუკიდებელი აღმოფხვრის, გარდაქმნის, ჩატვირთვის და შენახვის ინსტრუმენტების ჯგუფებისაგან, ანუ ყოველი ამ ჯგუფთაგანი შესაძლოა წარმოდგენილ იქნეს ცალკეულ მწარმოებლად. მეორე ჯგუფის ინსტრუმენტები უკეთესია ამორჩეულ იქნეს მონაცემთა ბაზის მართვის სისტემების მომწოდებლებიდან, ხოლო საქმიანი ანალიზის ინსტრუმენტები კი, მომწოდებლებისაგან, რომლებიც სპეციალიზირდებიან მონაცემთა საქმიანი ანალიზის ინსტრუმენტების სპექტრზე.

1.3. ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურის რეალიზაცია პლატფორმულ-ბაზირებული გადაწყვეტილებების საფუძველზე

საწყისისთვის განვიხილოთ ინფორმაციულ-ანალიტიკური სისტემის არქიტექტურის რეალიზაციის ვარიანტი, როდესაც გამოიყენება ერთი მწარმოებლის კომპონენტები - პლატფორმის მომწოდებლისაგან. თუ გამოვიყენებთ Gartner კვლევითი ცენტრის ტერმინოლოგიას, პლატფორმულ-ბაზირებული გადაწყვეტილება საჭიროა ვეძებოდ ეგრეთ წოდებულ BI-პლატფორმების (Business Intelligence Platforms)²¹ მწარმოებელ ფირმებს შორის. ინფორმაციული ტექნოლოგიების ბაზრის მოცემული სეგმენტი წარმოდგენილია 16

²¹ В. Артемьев Что такое BUSINESS INTELLIGENCE? Открытые системы, #04/2003

კომპანიის მიერ, რომლებიც აქ ჩამოთვლილია ალფავიტის მიხედვით: AlphaBlox, Arcplan, CA, Comshare, Crystal, Hyperion, Info Builders, Microsoft, Microstrategy, Oracle, PeopleSoft, ProClarity, Sagent, SAP, SAS, Whitelight. მათ შორის გამოიყოფა 7 ლიდერი და 7 პრეტენდენტი ლიდერობაზე მოცემულ სფეროში: Microsoft, SAS, Oracle, SAP, PeopleSoft, Info Builders, Hyperion. ჩამოთვლილი მწარმოებლებიდან ორ მწარმოებელს Microsoft-სა და Oracle-ს შესაძლებლობა გააჩნია ინფორმაციულ-ანალიტიკური სისტემის ყველა დონის რეალიზაციის თავისი ძალებით ისე, რომ არ მოიშველიოს მესამე ფირმა. გადამწყვეტი კრიტერიუმი ამ მწარმოებლების გამოსაყოფად არის კერძო მონაცემთა ბაზების დამუშავების სისტემის არსებობა.

განვიხილოთ ინფორმაციულ-ანალიტიკური სისტემის რეალიზაციის მაგალითი Oracle ინსტრუმენტების მეშვეობით. პროდუქტების შესაბამისობის სია წარმოდგენილია ცხრილ 2-ში²².

ინფორმაციულ-ანალიტიკური სისტემის კომპონენტები		პროდუქტის დანიშნულება	პროდუქტის დასახელება
მონაცემთა ანალიზი	ინტერნეტ და ინტრანეტ ქსელებში	Web-პორტალი	Oracle Portal
		რეგლამენტირებული ანგარიშგება Web-ში	Oracle Reports Services
		თავისუფალი მოთხოვნები Web-ში	Oracle Discoverer
		(OLAP) მრავალგანზომილებიანი ანალიზი Web-ში	OracleJDeveloper Oracle Business Intelligence Beans კომპონენტთან
	(LAN) ლოკალურ	რეგლამენტირებული ანგარიშგება	Oracle Reports

²² <http://www.oracle.com> საიტის მასალებიდან

	ქსელში		თავისუფალი მოთხოვნები	Oracle Discoverer
			(OLAP) მრავალგანზომილებიანი ანალიზი, კლიენტთა ადგილების დამუშავების ინსტრუმენტი	Oracle JDeveloper Oracle Business Intelligence Beans კომპონენტთან
			ცოდნის მოპოვება (data mining)	Oracle Data Mining Suite
აღმოფხვრა, გარდაქმნა, ჩატვირთვა და მონაცემთა შენახვა	მონაცემთა ვიტრინები	მრავალგანზომილებიანი	(OLAP) მრავალგანზომილებიანი მონაცემთა ვიტრინების მართვის	Oracle OLAP
		რელაციური	(OLAP) მრავალგანზომილებიანი მონაცემთა ვიტრინების პროექტირება და შექმნა	
	მონაცემთა საცავები		ETL – კომპონენტი მონაცემთა საცავებისათვის	Oracle Warehouse Builder
			მონაცემთა ბაზების მართვის სისტემა მონაცემთა საცავებისათვის	Oracle Database
			მონაცემთა საცავების პროექტირება და შექმნა	Oracle Warehouse Builder

ცხრილი 2. ინფორმაციულ-ანალიტიკური სისტემის რეალიზაცია Oracle ინსტრუმენტების მეშვეობით. პროდუქტების შესაბამისობის სია

მოკლედ ავღწეროთ ცხრილში მოცემული პროგრამული პროდუქტები მათი გამოყენების თვალსაზრისით ინფორმაციულ-ანალიტიკური სისტემის რეალიზაციისათვის.

Oracle Warehouse Builder- ეს მონაცემთა საცავებისა და ვიტრინების რეალიზაციისა და დამუშავების სფეროა. Common Warehouse Model ღია ბაზის არქიტექტურაზე შექმნილი Warehouse Builder-ი უზრუნველყოფს მთელ რიგ ამოცანებს. მათ რიგს მიეკუთვნება პროექტირება, შექმნა და მონაცემთა საცავების ადმინისტრირება, დამუშავება და აღმოფხვრის პროცედურების გენერაცია, გარდაქმნა, მონაცემთა ჩატვირთვა სხვადასხვა წყაროებიდან, მეტამონაცემების მართვა და შედწევის ინსტრუმენტალური საშუალებების ინტეგრაცია.

Warehouse Builder იყენებს ყოველივე ახალ Oracle8i მონაცემთა ბაზის დამუშავების სისტემის შესაძლებლობებს, რომლებიც დაკავშირებულნი არიან მონაცემთა საცავებთან, ისეთები, როგორც მატერიალიზებული წარმოდგენები, განზომილების ტიპის (dimension) ობიექტები და სხვა. ადმინისტრირების მოცანების გადაწყვეტისას და დამოკიდებული პროცესების მართვისას Warehouse Builder ეყრდნობა Oracle Enterprise Manager და Oracle Workflow-ზე, რომლებიც ინტეგრირებულნი არიან Oracle8i მონაცემთა ბაზასთან.

Warehouse Builder-ის Java-ზე გადწყვეტილი და Wizard-ებით შემკული სამომხმარებლო ინტერფეისი საკმაოდ აადვილებს შექმნის დამუშავებისა და მხარდაჭერის მონაცემთა საცავის ამოცანების გადაწყვეტას. პროდუქტი იძლევა საცავის სქემის ვიზუალური მოდელირების საშუალებას ან Oracle Designer რეპოზიტორიიდან მოახდინოს მისი აღმწერი მეტამონაცემების იმპორტირება. Warehouse Builder-ში შედის ვიზუალური განმარტებები წყაროებს, საცავებსა და მომდევნო გენერაციებს შორის მათი მონაცემების მოდულების ჩატვირთვის საფუძველზე PL/SQL-ის პროცედურების ან SQL*Loader-ის გამოყენებით. ამასთან შესაძლებელია მონაცემთა გარდაქმნის ჩაშენებული ბიბლიოთეკის გამოყენება და აგრეთვე PL/SQL

პროცედურების ჩვენს მიერ შექმნა (ნახ.4). ამას გარდა, Oracle Pure*Integrate პროდუქტი, რომელიც გამოიყენება Warehouse Builder-თან ერთობლივად გვაწვდის მონაცემთა ავტომატური გასუფთავების მძლავრ საშუალებებს, რომლებიც მოიცავენ არამკვეთრი ლოგიკის ალგორითმებს, სახელებისა და მისამართების სინტაქსურ გარჩევასა და ალბათობის მოდელებს და ა.შ. ეს მეთოდები, როგორც წესი გამოიყენება ინფორმაციის დამუშავებისათვის მარკეტინგულ კომპანიებში და გვეხმარება მოვიპოვოთ ინტეგრირებული წარმოდგენები კლიენტებზე და მათ კატეგორიებზე.

ნახ.4 Data Warehouse Builder-ი გვაძლევს საცავების მოდელირების სქემის შექმნის საშუალებას.

Oracle Pure*Extract), აგრეთვე ERP აპლიკაციები (Integrator for Oracle Applications, SAP R/3 კომპონენტებით და ა.შ.)

Common Warehouse Metadata ღია არქიტექტურა მეტამონაცემების გაცვლის საშუალებას იძლევა Warehouse Builder-სა და Discoverer და Express-ის ბიზნეს ანალიზის საშუალებებს შორის. სპეციალური დამხმარეს საშუალებით შეიძლება მეტაინფორმაციის გადატანა Warehouse Builder რეპოზიტორიდან შესაბამის წვდომის საშუალებების ინსტრუმენტალურ მეტამონაცემების ფენებში, რითაც მოახდენს შექმნასა და სისტემის მოვლიანობაში ადმინისტრირებაზე დანახარჯების მინიმიზირებას.

Oracle კორპორაცია, ელექტრონული ბიზნესისათვის პროგრამული უზრუნველყოფის უმსხვილესი მწარმოებელია. მან ანონსირება მოახდინა Oracle Warehouse Builder-ის ბოლო ვერსიის. იგი საკმაოდ ვრცელი, ღია გადაწყვეტილებაა, რომელიც განეკუთვნება დამუშავებას და კორპორატიული მონაცემთა საცავების დანერგვას, ინფორმაციული ჯიხურებისა და ელექტრონული ბიზნესის ინტელექტუალური აპლიკაციებისა.

Oracle Warehouse Builder-ში შეტანილი სიახლეები სამომხმარებლო ინტერფეისსა და წარმოებულობაში, რომლებიც მიმართულნი არიან იმაზე, რომ დამუშავებლები კარგავდნენ ნაკლებ დროს მონაცემთა ბაზების სტრუქტურის მოდერნიზაციაზე, ასევე მონაცემთა ინტეგრაციაზე სხვადასხვა წყაროებიდან და მოქმედი სისტემებიდან და ამავე დროს ინარჩუნებდნენ აქტუალობას და თავისი მონაცემების ზუსტ წარმოდგენას. გაფართოვებული მონაცემთა საცავების მართვის ფუნქციების და მათი გაერთიანების მეშვეობით, რომელიც რეალიზებულია Oracle Warehouse Builder-ში. ელექტრონულ წარმოებებს შესაძლებლობა მიეცემათ მოახდინონ ოპერატიული რეაგირება ბაზრის ცვლილებებზე.

"Xerox Business Services" ესაჭიროებოდა რეპორტების დეტალური სისტემა და მოგებისა და ზარალის დასათვლელი სისტემა, ესე იგი საჭირო იყო სხვადასხვა სისტემებში მყოფი ფინანსური ინფორმაციის გაერთიანება. Oracle Warehouse Builder-ი აღმოჩნდა ყველაზე

მოსახერხებელი ინსტრუმენტი: მისი საშუალებით ჩვენ სწრაფად და მინიმალური დანახარჯებით შევქმნით მონაცემთა საცავებს, რომელშიც დავიყვანთ მთელ ჩვენ ინფორმაციას. ეს შესანიშნავი გადაწყვეტილება აკმაყოფილებს, როგორც IT პროფესიონალების მოთხოვნებს, ასევე ხელმძღვანელებისა და ბიზნეს ქვედანაყოფებს.

გრაფიკული ინტერფეისი, დამხმარების საფუძველზე, მზა გარდაქმნის მოდულებით საკმაოდ აადვილებს და აჩქარებს საცავების პროექტირებისა და დანერგვის პროცესს. დანაწევრებული კონკურენტი გადაწყვეტილებებისაგან განსხვავებით, Oracle Warehouse Builder-ი უზრუნველყოფს მონაცემთა საცავების სრულ მართვას, რაც შეძლებს საშუალებას აძლევს იოლად მოიპოვონ მონაცემები, გარდაქმნან და ჩატვირთონ Oracle9i მონაცემთა ბაზების მართვის სისტემაში და ამასთან ეფექტურად გაუწიონ კონტროლი მონაცემთა საცავების მუშაობის ყველა ასპექტს. ამას გარდა Oracle9i Database-ის ინტეგრაცია Oracle9i Application Server-თან იძლევა მაღალი წარმომადგენლობისა და პრაქტიკულად განუსაზღვრელ Oracle Warehouse Builder-ის მასშტაბირების გარანტიას დაბალი საფასურის საფუძველზე.

Oracle მიდგომა გასხვავდება იმისაგან რასაც გვთავაზობს სხვა კონკურენტები, იმისათვის რომ უზრუნველყოთ დასრულებული ბიზნესის მართვის ინტელექტუალური გადაწყვეტილება. მათ უწევთ ურთიერთქმედება მრავალ სხვა პარტნიორებთან.

Oracle Database, მონაცემთა ბაზების მართვის სისტემა – Oracle კომპანიის პლატფორმული გადაწყვეტილებების საფუძველია. Oracle Database გამოიყენება მონაცემთა ბაზების დამუშავების სისტემის როლში მონაცემთა ვიტრინების ირელაციური საცავებისათვის.

Oracle Database 10g –მსოფლიოში პირველი მონაცემთა ბაზაა, რომელიც დამუშავებულია სპეციალურად ქსელში განაწილებული გამოთვლების ჩასატარებლად. Oracle Database 10g წარმოდგენილია სხადასხვა ტიპის მოწყობილობების ბაზაზე ეფექტური გაშლისათვის საშუალო სერვერებიდან Oracle Enterprise Grid მძლავრ მრავალპროცესორიან სერვერის სისტემებამდე, ცალკეული

კლასტერებიდან კორპორატიულ განაწილებულ გამოთვლით სისტემებამდე.

Oracle Database 10g მომხმარებლებს აძლევს საშუალებას გამოიყენოს აპარატული უზრუნველყოფა – სერვერებისა და მონაცემთა საცავების სისტემების. Oracle Database 10g, გააჩნია ტექნოლოგიები, რომლებიც ადმინისტრატორებს აძლევს საშუალებას შეინახონ, სწრაფად გადაანაწილონ და აღმოფხვრან მონაცემები მომხმარებლებისათვის და აპლიკაციებისათვის, რომლებიც მუშაობენ Grid ქსელებში. Oracle Database 10g საკმაოდ ამაღლებს მონაცემთა დამუშავების მწარმოებლობას და მოიცავს ადმინისტრირების ყველა მოსახერხებელ საშუალებებს.

Oracle Database 10g იძლევა ავტომატური მართვისა და მორგების საშუალებას, რომელიც მის გამოყენებას ხდის მარტივს და ეკონომიურად გამოსადეგს. მისი უნიკალური შესაძლებლობები წარმოების ყველა მონაცემების მართვის განხორციელებაა – ჩვეულებრივი ბიზნესს-ინფორმაციის ოპერაციებიდან მონაცემთა დინამიურ მრავალგანზომილებიან ანალიზამდე (OLAP), XML ფორმატის დოკუმენტებთან ოპერაციები, ლოკალური განაწილებული ინფორმაციის მართვა, რომელიც მას ხდის იდეალურს აპლიკაციების შესასრულებლად, რომლებიც უზრუნველყოფენ ტრანზაქციების ოპერატიულ დამუშავებას, ინფორმაციის ინტელექტუალურ ანალიზს, მონაცემთა შენახვას და ინფორმაციული ნაკადის მართვას.

Oracle Database 10g ზოგიერთი საკვანძო შესაძლებლობები:

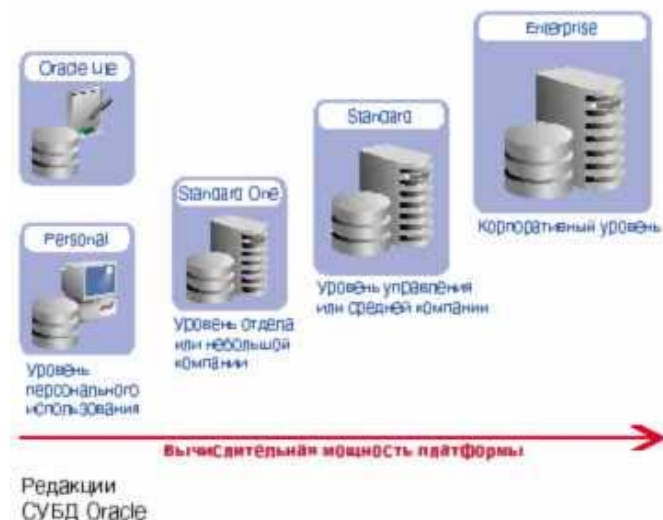
- **Real Application Cluster (RAC)** უზრუნველყოფს მონაცემთა ბაზების ერთი ექსემპლიარის მუშაობას რამოდენიმე grid კვანძებზე. ამავე დროს იძლევა საშუალებას ვმართოთ დატვირთვა და მოქნილად მოვახდინოთ სისტემის მასშტაბირება საჭიროების შემთხვევაში;
- **Automatic Storage Management (ASM)** უზრუნველყოფს ავტომატურ მონაცემთა განაწილებას არსებული

მონაცემთა შენახვის სისტემების რესურსებში, რაც ადაბლებს ძირითად კუთვნილების საფასურს (TCO);

- **წარმოებლობა.** Oracle Database 10g გვაძლევს საშუალებას ვმართოდ სერვისის დონეები და მოვახდინოთ ეტალონური კონფიგურაციების ტირაჟირება მთელი ქსელის არეში;
- **დამუშავების საშუალებები.** HTML DB აპლიკაციების დამუშავების ახალი ინსტრუმენტი ჩვეულებრივ მომხმარებლებს აძლევს ეფექტური აპლიკაციების შექმნის საშუალებას მონაცემთა ბაზებთან მუშაობისათვის მოკლე ვადებში;
- **თვითმართვა.** Oracle Database 10g სპეციალური მექანიზმები გვაძლევს საშუალებას დამოუკიდებლად გადაანაწილონ დატვირთვა სისტემაში, გაუკეთონ ოპტიმიზაცია და მოახდინონ SQL-მოთხოვნების კორექტირება, გამოავლინონ შეცდომები და მოახდინონ შეცდომების პროგნოზირება;
- **მონაცემთა დიდი ბაზები.** ახლა Oracle მონაცემთა ბაზის ეგზემპლარის მაქსიმალური ზომა შესაძლოა აღწევდეს 8 ეკზაბაიტამდე.
- **საშუალო ფასის სერვერული სისტემები.** Oracle Database 10g –თვის შესაძლოა გამოიყენოს საშუალო ფასების ერთპროცესორიანი კომპიუტერები ან მოდულური სისტემები;
- **მონაცემთა ბაზების ახალ ვერსიაში რეალიზებულია** სივრცეში გადასატანი ცხრილების მხარდაჭერა, **Oracle Streams** მონაცემთა ნაკადების მართვის სისტემა და **SQL** მოთხოვნების გადანაწილებული მოდელი. Grid-სფეროში არსებული მონაცემთა ბაზების გადასატანად არ არის საჭირო მათში ცვლილებების შეტანა, რაც საშუალებას იძლევა სწრაფად გამოვიყენოდ Oracle Database 10g-ს ყველა უპირატესობები;

ინფორმაციული სისტემის მასშტაბი და Oracle მონაცემთა მართვის სისტემის რედაქციები.

მონაცემთა ბაზების მართვის სისტემის ბირთვს წარმოადგენს მონაცემთა ბაზების სერვერი. იგი მიეწოდება ოთხიდან ერთ-ერთ რედაქციას (Oracle Database 10g Enterprise Edition, Oracle Database 10g Standard Edition, Oracle Database 10g Standard Edition One, Oracle Database 10g Personal Edition) ინფორმაციული სისტემის მასშტაბთან დამოკიდებულებაში, სადაც განსაზღვრულია მისი გამოყენების არეები (ნახ.5). დიდი ორგანიზაციის მასშტაბებისათვის გამოიყენება Oracle Database Enterprise Edition სისტემა (კორპორატიული რედაქცია), რომლისთვისაც არსებობს მთელი ოპერაციების რიგი, იგი არქიტექტურულად და ფუნქციონალურად აფართოვებს სერვერის შესაძლებლობებს.



ნახ.5 მონაცემთა ბაზების მართვის სისტემა²³

Oracle Database Standard Edition პროდუქტი (სტანდარტული რედაქცია) ორიენტირებულია საშუალო მასშტაბის ორგანიზაციებზე ან/და მსხვილი ორგანიზაციების ქვედანაყოფებზე. Oracle მონაცემთა ბაზის მართვის სისტემის მეათე ვერსიის ფარგლებში გახდა

²³ <http://www.interface.ru/fset.asp?Url=/oracle/news/n010809967.htm>

ხელმისაწვდომი კიდევ ერთი რედაქცია- Standard Edition One, რომელიც შეესაბამება Standard Edition-ის ფუნქციონალურ შესაძლებლობებს, მაგრამ ხელმისაწვდომი ლიცენზირებისათვის კომპიუტერებზე არაუმეტეს ორი პროცესორისა. პერსონალური რედაქცია (Personal Edition) განეკუთვნება, როგორც სახელიდან ჩანს, პერსონალური მოხმარებისათვის. სტანდარტულ და პერსონალურ რედაქციებში ძირითადი აქცენტი გაკეთებულია დაბალ საფასურზე, დაყენების სიმარტივეზე. ამასთან Oracle სერვერის ყველა ვარიანტს თავის საფუძველში უდევს ერთი და იგივე კოდი და ფუნქციონალურად იდენტურნი არიან, დამატებითი მოდულებისა და ოფციების გარდა, რომლებიც აუცილებელნი არიან სპეციფიკური კონფიგურაციებისათვის. ასეთი მიდგომის ძირითადი უპირატესობა, მონაცემთა ბაზების მართვის სისტემის შექმნაში, მდგომარეობს კოდის იდენტურობაში მონაცემთა ბაზების სერვერის ყველა ვარიანტისათვის. ყველა კომპიუტერული პლატფორმისათვის და არქიტექტურისათვის არსებობს ერთიანი Oracle მონაცემთა საცავების მართვის სისტემა, რომელიც მოგვეწოდება სხვადასხვა ვერსიებში და რომელიც გვაწვდის ერთნაირ ბაზურ ფუნქციონალურობას პლატფორმისაგან დამოუკიდებლად, რომელზეც იგი არის დაყენებული.

Oracle მონაცემთა ბაზის მართვის სისტემის ერთ-ერთ ძირითად დახასიათებას წარმოადგენს ფუნქციონირებული სისტემები მრავალ პლატფორმებზე, აგრეთვე დიდ ელექტრონულ გამომთვლელ მანქანებზე, UNIX-სერვერებზე, პერსონალურ კომპიუტერებზე და ა.შ. სხვა მნიშვნელოვან დახასიათებას წარმოადგენს Oracle-ის მიერ ყველა არქიტექტურის შესაძლო ვარიანტების მხარდაჭერა: სიმეტრიული მრავალპროცესორიანი სისტემების, კლასტერების, მასობრივი პარალელიზმის სისტემების, მეინფრეიმების არქიტექტურების. რა თქმა უნდა ნათელია ამ დამახასიათებლების აუცილებლობა თანამედროვე ორგანიზაციებისათვის, სადაც ხდება მრავალი სხვადასხვა მოდულების და მწარმოებლების კომპიუტერების ექსპლუატაცია. ასეთ შემთხვევაში წარმატების ფაქტორს წარმოადგენს წარმოდგენილი მოთხოვნების მაქსიმალური ტიპიზაცია, რომელიც მიზნად ისახავს პროგრამული უზრუნველყოფის თვითღირებულების შემცირებას. მონაცემთა ბაზების

მართვის სისტემების უნიფიკაცია წარმოადგენს ერთ-ერთ აუცილებელ ნაბიჯს ამ მიზნის მისაღწევად.

მრავალი პოპულარული კომპიუტერული პლატფორმებისა და არქიტექტურების არსებობის მიუხედავად, Oracle მხარდაჭერა მიიღწევა მკაცრი მონაცემთა ბაზების მართვის სისტემის კოდის მკაცრი ტექნოლოგიური სქემის დამუშავების საფუძველზე. სერვერული პროდუქტების დამუშავება ხდება Oracle კორპორაციის ერთიანი ქვედანაყოფის მიერ, ცვლილებები შედის ცენტრალიზებულად, ამის შემდეგ ყველა ვერსია გადის მკაცრ ტესტირებას საბაზო ვარიანტში, შემდგომ გადაიტანება ყველა პლატფორმაზე, სადაც აგრეთვე განიხილება დეტალურად. Oracle-ის გადატანის შესაძლებლობა გვეძლევა სათავე მონაცემთა ბაზების პროგრამული კოდის სტრუქტურის მეშვეობით. Oracle პროგრამული კოდის მიახლოებით 80%-„C“ ენაზე დაწერილი პროგრამებია, რომელიც წარმოადგენს პლატფორმა-დამოუკიდელს. კოდის მიახლოებით 20%, რომელიც წარმოადგენს მონაცემთა ბაზების დამუშავების მართვის სისტემის ბირთვს, რეალიზებულია მანქანაზე დამოკიდებულ ენებზე, და ამ კოდის ნაწილი მუშავდება სხვადასხვა პლატფორმაზე. Oracle მონაცემთა ბაზების მართვის სისტემა მოიცავს მონაცემთა მართვის რეალიზაციის მექანიზმების დეტალებს, რაც გვაძლევს უფლებას ვისაუბროთ პროგრამული უზრუნველყოფის პრაქტიკულად მთლიანი ბაზის უნიფიკაციაზე. ამასთან, Oracle გვაძლევს თანდართული სისტემების (რომლებიც რეალიზებულნი არიან ერთ პლატფორმაზე) სხვა პლატფორმაზე უცვლელად გადატანის საშუალებას, როგორც მონაცემთა ბაზების სტრუქტურაში, ასევე აპლიკაციების კოდებში.

Oracle მონაცემთა ბაზების მართვის სისტემა ერთ დონეზე არის ოპტიმიზირებული როგორც აპლიკაციების ტრანზაქციების ოპერატიული დამუშავების, ასევე ანალიტიკური აპლიკაციებისათვის. პრაქტიკაზე ეს ნიშნავს, რომ ერთი და იგივე პროდუქტი (მაგალითად, Oracle Database Enterprise Edition) შესაძლებელია გამოვიყენოთ, როგორც ოპერატიული მონაცემთა ბაზების სერვერი, რომელიც ამუშავებს მოკლე და გაცილებით მარტივი ტრანზაქციების ინტენსიურ ნაკადს, მომხმარებლებისაგან შემოსულს, ასევე მონაცემთა საცავების სერვერის

სახით, რომელიც დიდი მოცულობის მონაცემების კონცენტრირების და მათზე რთული ანალიტიკური გამოთვლების ჩატარების საშუალებას იძლევა.

Oracle ეყრდნობა SQL-3 სტანდარტს, რომელიც გვაძლევს ახალი ტიპის ობიექტების აღწერვის საშუალებას და რომელიც შედგება ატრიბუტებისაგან (სკალარული-ანუ სხვა ტიპების, მრავალი ობიექტების, ობიექტებზე მიმართვის) და მათთან ასოცირებული მეთოდების საშუალებას. ცხრილის ყოველივე მწკრივს შეუძლია შეიცავდეს ბაზურ მონაცემებს და რთულ ტიპებს, აგრეთვე გააჩნია თანდართული ცხრილების მხარდაჭერა და გააჩნია ცვალებადი სიგრძის ობიექტების მასივები.

Oracle-ის ერთ-ერთი განმასხვავებელი ნიშანია სხვადასხვა წინასწარ განსაზღვრული ტიპების მონაცემების შენახვა და დამუშავება. მოცემული ფუნქციონალურობა ინტეგრირებულია მონაცემთა ბაზების მართვის სისტემის ბირთვში და Oracle Database-ში შემავალი interMedia მოდულის მხარდაჭერით. ის უზრუნველყოფს მუშაობას ტექსტურ დოკუმენტებთან, სხვადასხვა ძეგლის საშუალების მეშვეობით, ასევე კონტექსტური; 2D-ზე მეტ გრაფიკულ ფორმატთან მუშაობა; აუდიო და ვიდეო ინფორმაციასთან მუშაობა. Oracle მონაცემთა ბაზების მართვის სისტემა წარმოადგენს ჩაშენებული ტიპების მონაცემების ერთობლიობას, ამასთან ერთად იძლევა ახალი ტიპის მონაცემების კონსტრუირების საშუალებას მათთანვე სპეციფიკური მეთოდების წვდომის წარმოდგენით. ეს იმას ნიშნავს, რომ დამმუშავებლები უპირველეს ყოვლისა ღებულობენ სისტემას არა მარტო მონაცემების შენახვისათვის და ატრიბუტული მონაცემების დამუშავებისათვის ცხრილების სახით, არამედ ინსტრუმენტს, რომელიც აძლევთ საშუალებას შექმნან სტრუქტურირებული მონაცემთა ტიპები, რომლებიც გვიჩვენებენ საგნობრივი სფეროს არსს.

InterMedia მოდული უზრუნველყოფს ყველა ტიპის მონაცემების მხარდაჭერას, ასევე ძეგლის ოპერაციების შესრულებას სხვადასხვა ტიპის ფორმატების ტექსტურ დოკუმენტებში.

Oracle Enterprise Manager კომპონენტი წარმოადგენს მონაცემთა ბაზების ადმინისტრირების უნივერსალურ საშუალებას, რომელსაც

გააჩნია მოხერხებული გრაფიკული ინტერფეისი და ბაზების ადმინისტრატორს აძლევს მრავალი Oracle მონაცემთა ბაზების ფართო სპექტრის ამოცანების შესრულების საშუალებას. ამასთან მოიცავს შექმნის, მოდიფიკაციის და მასში მეოფი ობიექტების წაშლის საშუალებას.

Advanced Replication Option მოდული გვაძლევს მონაცემების რეპლიკაციის საშუალებას შესაძლებლობების ფართო დიაპაზონში, სინქრონული, ასინქრონული, კასკადური და სხვა ტიპის რეპლიკაციების თანდართვით.

მერვე ვერსიიდან დაწყებული, Oracle მონაცემთა ბაზების დამუშავების სისტემა წარმოადგენს ობიექტ-რელაციურ სისტემას. **Objects Option** მოდულს გააჩნია ობიექტზე-ორიენტირებული შესაძლებლობების მხარდაჭერა: ობიექტური ტიპები, კოლექციები, მასივები, ჩაშენებული ცხრილები, წვდომები ობიექტებზე და დიდ ბინარულ ობიექტებზე (BLOB).

64 Bit Option მოდულის Oracle სერვერში ჩაშენების გამო, Oracle Database მუშაობს არა მარტო 32-ბიტიანი, არამედ 64-ბიტიანი კომპიუტერებთანაც, რაც საკმაოდ აფართოებს მის შესაძლებლობებს, მონაცემთა დამუშავების სიჩქარეში და ასევე დამუშავებული მონაცემების ზომებში.

Oracle Advanced Quering (AQ) წარმოადგენს Oracle Database –ში ჩაშენებული შენახვის მექანიზმს და შეტყობინებების რიგის დამუშავებას. AQ კომპონენტი **MOM (Message Oriented Middleware)** კლასს მიეკუთვნება. ასეთი კომპონენტის არსებობა გვაძლევს სრულ ფუნქციონალური სერვერის ბაზაზე ინფრასტრუქტურის შექმნის შესაძლებლობას შეტყობინებების დამუშავებისათვის და გამორიცხავს დამატებით მესამე ფორმების საშუალებების შეძენას ამ მიზნებისათვის, და ამავე დროს უზრუნველყოფს მათთან კავშირს **Oracle Messaging Gateway** პროდუქტის მეშვეობით. AQ გვთავაზობს შეტყობინებების გადაგზავნის ორ რეჟიმს: პირველი- **point-to-multipoint** და მეორე **publish/subscribe**. AQ გვაძლევს შეტყობინებების პრიორიტეტების მიცემის საშუალებას, ასევე შეტყობინებების მიცემის რიგს (**FIFO** ან და

პრიორიტეტების საფუძველზე), შეტყობინებების დაჯგუფება, შეტყობინებების მიწოდების განსაზღვრა და მათი არსებობის დრო, შეტყობინებების ფორმატის ავტომატურად გარდაქმნა, ელექტრონული ფოსტით ასინქრონული შეტყობინების მიღება და შეტყობინებების გადაცემა HTTP(S) საშუალებით. Oracle8i ვერსიიდან დაწყებული სერვერს გააჩნია ვირტუალური Java-მანქანა (JServer Enterprise Edition).

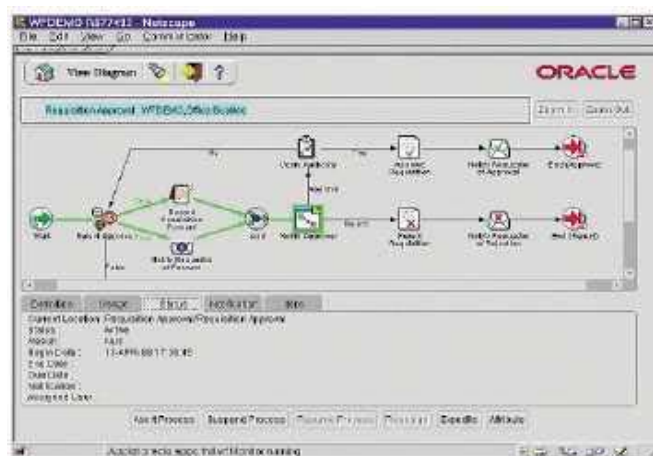
Oracle Database გააჩნია ყველა საჭირო საშუალება კლიენტების სამუშაო ადგილების მისაერთებლად Net8 პროტოკოლის მიხედვით (**Networking Kit** მოდული), აგრეთვე OLE ტექნოლოგიით კლიენტების მუშაობის უზრუნველყოფა (**Objects for OLE** მოდული), ODBC-დრაივერების ერთობლიობა (**ODBC Driver**) და პროგრამების დამუშავებისათვის ბიბლიოთეკები მესამე დონის ენებზე, რომლებიც მონაცემთა ბაზებთან წვდომისათვის გამოიყენება Oracle Call Level Interface (OCI). **Oracle Call Interface** უზრუნველყოფს პროგრამების დამუშავებას დაბალი დონის საფეხურების გამოძახების გამოყენებით მონაცემთა ბაზებთან წვდომისათვის. ეს გვაძლევს ეფექტური პროგრამების შექმნის საშუალებას, რომლებიც მოითხოვენ მინიმალურ რესურსებს. აპლიკაციების დამუშავების შესაძლებლობა, რომლებიც ოპტიმიზირებულნი არიან სიჩქარით და გამოყენებული მესხიერებით, მიიღწევა ფუნქციების გამოძახების გამოყენებით, რომელიც გვაძლევს SQL-სა და PL/SQL-ის ოპერატორების მოქმედების სრულ კონტროლს.

Oracle Objects for OLE კომპონენტი გავძლევს მონაცემთა ბაზებთან Oracle აპლიკაციებთან წვდომის საშუალებას, რომლებიც დამუშავებულნი არიან C++, Microsoft Visual Basic, OLE 2.0-ზე. Visual Basic ენის მაკროგანსაზღვრების სრული მხარდაჭერა იძლევა Oracle მონაცემთა ბაზებიდან მონაცემების მიღების საშუალებას, უშუალოდ Microsoft Excel-ის ცხრილების სახით.

Oracle Workflow- ეს არის ორგანიზაციის ბიზნესპროცედურების სტანდარტული ავტომატიზაციის საშუალება, რომელიც ორიენტირებულია კორპორატიული აპლიკაციების დამუშავებლებზე. Oracle ტექნოლოგიებზე დაფუძნებული. Oracle Workflow წარმოგვიდგენს ინფრასტრუქტურას და პროექტირების საშუალებას (Workflow Builder)

ყველა ტიპის ინფორმაციის ავტომატიზაციისათვის, როული ბიზნესწესების ფორმალიზაციისათვის და მომხმარებლის ჩართვას როული გადაწყვეტილებების მიღების პროცესში.

აპლიკაციების დამუშავება ნაკადების მართვისათვის იწვება Workflow Builder გრაფიკულ გარემოში ალგორითმის პროექტირებით (ნახ.6).



ნახ.6 Workflow Builder გრაფიკული არე

პროცესი შედგება სტანდარტული მოქმედებისაგან, ისეთი როგორიც არის შესვლის წერტილები, გამოსვლის, დანაწევრების, შეტყობინების, ჩაშენებული პროცესის, კონკრეტული აპლიკაციებისათვის სპეციფიკური მოქმედებების, რომელთა ფუნქციონალურობა განისაზღვრება დამუშავებლებით. ამის შემდეგ, როდესაც პროცესების აღწერა შენახულია რეპოზიტორიებში, ისინი შესაძლოა გამოყენებულ იქნან პროგრამული ინტერფეისით აპლიკაციებისათვის. დამატებით საშუალებას წარმოადგენს საფოსტო შეტყობინებების დაგზავნა პროცესების შესრულებისა და მუშაობის შესახებ და მომხმარებლების პროცესებთან ინტერაქტიული ურთიერთქმედების ფორმების წარმოდგენისა და პროცესების მუშაობის რეზულტატების წარმოდგენა.

Oracle Database Lite (ODL)- პროგრამული პროდუქტი წარმოდგენილია მობილური აპლიკაციების სისტემის

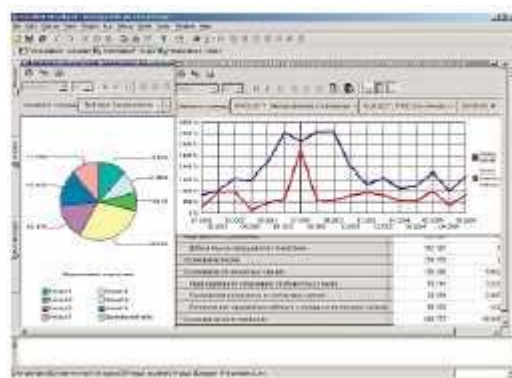
ინფრასტრუქტურის შესაქმნელად. პროდუქტის ჯგუფში შედის დამუშავებისათვის ყველა საჭირო კომპონენტი, დაყენება და აპლიკაციების მართვა ყველა მობილური მოწყობილობებისათვის ყველა პოპულარულ ოპერაციულ სისტემაზე: Linux, Unix, Palm OS, Microsoft Windows CE/PPC, Microsoft Windows NT/2000/XP. წარმოდგენილი ინფრასტრუქტურის ძირითად ამოცანას წარმოადგენს მონაცემთა უსაფრთხო და დაცულ სინქრონიზაციას Oracle Database კორპორატიულ მონაცემთა ბაზასა და მობილურ კლიენტებს შორის. სინქრონიზაციის პირველი სეანსის შემდეგ მომხმარებლები რომლებიც მუშაობენ კომპიუტერებთან, სადაც არ იყო დაყენებული არანაირი სპეციალური პროგრამული უზრუნველყოფა, დებულობენ მუშა აპლიკაციებს და ODL მონაცემთა ბაზებს აქტუალური კორპორატიული მონაცემებით. შემდეგი კავშირების სეანსების დროს მომხმარებლებს გადაეცემა მხოლოდ შეცვლილი ინფორმაცია. ODL არც თუ ისე დიდი, მაგრამ სრულფუნქციონალური რელაციური ბაზა გააჩნია, რომელიც პროექტირებულია სპეციალურად მობილური საშუალებების მუშაობისათვის, რომელშიც მთლიანად არის რეალიზებული ტრანზაქციების მექანიზმები, მიმართვის ერთობლიობა და SQL ენის სპეციფიკაცია.

ბიზნესლოგიკა, შენახული პროცედურები და ტრიგერები მუშავდება Java-ზე. Mobile Server Oracle AS 10g-ს გაფართოებაა. ეს კომპონენტი უზრუნველყოფს მობილური აპლიკაციების ურთიერთქმედებას Oracle Database 10g-სთან ან და სხვადასხვა ინტერნეტ-აპლიკაციებთან. მონაცემთა სინქრონიზაციის დროს კავშირის გაწყვეტისას ინფორმაციის გადაცემა მობილურ მოწყობილობებზე აღდგება კავშირის აღდგენის შემდეგ იმ წერტილიდან, სადაც გაწყდა. Mobile Server-ის გამოყენება გარანტირებულს ხდის მონაცემთა მიწოდებას. ინფორმაცია, რომელიც გადაეცემა ქსელში და ინახება მონაცემთა ბაზაში, შესაძლოა დაშიფრული იყოს FIPS-140 ალგორითმის მიხედვით, რომელიც აკმაყოფილებს AES. მონაცემთა სინქრონიზაცია Oracle Lite 10g –სა და Oracle Database მონაცემთა ბაზების შორის ხორციელდება TCP/IP, HTTP, CDPD, 802.11b Wireless LAN, PPP, GPRS,

HotSync, ActiveSync პროტოკოლების მეშვეობით. Open Transport API გვაძლევს საშუალებას გამოვიყენოთ ქსელის გარეშე სატრანსპორტო პროტოკოლი სინქრონიზაციისათვის. მობილური აპლიკაციები მუშავდებიან Mobile Development Kit-ის მეშვეობით C, C ++, Java, Visual Basic პროგრამირების ენებზე, ActiveX Data Objects (ADO) გამოყენებით Oracle JDeveloper 10g, Microsoft Visual Studio.Net 2003, Microsoft EVT 3.0, Borland Delphi, Sybase Power Builder, Metroworks CodeWarrior 8+, Rrapid Software Formation ინსტრუმენტალურ არეებში. აპლიკაციებს, რომლებიც მუშაობენ მობილურ მოწყობილობებზე გააჩნიათ წვდომა Oracle Lite 10g – სთან სხვადასხვა პროგრამული ინტერფეისების გამოყენებით (JDBC, ODBC, ADOCE, ADO.Net, SODA Stateless Object Database Access).

მრავლგანზომილებიანი ვიტრინებისათვის გამოიყენება **Oracle OLAP**, რომელიც განუწყვეტლივ არის დაკავშირებული Oracle Database მონაცემთა ბაზების დამუშავების სისტემასთან.

Oracle OLAP ოპცია წარმოდგენილია ანალიტიკური სისტემების შექმნისათვის, რომლებიც აწეობილია მრავალგანზომილებიანი ანალიზის პრინციპზე და OLAP ტექნოლოგიაზე (ნახ.7). ეს კომპონენტი გვაძლევს შენახვის და დამუშავების საშუალებას ერთი და იგივე მონაცემთა ბაზებში, არამარტო რელაციურისათვის, არამედ მრავალგანზომილებიანი ინფორმაციისათვის.



ნახ.7 ანალიტიკური სისტემების შექმნა, რომლებიც შექმნილი არიან მრავალგანზომილებიანი ანალიზისა და OLAP ტექნოლოგიის საფუძველზე

ოპციის საფუძველს წარმოადგენს სრულმასშტაბიანი მრავლგანზომილებიანი მონაცემთა მოდელი, რომლის მეშვეობითაც

შესაძლოა განსაზღვრა და მუშაობა ისეთ მცნებებთან როგორიც არის განზომილება, იერარქია, მრავალგანზომილებიანი მაჩვენებლები და ა.შ. OLAP Catalog-ის რეპოზიტორიებში ყველა ასეთი ობიექტების განსაზღვრა და აღწერა ინახება, ხოლო თვით მონაცემები კი შესაძლებელია ინახებოდეს როგორც რელაციურ ბაზაში, ასევე სპეციალიზირებულ მრავალგანზომილებიან სტრუქტურაში. OLAP აპლიკაციების შესაქმნელად ოპციის სტრუქტურაში შედის Java OLAP API, Java კლასების ერთობლიობა, რომელიც აწარმოებს ყველა ოპერაციას შექმნისათვის, მოდიფიკაციისათვის და მრავალგანზომილებიანი ობიექტების მანიპულირებისათვის მათი შენახვის წესის გარეშე. აგრეთვე წარმოადგენს ფართო მათემატიკურ, სტატიკურ და ფინანსური ფუნქციების ერთობლიობას, რომელიც წარმოგიდგება ისეთი ამოცანებისათვის, როგორცაა პროგნოზირება, მოდელირება, სიტუაციების დაგეგმვა და შეფასება. მრავალგანზომილებიან ინფორმაციასთან წვდომა შესაძლებელია Java აპლიკაციებიდან, აგრეთვე SQL-ზე ორიენტირებულ სპეციალურ პროცედურებიდან.

Oracle Database-ში ჩაშენებული OLAP საშუალებები უზრუნველყოფს მაღალი დონის უსაფრთხოებას, მაშტაბირებას და უკეთ მუშაობას ერთდროულად მრავალი მომხმარებლისათვის. OLAP ტექნოლოგიის მთლიანი ინტეგრაციის ძირითად უპირატესობას Oracle მონაცემთა ბაზების რელაციურ სერვერთან წარმოადგენს ერთიან ადმინისტრირების მიდგომას, როგორც რელაციურ, ასევე მრავალგანზომილებიან მონაცემთა ბაზებთან, მომხმარებლების მართვით და წვდომის უფლებების მინიჭებით, მოთხოვნების ოპტიმიზაციის შესრულებით, კლასტერული ტექნოლოგიების გამოყენებით, პარალელირების საშუალებებით, სანქციონირებით და აგრეგირებული მონაცემების მართვით. Oracle OLAP-ის საფუძველზე ანალიტიკოსების, ხელმძღვანელების და სხვა საბოლოო მომხმარებლების სამუშაო ადგილების შესაქმნელად გამოიყენება Oracle JDeveloper-ის სპეციალური სფერო, სპეციალური Oracle Business Intelligence Java Beans სპეციალური კომპონენტით. ამის გარდა,

მრავალგანზომილებიან ობიექტებთან OLAP ოპერაციებთან მომხმარებლებს თავისუფლად შეუძლიათ იმუშაონ MS Excel-დან ან/და Oracle Discoverer-ის მეშვეობით.

Oracle ფირმის პლატფორმა მონაცემთა საქმიანი ანალიზის ინსტრუმენტების ნაწილში განსხვავდება არამარტო სამაგიდო ინსტრუმენტალური საშუალებებით, წინასწარ განსაზღვრული ინტერფეისით და ფუნქციონალური შესაძლებლობების ერთობლიობით, არამედ ძლიერი ინსტრუმენტებით საბოლოო მომხმარებლების კლიენტური ადგილების დამუშავებისათვის. ეს იძლევა ზუსტი ანალიტიკური აპლიკაციების შექმნის საშუალებას ყოველივე ცალკეული მომხმარებლისათვის.

რეგლამენტირებული რეპორტების პლატფორმას Oracle-ში წარმოადგენს **Oracle Reports** პროდუქტი - ვიზუალური დამუშავების საშუალება და სტანდარტული რეპორტების ფორმების პუბლიკაცია. ვიზუალური ინტერფეისი და ჩაშენებული wizard-ები გვაძლევს რთული რეპორტების დამუშავების დროის შემცირების საშუალებას. რეპორტების ფორმების Oracle Reports-ის საფუძველზე შექმნა ხორციელდება კვალიფიცირებული სპეციალისტების მეშვეობით, ხოლო უკვე ფორმირებული რეპორტები შესაძლებელია გაეშვას საბოლოო არაპროგრამისტ მომხმარებლების მიერ.

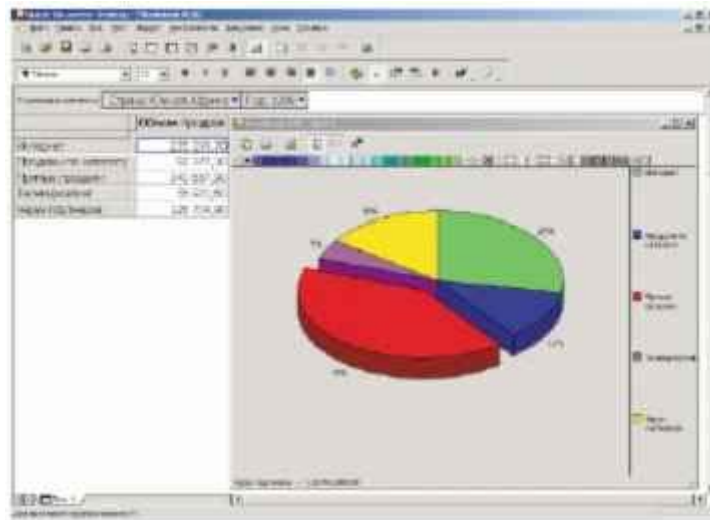
რეპორტების ვიზუალური პროექტირება და გენერაცია შესაძლებელია განხორციელდეს სხვადასხვა პროგრამულ პლატფორმაზე. მომხმარებლები Report-ის დამუშავების შედეგს დებულობენ ნაბეჭდი სახით, ან და HTML, PDF, RTF , XML სტანდარტული ფორმატის ელექტრონული ფაილების სახით, რომლებიც მიწოდებულნი არიან ელექტრონული ფოსტით, ან ფაილ-სერვერებზე ინახებიან. ოპერატიულ რეჟიმში მომხმარებელთა წვდომა რეპორტებთან შესაძლოა ორგანიზებულ იქნას Web-ის შიდა ან გარე ქსელში. ამ შემთხვევაში საკმარისია ჩვეულებრილი ნავიგატორი, იმისათვის რომ მივიღოთ მზა რეპორტები. რეპორტები აგრეთვე შესაძლოა

ფორმირდებოდეს წინასწარ დაგეგმილი მოვლენის წარმოქმნისას, მაგალითად მონაცემთა ბაზაში ახალი ინფორმაციის მიღებისას.

Reports ტექნოლოგია JSP-ს იყენებს (Java Server Pages) მომხმარებლისათვის მოსახერხებელ ფორმებში. შექმნილ რეპორტზე ბმული შესაძლოა მოთავსდეს ყველანაირ Web-გვერდზე. უნდა აღინიშნოს, რომ Reports Developer-ი არ მოითხოვს დამმუშავებლებისაგან სხვადასხვა მიდგომების გამოყენებას რეპორტების ფორმირებისათვის, რომლებიც სრულდებიან სხვადასხვა გარემოში. ერთხელ შექმნილი რეპორტი შესრულდება მრავალნაირ პლატფორმაზე, ხოლო მისი გამომავალი ფორმები კი წარმოიდგინება სხვადასხვა ფორმატებში-ტექსტური ფაილებიდან, დინამიურ Web გვერდებამდე. თუ რამოდენიმე მომხმარებელი ერთი და იგივე მონაცემებს მოითხოვს ერთდროულად, რეპორტები არ იქმნება თავიდან ყოველი მოთხოვნისას. მომხმარებელს მიეწოდება პირველი რეპორტის ასლი აპლიკაციების სერვერის ბუფერიდან. Oracle დამუშავების საშუალებების ტრადიციების მიხედვით, Reports Developer-ი მჭიდროდ ინტეგრირებულია Oracle Database-სთან და იყენებს დამუშავების ერთსადაიმთხვე ენებს. ამავე დროს Reports გვაძლევს საშუალებას მივიღოთ მონაცემები Express Server მონაცემთა ანალიტიკური დამუშავების სერვერიდან, აგრეთვე სხვა ინფორმაციული წყაროებიდან - სტანდარტული ინტერფეისებიდან (ODBC და JDBC).

არარეგლამენტირებულ რეპორტებზე პასუხს აგებს პროდუქტი Oracle Discoverer – თავისუფალი ტიპის რეპორტების ფორმირებისათვის, არარეგლამენტირებული მოთხოვნებისათვის და მონაცემთა ანალიზისათვის (ნახ.8). ის გვაძლევს რელაციურ საცავეებს, მონაცემთა ვიტრინებში ინფორმაციასთან წვდომის საშუალებას და ტრანზაქციულ სისტემებში, არც თუ ისე აუცილებელია Oracle მონაცემთა ბაზის მართვის სისტემის მეშვეობით იმუშაოს. იმისათვის რომ საბოლოო მომხმარებელ არაპროგრამისტს შეეძლოს მისთვის საინტერესო ინფორმაციების ნაწყვეტების ნახვა, პროგრამულ პროდუქტში გამოიყენება მეტამონაცემების ნაწილი, რომელიც აღიწერება დამმუშავებლის მიერ, და რომელიც იძლევა მომხმარებლისაგან

მუშაობის სიძნელეების დაფარვის საშუალებას პროგრამულ დონეზე, ინფორმაციის საქმიან ენაზე წარმოდგენის მეშვეობით.

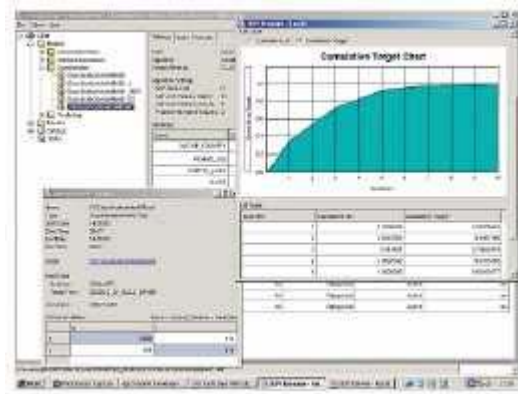


ნახ.8 თავისუფალი ფორმის რეპორტების ფორმირება Oracle Discoverer-ში

Oracle OLAP წარმოდგენილია ანალიტიკური სისტემების შესაქმნელად, რომლის საფუძველს წარმოადგენს მრავალგანზომილებიანი ტექნოლოგიები. ეს კომპონენტი გვაძლევს ერთი და იგივე მონაცემთა ბაზაში რელაციური და მრავალგანზომილებიანი ინფორმაციის შენახვის და დამუშავების საშუალებას. Oracle9i Java OLAP API საშუალებები იძლევა მრავალგანზომილებიანი მოდელის მხარდაჭერის საშუალებას, მატემატიკური, სტატისტიკური და ფინანსური ფუნქციების წარმოდგენით პროგნოზირების ამოცანების გადასაწყვეტად, სიტუაციების შესაფასებლად და ტენდენციების გამოსავლენად. ძირითადად OLAP API წარმოადგენს Java ობიექტ-ორიენტირებულ პროგრამულ ინტერფეისს OLAP მოთხოვნებისათვის. OLAP აპლიკაციების დამუშავებლებს კომპანია Oracle Web-სათვის წარმოადგინეს Oracle JDeveloper პროგრამულ მოდული Oracle Business Intelligence Beans კომპონენტით.

Oracle Data Mining (წარსულში - Darwin-ი) ინსტრუმენტალური არე წარმოდგენილია მონაცემთა ანალიზისათვის, მონაცემთა მოძიების ტექნოლოგიის მეთოდებით. Oracle Data Mining-ი მხარს უჭერს მონაცემთა

მოკრების ტექნოლოგიების ისეთ ეტაპებს, როგორიც არის ამოცანის დასმა, მონაცემების მომზადება, მოდელების ავტომატური შექმნა, ანალიზი, რეზულტატების ტესტირება და მოდელების გამოყენება რეალურ აპლიკაციებში (ნახ.9).



ნახ.9 Oracle Data Mining ინსტრუმენტალური არე

Oracle Data Mining (ODM) ოპცია წარმოდგენილია მონაცემთა ანალიზისათვის, ცოდნის აღმოფხვრის ტექნოლოგიების მეთოდებით. ამ ტექნოლოგიის ძირითადი ამოცანაა ფარული კანონზომიერებების, დამოკიდებულებების და კავშირების აღმოჩენა მონაცემთა დიდ ნაკრებში, რომლებიც საჭიროა გადაწყვეტილებების მისაღებად მართვის სხვადასხვა დონეებზე. ასეთი კანონზომიერებები წარმოიშობება სხვადასხვა ტიპების მოდელებით, რომელიც იძლევა სიტუაციების ან ობიექტების კლასიფიკაციის ჩატარების საშუალებას, მათი ქმედებების პროგნოზირებას, საერთო ჯგუფის ობიექტების გამოვლენას და ა.შ. მოდელები იქმნება ავტომატურად ობიექტებზე მოცემული მონაცემების ანალიზის საფუძველზე, დაკვირვებების და სიტუაციებს სპეციალური ალგორითმების საშუალებით. ცოდნის აღმოფხვრის მეთოდები გამოიყენება სხვადასხვა სფეროებში ისეთი პრაქტიკული ამოცანების გადასაწყვეტად, როგორიც არის, ახალი კლიენტების ძიება, მათი მოზიდვა, მათი შენარჩუნება, ეფექტურობის ამაღლება მარკეტინგულ საქმიანობაში პროდუქტებისა და მომსახურების წინსვლით, აპარატურის შეცდომების აღმოფხვრა და ა.შ. ოპციის საფუძველს წარმოადგენს პროცედურები, რომლებიც ქმნიან კლასიფიკაციის

მოდელის შექმნის ალგორითმებს, რეგრესიას, კლასტერიზაციას. Data Mining 10g ვერსია მხარს უჭერს ასეთი ალგორითმების ფართო სპექტრს (ცხრილი 3).

კლასიფიკაცია და რეგრესიული მოდელები	Support Vector Machine
არსებული ატრიბუტების ძიება	Minimal Descriptor Length
კლასტერიზაცია	Enhanced K-means, O-cluster
ასოციაციების ძიება	Apriori Algorithm
ნიშნების გამოყოფა	Non-Negative Matrix Factorization

ცხრილი 3. Oracle Data Mining –ში რეალიზებული ალგორითმები

ალგორითმების ერთ-ერთ ძირითად თავისებურებას წარმოადგენს ის, რომ ისინი მუშაობენ რელაციურ მონაცემთა ბაზებთან და არ საჭიროებენ მონაცემების აღმოფხვრას და შენახვას სხვადასხვა ფორმატებში. ალგორითმების გარდა, ODM-ში შედის მონაცემების მომზადების სფერო, რეზულტატების შეფასების და ახალი მოდელების გამოყენება ახალ მონაცემთა ნაკრებისათვის. ამ ყველა საშუალებების გამოყენება შესაძლებელია როგორც პროგრამულ Java API, PL/SQL API დონეზე, ასევე ODM Client გრაფიკული საშუალებით, რომელიც ორიენტირებულია ანალიტიკოსების მუშაობაზე და რომლებიც დაყრდნობილია პროგნოზირების ამოცანების გადაწყვეტაზე, ტენდენციების გამოვლენაზე, სეგმენტაციაზე და ა. შ. Oracle Portal პროგრამული პროდუქტი გვაძლევს საშუალებას წარმოვიდგინოთ საქმიანი ანალიზის ცალკეული კომპონენტები Web-გარემოში, როგორც ერთიანი კომპონენტი. Oracle Portal წარმოდგენილია ორგანიზაციის ინფორმაციული პორტალის სწრაფი გაშლისათვის და გამოიყენება როგორც მზა გადაწყვეტილება. მისი გამოყენების მთავარ მიზანს წარმოადგენს –მოცულობის შემცირებას ინფორმაციული პორტალის მზა კომპონენტების გამოყენებით, ხელით პროგრამირებისას, ასევე ვიზუალური ინტერაქტიული დამუშავების საშუალებები (wizards). Oracle Portal-ში გვაქვს შესაძლებლობა შევქნათ პორტალები და შევიმუშავოთ

მათი მხარდაჭერა Web დიზაინერების გარეშე. პლატფორმულ-ბაზირებული გადაწყვეტილებების საფუძველზე შექმნილი ინფორმაციულ-ანალიტიკური სისტემის რიგი უპირატესობების მიუხედავად საჭიროა შევნიშნოთ, რომ არსებობს რიგი შეზღუდვები, როდესაც ინფორმაციულ-ანალიტიკური სისტემის შექმნისათვის ასეთი მიდგომა მიუღებელია. იშვიათად, თუ ორგანიზაცია ინფორმაციულ-ანალიტიკურ სისტემას ქმნის ნულიდან. რეალურ პირობებში კომპლექსური გადაწყვეტილებების დროს ფირმის შემქმნელები ხშირად ხვდებიან ისეთ პრობლემას როგორიც არის უკვე არსებული სიტემების ნაწილების შენარჩუნების სურვილი, რომლებიც სხვადასხვა დროსა და პირობებში იყო შექმნილი. ძალიან ხშირად შეუძლებელია ძველ სისტემებზე უარის თქმა. ამავე დროს, სხვა სისტემებზე გადასვლა დაკავშირებულია დიდ თანხებთან. ამის გარდა ერთი შემქმნელის კომპლექსური გადაწყვეტილებები მონაცემთა ბაზების მართვის სიტემაში დღევანდელ დღეს დამოუკიდებელი არ არის. ეს იხსნება შემდეგით: ინფორმაციულ-ანალიტიკური სისტემისათვის პროგრამული უზრუნველყოფის ძირითადი შემქმნელები მიისწრაფიან მათ მიერ წარმოდგენილი გადაწყვეტილებების მაქსიმალური ინტეგრაციისაკენ. ამის გამო ერთი ან რამოდენიმე ინსტრუმენტის გამოყენების სურვილი ორგანიზაციას აიძულებს ამ მომწოდებლის სხვა პროდუქტების გამოყენებასაც, რაც ხშირად არ შეესაბამება დამკვეთის მოთხოვნას. მაგალითად, საქმიანი ანალიზის ინსტრუმენტები, შეუძლებელია არ იყოს ინტეგრირებული მონაცემთა დამუშავების მართვის სისტემასთან.

ნაშრომში რამოდენიმეჯერ უკვე ნახსენები იყო ტერმინი Data Mining, რომელიც ჩვენ უკვე მოკლედ აღვწერეთ. მოდით კიდევ ერთხელ განვიხილოთ, თუ რა არის ეს, Data Mining პრობლემის მდგომარეობა და ახალი გადაწყვეტილებები.

II თავი. DATA MINING-ტექნოლოგიის არსი და მისი გამოყენების სფეროები

2.1 Data Mining-ის შინაარსი

პირველ თავში ჩვენ აღვწერეთ თუ რა სისტემები და პროდუქტები არსებობს ინტელექტუალური ანალიზისათვის. როგორც უკვე ვთქვით ერთ-ერთი Data Mining-ა.²⁴ Data Mining ითარგმნება როგორც “მონაცემთა მოძიება”. Data Mining ხშირად აგრეთვე გამოიყენება (knowledge discovery in database) “მონაცემთა მოძიება მონაცემთა ბაზაში”-ს სინონიმად. ეს ახალი ტოტია მონაცემთა დამუშავებასა და მოძიების სფეროში.

Data Mining-ის მიზანი არის დაფარული მონაცემებისა და კანონზომიერებების აღმოფხვრა მონაცემთა შეგროვებისას.

საქმე იმაშია, რომ ადამიანის გონება არ არის მიმართული სხვადასხვა სახის ინფორმაციის მასივების მიღებაზე. ამის გარდა ინფორმაციის აღმოფხვრისას ადამიანის გონებას მხოლოდ ორი-სამი დამოკიდებულების აღმოჩენის საშუალება აქვს. აგრეთვე მათემატიკური სტატისტიკა დროის დიდი პერიოდის განმავლობაში ვერ ასრულებდა რეალურად რთული ცხოვრებისეული ამოცანების გადაწყვეტას, მიუხედავად იმისა, რომ ის ითვლებოდა ძირითად მონაცემთა ანალიზის ინსტრუმენტად. ის ოპერირებს გასაშუალოებული მონაცემებით, რომლებიც წარმოადგენს ფიქტიურ საზომს (მაგალითად, საავადმყოფოში პაციენტების საშუალო ტემპერატურა, ქუჩაზე სახლის საშუალო სიმაღლე, და ა.შ.). ამის გამო მათემატიკური სტატისტიკის მეთოდები სარგებლიანია მხოლოდ წინასწარ ფორმულირებული ჰიპოთეზების გადასამოწმებლად (verification-driven data mining).

თანამედროვე Data Mining ტექნოლოგიები (discovery-driven data mining) ათვალიერებს მონაცემებს ავტომატური შაბლონების (პატერნების) ძებნის მიზნით. ისინი სახასიათოა რაიმე მრავალგანზომილებიანი არაერთგვაროვანი მონაცემების ფრაგმენტებისათვის. ოპერატიული, ანალიტიკური მონაცემების დამუშავებისაგან განსხვავებით (online analytical processing, OLAP) Data

²⁴ Глоцарий// <http://www.rdtex.ru/docs/glossary/D57123.html>

Mining-ში ჰიპოთეზების ფორმულირების დრო და უჩვეულო (unexpected) შაბლონების აღმოჩენა გადატანილია ადამიანიდან კომპიუტერზე (ცხრილი 4).

OLAP	Data Mining
როგორია ტრანზაქციის საშუალო მაჩვენებელი მწველთათვის და არა მწველთათვის?	რომელი ფაქტორები უკეთ წინასწარმეტყველებენ უბედურ შემთხვევებზე?
როგორია საშუალო მაჩვენებელი ყოველ-დღიური შენაძენებისა მოპარული და არა მოპარული საკრედიტო ბარათებით?	რა სქემებია სახასიათო საკრედიტო ბარათებით თაღლითობისთვის?

ცხრილი. 4 OLAP და Data Mining – ში ამოცანების ფორმულირების და მეთოდების გამოყენების მაგალითები²⁵

პრინციპში Data Mining –ის ამოცანის დასმაში ახალი არაფერია. ტექნოლოგიების განვითარებამ მოიტანა კოლოსალური ინფორმაციული ნაკადი სხვადასხვა სფეროებში, რომლებიც პროდუქციული დამუშავების გარეშე შესაძლოა გადაიქცეს ნაგვის ზეავად. და მეორე, საშუალებები და მონაცემთა დამუშავების მეთოდები გახდა ყველასათვის მისაწვდომი და მოსახერხებელი, ხოლო მათი რეზულტატები კი გასაგები ყოველი ადამიანისათვის.

Data Mining- ის გამოყენების სფერო არ არის განსაღვრული. ის ყველგან არის სადაც კი არსებობს მონაცემები. მაგრამ პირველ რიგში Data Mining –ის მეთოდებმა დააინტრიგა კომერციული საწარმოები, რომლებიც თავიანთ პროექტებს ახორციელებენ ინფ. მონაცემთა საცავის (Data Warehousing) საფუძველზე. ასეთი საწარმოების მაგალითი გვიჩვენებს Data Mining-ის გამოყენების ეფექტიანობას, რომელიც აღწევს

²⁵ ცხრილი შექმნილია შემდეგი მასალის მიხედვით: Дюк В., "Data Mining – состояние, проблемы, новые решения", http://on.wplus.net/sparm/science/Data_mining.html, 1999.

100%-ს. წარმოდგენილია შეტყობინებები 20 მილიონ დოლარიან პროექტზე რომელმაც თანხა ამოიღო 4 თვეში. სხვა მაგალითი: წლიური ეკონომია 700 ათასი დოლარი Data Mining-ის დანერგვის შედეგად დიდი ბრიტანეთის სუპერმარკეტებში.

Data Mining-ი წარმოადგენს დიდ ფასეულობას ანალიტიკოსებისათვის მათ ყოველდღიურ საქმიანობაში. საქმიანმა ხალხმა კარგად გაითავისა, რომ Data Mining-ის მეთოდების მეშვეობით მათ შეუძლიათ მიიღონ საგრძნობი უპირატესობა კონკურენტიან ბრძოლაში. ახლა კი განვიხილოთ Data Mining-ის ზოგიერთი აპლიკაციები.

საცალო ვაჭრობის სხვადასხვა წარმოებები დღესდღეობით აგროვებენ ინფორმაციას ცალკეული შენაძენის შესახებ, მაღაზიის საკრედიტო ბარათის გამოყენებით, ან და კონტროლის კომპიუტერიზირებული სისტემების მეშვეობით. წარმოადგენთ ამოცანებს, რომლებიც შესაძლებელია გადავწყვიტოთ საცალო ვაჭრობაში Data Mining –ის მეშვეობით:

- მყიდველის კალათის ანალიზი (დამთხვევის ანალიზი), წარმოდგენილია იმისათვის, რომ გამოიყოს ის საქონელი, რომელსაც მყიდველი ყიდულობს ერთად. მყიდველის კალათის ცოდნა, საჭიროა იმისათვის, რომ გაუმჯობესდეს რეკლამა და იმ საქონლის მომარაგება, რომელიც უფრო მოხმარებაშია და მოხდეს მათი დალაგება გამოსაჩენ ადგილზე;
- დროებითი შაბლონების კვლევა ეხმარება სავაჭრო ობიექტებს გათვალოს იმ საქონლის მომარაგების დონე, რომელიც საჭიროა. ის იძლევა ისეთი სახის პასუხებს, როგორიცაა: თუ მომხმარებელმა იყიდა ვიდეოკამერა, რამდენხანში იყიდის ის ახალ ბატარეებსა და ფირს?;
- საპროგნოზო მოდელების შექმნა ეხმარება სავაჭრო ობიექტებს გათვალოს სხვადასხვა კატეგორიის

მომხმარებლების მოთხოვნები. მაგალითად არსებობენ ისეთი კლიენტები, რომლებიც ყიდულობენ მხოლოდ და მხოლოდ ცნობილი მარკების საქონელს ან/და კლიენტები, რომლებიც ყიდულობენ საქონელს მხოლოდ ფასდაკლებებზე. ეს ცოდნა საჭიროა ზუსტად მიმართული ეკონომიური ღონისძიებების ჩასატარებლად, სხვადასხვა ტიპის საქონლის წინა რიგებზე წამოსაწევად.

Data Mining- ის ტექნოლოგიის მიღწევები გამოიყენება საბანკო საქმეში შემდეგი გავრცელებული ამოცანების გადასაწყვეტად:

- *საკრედიტო ბარათებით თაღლითობის აღმოფრხერა.* ძველი ტრანზაქციების ანალიზის მეშვეობით, რომლებიც აღმოჩნდა თაღლითური, ბანკს აქვს შესაძლებლობა გამოავლინოს ასეთი თაღლითობის ზოგი სტერეოტიპი. მაგალითად, შეიძლება დადგინდეს, რომ ერთ-ერთი შემატყობინებელი სიგნალია მრავალი ტრანზაქციები ელექტრო საყოფაცხოვრებო ტექნიკის მაღაზიებში მოკლე დროის პერიოდში. მიღებული ცოდნა, ბანკს შეუძლია გამოიყენოს თავის მოქმედ სისტემებში, ტრანზაქციის აკრძალვის მეშვეობით, თაღლითობის სტერეოტიპის დამთხვევის შემთხვევაში, მყიდველთან წინასწარი საუბრის გარეშე;
- *კლიენტების სეგმენტაცია.* კლიენტების სხვა და სხვა კატეგორიებად დაყოფის მეშვეობით, ბანკს აქვს შესაძლებლობა აწარმოოს თავისი მარკეტინგული პოლიტიკა უფრო მიზანმიმართულად და ნაყოფიერად, სხვადასხვა ჯგუფის კლიენტებისათვის სხვადასხვა მომსახურების შეთავაზებით. მაგალითად, ბანკს შეუძლია შესთავაზოს კლიენტებს, რომლებიც ბევრს მოგზაურობენ ერთი საერთო ტიპის ბარათი (affinity) და მეორე, კლიენტებს, რომლებიც ყოველთვის დროზე იხდიან გადასახადს თავის ანგარიშიდან შენაძენებში საკრედიტო ბარათებით. სეგმენტაცია აგრეთვე სასარგებლოა

გასარკვევად, საგარაუდოდ თუ რომელი განყოფილება მიიღებს მოგებას გარკვეული სარეკლამო აქციებიდან;

- *კლიენტური ცვალებადობის პროგნოზირება.* Data Mining-ი იძლევა ბანკებისათვის საშუალებას შეექმნათ ბანკის სასურველი კლიენტების საპროგნოზო მოდელები, და შესაბამისად მოვემსახუროთ ამა თუ იმ კატეგორიას. ამისათვის ბანკი არკვევს დღეისათვის მომგებიანი კლიენტების მოთხოვნების ხასიათს და Data Mining-ის მეთოდების მეშვეობით არკვევს საერთო ხასიათს, რომლებითაც ისინი გამოირჩეოდნენ რამოდენიმე წლის წინ. შემდგომ ის ავლენს კლიენტებს, რომლებსაც გააჩნიათ ეს მახასიათებლები დღეს და ითვალისწინებს ამ კლიენტებს როგორც ახლო მომავალში მომგებიანებს. ბანკს შეუძლია ზემოთ აღნიშნული კლიენტები ჩართოს კლიენტების შენარჩუნების გამიზნულ პროგრამებში, მაგალითად სპეციალური შეთანხმებების შეთავაზებით ან/და საჯარიმო სანქციების შემოტანით მომდევნო მომსახურების შეთავაზებისგან თავის არიდების შემთხვევაში.

ტელეკომუნიკაციის სფეროში, დამახასიათებელია კონკუნკურენციის ზრდის ტენდენცია. აქ Data Mining ტენდენცია, კომპანიებს ეხმარება ეფექტურად წინ წაწიონ თავისი მარკეტინგული პროგრამების ფასწარმოქმნები, იმისათვის, რომ შეძლონ კლიენტების შენარჩუნება და ახალი კლიენტების მოზიდვა. ტიპიურ ღონისძიებებს წარმოადგენს:

- *ჩანაწერების ანალიზი გამოძახებების დახასიათების შესახებ.* ასეთი ანალიზის საფუძველს წარმოადგენს - კლიენტების მოძიების საშუალება მოხმარების ერთნაირი სტერეოტიპებით მომსახურებით, მიმზიდველი ფასების კრებულით და მომსახურებით;
- *კლიენტების ღოიადურობის გამოვლენა.* ზოგიერთი კლიენტი ძალიან ხშირად იცვლის პროვაიდერებს, ახალი კომპანიების პროგრამების სარგებლობით, რაც ასტიმულირებს ახალი კლიენტების მოზიდვას. Data Mining შესაძლებელია გამოყენებულ იქნას სხვადასხვა ტიპის

კლიენტების გამოსავლენად. მაგალითად ისეთი კლიენტების, რომლებმაც ერთხელ გამოიყენეს ამა თუ იმ კომპანიის მომსახურება, და ყოველთვის მათი ერთგულები დარჩებიან. საერთო ჯამში, მარკეტინგული დანახარჯები უნდა გაკეთდეს იქ სადაც შემოსავალი მეტი იქნება.

სადაზღვევო კომპანიები, რამოდენიმე წლის განმავლობაში აგროვებენ ინფორმაციის დიდ რაოდენობას.

- *თაღლითობის გამოვლენა.* სადაზღვევო კომპანიებს აქვს შესაძლებლობა შეამცირონ თაღლითობის რაოდენობა, სხვადასხვა განცხადებებში დანაკარგის ანაზრაურებით ერთნაირი სტერეოტიპების გამოვლენით, რომლებიც ხასიათდება იურისტებს შორის ურთიერთობით, ექიმებსა და განმცხადებლებს შორის, მაღალი თანხის მოთხოვნის შემთხვევებში, მაგალითად ზარალის ანაზრაურების დროს;
- *პროდუქტების დამუშავება.* Data Mining-ი მოგვცემს კლიენტების კატეგორიების სახარბიელო კომბინაციების გამოვლენის საშუალებას, პოლისების ვარიანტებს და მათი დაფარვის საშუალებებს. ასეთი კომბინაციების ცოდნა საჭიროა ახალი პროდუქტების დამუშავებისათვის და აგრეთვე არსებული პროდუქტების ზუსტი მორგებისათვის მომავალში გასაყიდათ;
- *რისკის ანალიზი.* ფაქტორების შეთანხმების გამოვლენის გზით, რომლებიც დაკავშირებულია გადახდილ განცხადებებთან, დამზღვევებს შეუძლიათ შეამცირონ თავიანთი დანახარჯები მათი ვალდებულებისადმი. ცნობილია შემთხვევა, როდესაც შეერთებულ შტატებში მსხვილმა სადაზღვევო კომპანიამ შეამოწმა განცხადებები, რომელთა საფუძველზეც იყო გადახდილი დიდი თანხები ბოლო ორი წლის განმავლობაში. ამ დროს, გამოვლინდა, რომ დაქორწინებული პირებისათვის იყო გადახდილი ორჯერ მეტი თანხები, ვიდრე მარტოსულ განმცხადებლებისათვის.

კომპანიამ სასწრაფოდ მოახდინა რეაგირება ამ ახალ ცოდნაზე შემდეგი სახით – განიხილა ფასდაკლების პოლიტიკა ოჯახური წყვილებისათვის.

Data Mining შესაძლოა გამოყენებულ იქნას მრავალ სხვა სფეროში:

- *ბაზრის სეგმენტაცია.* Data Mining-ის მეთოდების გამოყენების საშუალება ეძლევა ყველა სფეროს თავიანთი კლიენტურის ცალკეული სეგმენტების გამოსავლენად. Data Mining-ი წარმოებებს აძლევს საშუალებას გაითვალისწინონ გაცილებით მეტი პარამეტრი, ვიდრე ეს შესაძლებელი იყო არასტრუქტურირებული ინფორმაციის შენახვის არატრადიციული მეთოდების საფუძველზე;
- *მანქანათმშენებლობის განვითარება.* მანქანის აწყობისას მწარმოებლები ითვალისწინებენ ყოველი ცალკეული კლიენტის მოთხოვნებს, ამიტომ მათ სჭირდებათ სხვადასხვა განსაზღვრული მახასიათებლების პროგნოზირება და იმის ცოდნა, თუ რა მახასიათებლების შეკვეთა ხდება ერთობლივად;
- *გარანტიების პოლიტიკა.* მწარმოებლებს ესაჭიროებათ ინფორმაცია სამომავლოდ კლიენტების რაოდენობის შესახებ, რომლებიც შემოიტანენ საგარანტიო განაცხადებს. საჭიროა აგრეთვე განაცხადების საშუალო საფასური;
- *ხშირად მფრინავი კლიენტებისათვის სპეციალური პროგრამები.* ავიაკომპანიებს შეუძლიათ გამოავლინონ კლიენტების ჯგუფი, რომელთათვის შესაძლებელია მომზადდეს პროგრამები, რომელთა საფუძველზეც ისინი უფრო მეტს იფრენენ. მაგალითად ერთმა ავიაკომპანიამ აღმოაჩინა კლიენტების შემდეგი კატეგორია: ისინი ხშირად დაფრინავდნენ მოკლე მანძილებზე, რის შედეგადაც ვერ აგროვებდნენ შესაბამის მანძილების რაოდენობას, რის საფუძველზეც გახდებოდნენ მათი კლუბის წევრები. ამის გამო ავიაკომპანიამ შეცვალა კლუბში მიღების წესები ასეთი კლიენტებისათვის.

ცნობილია უამრავი საექსპერტო სისტემა სამედიცინო დიაგნოზების დასასმელად. ისინი ძირითადად შექმნილნი არიან წესების საფუძველზე, რომლებიც წარმოადგენენ სხვადასხვა ავადმყოფობების სხვადასხვა

სიმპტომების ერთობლიობას. ასეთი წესების სახით, იგებენ არა მარტო იმას თუ რით არის დაავადებული პაციენტი, არამედ იმასაც თუ რა საშუალებებით უნდა უმკურნალონ მას. წესები გვეხმარება მედიკამენტოზური საშუალებების არჩევაში, მათი ჩვენებებისა და უკუჩვენებების განსაზღვრაში, ორიენტაცია მოახდინოს სამკურნალო პროცედურებში, შეიქმნას ეფექტური მკურნალობის პირობები, და გაირკვეს დრო და პირობები დანიშნული მკურნალობის და ა.შ. Data Mining-ი გვაძლევს საშუალებას აღმოვაჩინოთ წარმოდგენილი წესები სამედიცინო მონაცემებში. Data Mining-ის გამოყენების საუკეთესო მაგალითს ჯანდაცვის სისტემაში წარმოადგენს ‘მანჩესტერ იუნაიტედ’-ის მაგალითი. Clementine სისტემის მეშვეობით, რომლის შესახებაც ჩვენ გვექნება მომავალში საუბარი და რომელსაც ჯანდაცვის სისტემაში გააჩნია კლიენტების რაოდენობის 30%, შესაძლებელია გამოვლინდეს გენეტიკური ფაქტორები, რომლებიც ზემოქმედებას ახდენს სხვადასხვა ავადმყოფობებზე. მიახლოებით ასეთი მიზნებისათვის ამ სისტემას იყენებს ცნობილი ფეხბურთის კლუბი ‘მანჩესტერ იუნაიტედ’-ი, რომელსაც გააჩნია 5000 მოთამაშე სხვადასხვა წლოვანების კატეგორიებში. სისტემა იძლევა ყოველი ცალკეული მოთამაშეების ტრავმების პროგნოზირების საშუალებას (და ზოგჯერ თავიდან აცილების საშუალებას) და პერსონალური ვარჯიშების გეგმების შედგენის საშუალებას.²⁶

ყველაზე მძაფრად და ამასთან მკვეთრად დგას ამოცანა კანონზომიერების აღმოჩენის ექსპერიმენტალურ მონაცემებში მოლეკულარულ გენეტიკაში და გენურ ინჟინერიაში. აქ ის ფორმულირდება, როგორც ეგრეთ წოდებული მარკერების განსაზღვრა, რომლის ქვეშაც წარმოდგინება გენეტიკური კოდები, რომლებიც კონტროლირებას უწევენ ცოცხალ ორგანიზმებს ამა თუ იმ ფენოტიპური ნიშნების მიხედვით. ასეთ კოდებს შესაძლებელია ჰქონდეთ ასი, ათასი და მეტი დამაკავშირებელი ელემენტები.

გენეტიკური აღმოჩენების გასაფართოვებლად გამოიყოფა უამრავი თანხები. ამ ბოლო დროს ამ სფეროში წარმოიქმნა დიდი ინტერესი Data

²⁶ Н. Смирнов Система для «Манчестера» 18.08.2003 *Еженедельник "Computerworld"*, #30, 2003 год // Издательство "Открытые системы"

Mining-ის გამოყენების თაობაზე. ცნობილია რამოდენიმე მსხვილი კომპანია, რომლებიც სპეციალიზირდებიან ამ მეთოდების გამოყენებაზე - ადამიანისა და მცენარეების გენომის გაშიფრვა.

Data Mining მეთოდები ფართო გამოყენებას პოულობს გამოყენებით ქიმიაში (ორგანულ და არაორგანულში). აქ ძალიან ხშირად წარმოიშვება ხოლმე კითხვა ქიმიური წყობილების უცნაურობების მოძიებისას ამა თუ იმ ნაერთებში, რომლებიც განსაზღვრავენ მათ თვისებებს. კერძოდ, აქტუალურია ასეთი ამოცანა რთული ქიმიური ნაერთების ანალიზისას, რომლის აღწერაც მოიცავს 100 და 1000 სტრუქტურულ ელემენტს და მათ კავშირებს.

შესაძლებელია მოვიყვანოთ კიდევ უამრავი მაგალითი ცოდნის სხვადასხვა სფეროდან, სადაც Data Mining მეთოდები თამაშობს წამყვან როლს. შემდგომში მე აღვწერ ერთ-ერთ სფეროს, სადაც შესაძლებელია Data Mining ტექნოლოგიის წარმატებულად გამოყენება. ასეთი სფეროების განსაკუთრებულობას წარმოადგენს მათი რთული სისტემური ორგანიზაცია. ისინი ძირითადად ეხება სისტემის ორგანიზაციის კიბერნეტიკის ზედა დონეს, რომლის კანონზომიერებაც შეუძლებელია აღწერილი იყოს ზუსტად სტატისტიკური ან/და სხვა ანალიტიკური მათემატიკური მოდელების ენაზე. მონაცემები, ასეთ სფეროებში არაერთგვაროვანია, გეტეროგენულია, არასტაციონარულია და ხშირად გამოიყოფა მათი ფართო ზომიერებით.

2.2. ტიპების კანონზომიერება და ინტელექტუალური ანალიზის კლასების სისტემები

გამოიყოფა ხუთი კანონზომიერების სტანდარტული ტიპი, რომლებიც იძლევა საშუალებას გამოვააშკაროთ Data Mining-ის მეთოდები:

- ასოციაცია;
- თანმიმდევრობა;
- კლასიფიკაცია;
- კლასტერიზაცია;

- პროგნოზირება.

ასოციაცია ელინდება იმ შემთხვევაში, როდესაც რამოდენიმე მოვლენა ერთმანეთთან არის დაკავშირებული. მაგ. სუპერმარკეტში ჩატარებულმა დაკვირვებებმა აჩვენეს, რომ 65% სიმინდის “ჩიფსების” მყიდველები აგრეთვე ყიდულობენ “კოკა-კოლა“-საც, ხოლო ასეთი კომპლექსის შეძენისას ფასდაკლებით “კოლა“-ს ყიდულობს 85%. ამ შემთხვევაში ასეთი ტიპის ასოციაციის ინფორმაციით მენეჯერებს შეუძლიათ თვალნათლივ დაინახონ, თუ რამდენად ნაყოფიერია ასეთი ფასდაკლების შეთავაზება მყიდველისათვის.

თუ არსებობს დროში მოვლენების რაღაც ჯაჭვი, საუბარი არის “თანმიმდევრობითობაზე”. ასე მაგალითად ახალი სახლის შეძენისას 45%-ს, რაღაც დროის პერიოდში, უწინდებად სამზარეულოსთვის შენაძენების გაკეთების აუცილებლობა და ა.შ.

კლასიფიკაციის მიხედვით შესაძლო ხდება ამა თუ იმ ობიექტის ჯგუფის მახასიათებლებისა და მოვლენების გამოაშკარავება. ეს ხდება უკვე რაღაც კლასიფიცირებული ობიექტების ანალიზისა და საერთო წესების ფორმულირების მეშვეობით. Data Mining-ი თვითონ გამოყოფს რაღაც ერთგვაროვან მონაცემთა ჯგუფს.

ყოველი არსებული პროგნოზირების სისტემის საფუძველს წარმოადგენს ისტორიული ინფორმაცია, რომელიც ინახება მონაცემთა ბაზაში დროებითი რიგების სახით. თუ მოგვეცემა საშუალება, რომ შევქმნათ მათემატიკური მოდელი და ვიპოვოთ შაბლონები რომლებიც ადეკვატურად განსაზღვრავენ ამ დინამიკას, ამ შემთხვევაში გვეძლევა შესაძლებლობა მათი მეშვეობით ვიწინასწარმეტყველოთ სისტემის მოქმედება მომავალში. განვიხილოთ ინტელექტუალური ანალიზის კლასების სისტემები.

წარმოდგენილი კლასიფიკაციის საფუძველად ჩადებული გვაქვს სადისერტაციო ნამუშევარი.

საგნობრივ-ორიენტირებული ანალიტიკური სისტემები ძალიან მრავალფეროვანია. ყველაზე გავრცელებული ასეთი სისტემების ქვეკლასი, რომელიც ყველაზე გავრცელებულია საფინანსო ბაზრის კვლევაში, არის “ტექნიკური ანალიზი”. ის წარმოადგენს ფასების დინამიკის პროგნოზისა და ინვესტიციური პორტფელის ოპტიმალური

სტრუქტურის ამორჩევის ერთობლიობას, რომლებიც დაფუძნებულია ბაზრის დინამიკის სხვადასხვა ემპერიულ მოდელებზე. ეს მეთოდები შესაძლოა იყოს ძალიან მარტივი, მაგრამ ჰქონდეთ საკმაოდ ორიგინალური მათემატიკური საფუძველი.

მიუხედავად იმისა, რომ ბოლო ვერსიები ყველა ცნობილი სტატისტიკური პაკეტებისა შეიცავენ როგორც ტრადიციულ სტატისტიკურ მეთოდებს და Data Mining-ის ელემენტებს, ძირითადი ყურადღება მაინც ეთმობა კლასიკურ მეთოდებს – კორელაციურს, რეგრესიულს, ფაქტორულ ანალიზსა და სხვა. ამ კლასის ნაკლოვანებად მიიჩნევენ მოთხოვნას მომხმარებლის სპეციალური მომზადებისათვის. აგრეთვე თვლიან, რომ სტატისტიკური პაკეტები არის ძალიან “მძიმე” მასობრივი გამოყენებისთვის ფინანსებსა და ბიზნესში. ამავე დროს ეს სისტემები ძალიან ძვირია. არსებობს სტატისტიკური პაკეტების ერთი პრინციპიალური ნაკლოვანება, რომელიც განსაზღვრავს შეზღუდულ Data Mining-ის გამოყენებას. მეთოდების უმრავლესობა ეყრდნობა გამორჩევის დახასიათებას. რეალურ შემთხვევებში კი ეს დახასიათებლები არის ფიქტიური დახასიათებლები რთული ცხოვრებისეული ფენომენებისათვის.

მაგალითისათვის შეგვიძლია მოვიყვანოთ ისეთი ძლიერი და გავრცელებული პაკეტები, როგორიც არის SAS, SPSS, STATGRAPICS, STATISTICA, STADIA და სხვა.

ეს არის სისტემის დიდი კლასი, რომლის არქიტექტურაც ცდილობს გაუკეთოს ნერვულ ქსოვილს იმიტაცია ნეირონებისაგან. ერთ-ერთ ყველაზე გავრცელებულ არქიტექტურაში, მრავალფენიან პერცეპტრონში შეცდომის უკუგავრცელებით, ხდება ნეირონების ემულირება იერარქიული ქსელის შემადგენლობაში, სადაც ყოველი უფრო მაღალი საფეხურის ნეირონი თავისი შესასვლელებით დაკავშირებულია ქვედა საფეხურების ნეირონების გამოსასვლელებთან. ყველაზე ქვედა ნეირონების ფენას მიეწოდება შესასვლელი პარამეტრების მნიშვნელობები, რომელთა საფუძველზეც უნდა მივიღოთ გადაწყვეტილებები. ეს მნიშვნელობები განიხილება როგორც სიგნალები, რომლებიც გადაეცემა ზედა ფენაში მათი გამძლავრებისა თუ შენელების რიცხვითი მნიშვნელობების დამოკიდებულებით (წონით),

რომლებიც ენიჭება ნეირონებს შორის არსებულ კვანძებს. ყველაზე ზედა ფენის ნეირონების გამოსასვლელზე მუშავდება პასუხი, მთელი ქსელის რეაქცია, შესასვლელი პარამეტრების მნიშვნელობების შეტანაზე. იმისათვის, რომ ქსელის გამოყენება შესაძლებელი გახდეს შემდგომშიც, მას უნდა თავდაპირველად “ნეტრინირება” გავუკეთოდ ადრე მიღებულ მნიშვნელობებზე, რომლებისთვისაც ცნობილია შესასვლელი პარამეტრების მნიშვნელობები და სწორი პასუხები მათზე. ეს მცდელობა გამოყენებულია ნეირონებთაშორის კავშირების წონების შერჩევისათვის, რომელიც უზრუნველყოფს ცნობილ პასუხებთან ქსელის პასუხების მიახლოებას.

ნეირონული პარადიგმის მთავარ ნაკლოვანებას წარმოადგენს უზარმაზარი მოცულობის შემსწავლელი შერჩევითობის აუცილებლობა. სხვა ნაკლოვანება აისახება შემდეგში: ნეიტრირებული ნეირონული ქსელიც კი წარმოადგენს შავ ყუთს. ნეირონული სისტემების მაგალითები: BrainMaker(CSS), NeuroShell(Ward Systems Group), OWL(HyperLogic). ამ სისტემების საფასური მაღალია.

Case based reasoning CBR სისტემის იდეა ერთი შეხედვით ძალიან მარტივია. იმისათვის, რომ გაკეთდეს პროგნოზი მომავლისათვის ან/და შეირჩეს სწორი გადაწყვეტილება, ეს სისტემები პოულობენ წარსულში მოცემული სიტუაციის ახლო ანალოგებს და ირჩევენ იმავე პასუხს, რომელიც იყო მათვის სწორი. ამიტომ ამ მეთოდს აგრეთვე უწოდებენ “ახლო მეზობლის” მეთოდს (nearest neighbor).

ამ სისტემის მთავარ მინუსად ითვლება ის, რომ ის საერთოდ არ ქმნის არანაირ მოდელებს ან წესებს, რომლებიც აერთიანებს წინამორბედ გამოცდილებებს – გადაწყვეტილების შერჩევისას ის ეფუძნება ხელმისაწვდომ ისტორიულ მონაცემების მასივს, ამიტომ შეუძლებელია ითქვას, თუ რომელ კონკრეტულ CBR სისტემის ფაქტორებზე იქმნება თავისი პასუხები.

Data Mining-ის ამოცანების გადაწყვეტაში ერთ-ერთი ყველაზე პოპულარული მეთოდია გადაწყვეტილებების ხე. ისინი ქმნიან იერარქიულ სტრუქტურას კლასიფიცირებული წესებისათვის. მაგ.: “თუ ... მაშინ”, რომელსაც აქვს ხის სახე. იმისათვის, რომ მივიღოთ გადაწყვეტილება, თუ რომელ კლასს მივაკუთვნოთ ესა თუ ის ობიექტი

თუ სიტუაცია. იმისათვის, რომ მივიღოთ გადაწყვეტილება, თუ რომელ კლასს მივაკუთვნოთ ესა თუ ის ობიექტი თუ სიტუაცია, საჭიროა კითხვებზე პასუხი, რომლებიც მოიძებნება ხის კვანძებში მისი ფესვებიდან მოყოლებული. კითხვებს აქვთ შემდეგი სახე “პარამეტრ A მნიშვნელობა მეტია X –ზე?”. თუ პასუხი დადებითია, ამ შემთხვევაში ხორციელდება გადასვლა შემდეგი საფეხურის მარჯვენა კვანძისკენ, თუ უარყოფითია, მაშინ გადასვლა ხდება მარცხენა კვანძისკენ; შემდგომ მომდევნო კითხვა, რომელიც დაკავშირებულია შესაბამის კვანძთან. ამგვარი მიდგომის პოპულარობა მდგომარეობს თვალნათლივ პროცესებსა და გაიოლებულ აღქმაში. მაგრამ ხეებისათვის ძალიან მკაფიოდ დგას მნიშვნელობის პრობლემა. საქმე იმაშია, რომ ცალკეულ კვანძებს ყოველივე ახალ შექმნილ საფეხურზე სულ უფრო ცოტა და ცოტა მონაცემების რაოდენობა შეესაბამება - ხე აცალკევებს მონაცემებს ძალიან ბევრ კერძო შემთხვევებზე. რაც უფრო მეტია ასეთი კერძო შემთხვევები, მით უფრო ნაკლები შემსწავლელი შემთხვევა ხდება ასეთ კერძო შემთხვევებში და მით უფრო მკაფიო ხდება მათი კლასიფიკაცია. თუ შექმნილი ხე წააგავს ბუჩქს - ანუ შედგება გაუმართლებელი დიდი რაოდენობის ტოტებისაგან ის არ მოგვცემს სტატისტიკურად საფუძვლიან პასუხებს. გადაწყვეტილებების ხე იძლევა სასურველ პასუხებს მხოლოდ დამოუკიდებელი თვისებებისათვის.

გადავხედოთ ამ მიდგომას PolyAnalyst-ის²⁷ მეშვეობით. მოცემულ სისტემაში მიზნობრივი ცვლადების დამოკიდებულების ჰიპოთეზები სხვა ცვლადებთან ფორმულირდება პროგრამირების ზოგიერთი ენების საშუალებით. პროგრამირების შექმნის პროცესი იქმნება ევოლუციურ პროგრამირების სამყაროში. როდესაც სისტემა პოულობს პროგრამას, რომელიც ზუსტად გამოხატავს სასურველ დამოკიდებულებას, იგი იწყებს მასში მოდიფიკაციების შეტანას და შვილობილი პროგრამებს შორის დაზუსტებული პროგრამის არჩევას, ასეთი სახით სისტემა “ზრდის” რამოდენიმე პროგრამის გენეტიკურ ხაზს, რომლებიც კონკურენციას უწევენ ერთმანეთს სასურველი დამოკიდებულების გამოვლენაში. სპეციალური სატრანსლაციო სისტემის მოდულს

²⁷ Poly Analyst Решения для лучших в бизнесе/ Компания Мегапьютер Интеллидженс/ Аналитические программы и системы/ Москва 2001

PolyAnalyst-ს გადაჭყავს მოძებნილი დამოკიდებულებები სისტემის შიდა ენიდან მომხმარებლისათვის გასაგებ ენაზე (მათემატიკური ფორმულები, ცხრილები და ა. შ.) იმისათვის, რომ მიღებული რეზულტატები გავხადოდ უფრო გასაგები არამათემატიკოსისათვის, არსებობს ვიზუალიზაციის საშუალებების მდიდარი არსენალი სხვადასხვაგვარი დამოკიდებულებების აღმოსაჩენად.

გამოსასვლელი დამოკიდებულებების სტატიკური მნიშვნელობების კონტროლისათვის გამოიყენება თანამედროვე მეთოდების ნაკრები, მაგალითად რანდომიზირებული ტესტირება. ევოლუციური პროგრამირების სხვა მიმართულება დაკავშირებულია მიზნობრივი ცვლადების დამოკიდებულების ძებნასთან სხვა ცვლადებთან განსაზღვრული ტიპის ფუნქციების სახით. მაგალითად ასეთი ტიპის ალგორითმებში ერთ-ერთი ყველაზე საუკეთესო მაგალითია – არგუმენტების ერთობლივი დაჯგუფების მეთოდით დამოკიდებულებების ძებნა პოლინომების ფორმით.

Data Mining-ი შეიძლება ითქვას არ არის გენეტიკური ალგორითმების გამოყენების სფერო. ისინი უნდა განიხილოს, როგორც ძლიერი საშუალება სხვადასხვაგვარი კომბინატორული და ოპტიმიზაციის ამოცანების გადასაჭრელად. მიუხედავად ამისა გენეტიკური ალგორითმები შემოვიდა Data Mining-ის სტანდარტულ მეთოდების ინსტრუმენტებში, რის გამოც ისინი შეტანილია მოცემულ ნაწივეტში. გენეტიკურ ალგორითმებს გააჩნიათ უამრავი უარყოფითი თვისება. ისევე როგორც რეალურ ცხოვრებაში, “ევოლუცია” შეიძლება შეჩერდეს რომელიმე არაპროდუქტიულ ტოტზე.

ცხოვრებისეული მაგალითი: როდესაც ორ პერსპექტიულ მშობელს, რომლებიც გამორიცხულები იქნებიან ევოლუციიდან გენეტიკური ალგორითმით შესაძლოა ჰქონდეთ საშუალება შექმნან მაღალეფექტური შთამომავალი. ეს შესამჩნევი ხდება ფართომაშტაბიანი, რთული შიდა კვანძებიანი ამოცანების გადაწყვეტისას.

მაგალითად შეიძლება მოვიყვანოთ სისტემა GeneHunter Ward System Group ფირმის.

შეზღუდული გადარჩევის ალგორითმები მოგვაწოდა 60-იან წლებში მ.მ. ბონგარდმა მონაცემებში ლოგიკური კანონზომიერების გამოსავლენად. მას შემდეგ მათ გვიჩვენეს თავისი ეფექტურობა მრავალი ამოცანის გადაწყვეტისას სხვადასხვა სფეროში.

ეს ალგორითმები თვლიან კომბინაციებს მონაცემების ქვეჯგუფებში მარტივი ლოგიკური მოვლენების მეშვეობით.

მარტივი ლოგიკური მოვლენის მაგალითი: $x=a$; $x>a$; $a<x<b$ და სხვა, სადაც x რაიმე პარამეტრია, a და b – კონსტანტები. შეზღუდვას წარმოადგენს მარტივი ლოგიკური მოვლენების კომბინაციების სიგრძე. გამოთვლილი სიხშირეების ანალიზის საფუძველზე კეთდება დასკვნა ამა თუ იმ კომბინაციების სარგებლიანობის შესახებ მონაცემებში ასოციაციის დასადგენად, მათი კლასიფიკაციისთვის, პროგნოზირებისათვის და ა.შ.

ასეთი მიდგომის ყველაზე თანამედროვე წარმომადგენელია სისტემა WizWhy, WizSoft .

WizWhy-ის ავტორი გვარწმუნებს, რომ მისი სისტემა ავლენს ყველა ლოგიკურ IF.....THEN წესს მონაცემებში. რა თქმა უნდა ეს ასე არ არის. 1) WizWhy სისტემაში IF.....THEN წესში კომბინაციის მაქსიმალური სიგრძე უდრის 6-ს. 2) ალგორითმის მუშაობის დასაწყისში იწყება მარტივი ლოგიკური მოვლენების ძებნა ევრეტიკული წესით, რომელზეც შემდგომში მყარდება მომდევნო ანალიზი. WizWhy ამ ნაკლოვნებების აღმოჩენისას რთული არ იყო მოგვეწოდებინა ისეთი ამოცანა რომელსაც ის საერთოდ ვერ გადაწყვეტდა, აგრეთვე სხვა მომენტიც: სისტემა იძლევა პასუხს მხოლოდ სასურველ დროში და არც თუ ისეთი დიდი მოცულობის მონაცემებისათვის (არა უმეტეს 20-სა).

მიუხედავად ამისა WizWhy სისტემა დღესდღეობით რჩება ლიდერად Data Mining – ის ბაზარზე. ამას გააჩნია თავისი საფუძველი. სისტემა ყოველთვის იძლევა უფრო მაღალ მაჩვენებლებს პრაქტიკული ამოცანების გადაწყვეტაში ყველა სხვა ალგორითმებთან შედარებით. ნახ.10-ზე წარმოდგენილია პროდუქტები Data Mining-ისათვის.



ნახ.10 DATA MINING-ის პოპულარული პროდუქტები

ახლა კი გაგვაჩნია წარმოდგენა DATA MINING პრობლემის თაობაზე და აგრეთვე ახალ გადაწყვეტილებებზე. მაგრამ, იმისათვის, რომ გადავიდეთ ძირითადი მაგალითის გამოკვლევებზე მინდა აღვწერო ყველა საშუალებები, რომელსაც გვაწვდის კომპანია ORACLE DATA MINING-სათვის, კერძოდ პროდუქტების ერთობლიობა და DATA MINING კლასის სერვისები.

Oracle კორპორაცია წარმოუდგენს თავის მომხმარებლებს Data Mining (მონაცემთა ინტელექტუალური ანალიზი) კლასის და სერვისის პროდუქტების ერთობლიობას. ეს ერთობლიობა ეხმარება კომპანიებს კლიენტების ქცევის გაგებასა და წინასწარ გამოვლენაში (მყიდველებს, დამკვეთებს) და საშუალებას იძლევა შეიქმნას სრული ინტეგრირებული გადაწყვეტილებები კლიენტებთან ურთიერთობებში (CRM). ამ დროისათვის თქვენ შეგიძლიათ გამოიყენოთ Oracle Data Mining Suite კომპლექტი (რომელსაც ვიცნობდით, როგორც "Oracle Darwin")

წინასწარმეტყველების მოდელების შესაქმნელად (predictive models), რომელიც თვენს მონაცემებში ეძებს ფარულ კანონზომიერებებს და ახალ ინფორმაციას, რომელიც შესაძლებელია გამოყენებულ იქნას:

- სისტემებში ექსტრემალური სიტუაციების ადრეული აღმოჩენა;
- საბაზრო ეკონომიკის პირობებში კლიენტების მოთხოვნებისა და ქცევების სრული აღთქმისათვის;
- სხვადასხვა გაყიდვების ეფექტურობის განსაზღვრისათვის;
- როგორც ბოროტმოქმედებთან ბრძოლის საშუალება.

ინფორმაცია, აღმოფხვრილი მონაცემებიდან ასეთი სახით, უზრუნველყოფს ბიზნეს-ანალიზს, რომლის რეზულტატიც წარმოადგენს კლიენტებთან ურთიერთობის გაუმჯობესებას და ამის საფუძველზე შემოსავლის გაზრდას და ძირითად ბიზნესში ეკონომიას.

განვიხილოთ:

- Data Mining –ის უპირატესობა;
- Data Mining-ის განსაზღვრა;
- Oracle Data Mining Suite-ის კომპლექტის შესაძლებლობების განხილვა;
- Data Mining-ის ფუნქციონალურობა Oracle CRM 11I აპლიკაციებში;
- Data Mining-ი და ახალი პროდუქტი პერსონალიზაციისათვის Web გარემოში;
- Data Mining-ის მომავალი Oracle-საგან.

2.3 Data Mining უპირატესობა

მონაცემები გროვდება სხვადასხვა წყაროებიდან: კლიენტების მომსახურების სფეროდან, შეკვეთების შემოტანიდან, გაყიდვებიდან, აგრეთვე web-გვერდებიდან და ა.შ. მონაცემები ბევრია, მაგრამ როგორ გამოვიყენოდ ისინი სათანადოთ? როგორ მივიღოთ სრული წარმოდგენა კლიენტებზე? როგორ შევიტყოთ ინფორმაცია მყიდველზე, რომელმაც ახლახანს შეიძინა თქვენს მაღაზიაში რაიმე პროდუქცია, რომ იგივე აღადმიანია, ვინც დატოვა საჩივარი თვენს web-გვერდზე, ან და

ათვალიერებდა თქვენს web-გვერდს სხვა პროდუქტებისა და მომსახურების ძიებაში? როგორ უნდა მივიღოთ სრული წარმოდგენა კლიენტზე, რომლის მონაცემებიც წარმოდგენილია უამრავ და სხვადასხვა ტიპის მონაცემთა წყაროებში?

პასუხი თვენს მონაცემებშია! თქვენ უნდა მოახდინოთ მონაცემების კონსოლიდირება ერთ წყაროში უნიფიცირებული წარმოდგენით მომხმარებლებზე, რომლის ფარგლებშიც ხდება მონაცემების აგრეგირება. დიდი შესაძლებლობა იმალება კვლევაში, ამ მონაცემების მოძიებაში თავიანთი კლიენტების უკეთ შესწავლისათვის და მათი მოთხოვნების წინასწარმეტყველებისთვის. ამ მონაცემების კვლევისას, თქვენ შეგიძლიათ მოიძიოთ თქვენი საუკეთესო კლიენტების სრული პროფილები, და შემდგომში მოიძიოთ სხვა კლიენტები ასეთი პროფილებით. თქვენ შესაძლებლობა მოგეცემათ იწინასწარმეტყველოთ მათი მოთხოვნები და შესთავაზოთ მათ ის პროდუქცია და მომსახურება, რომელიც მათ დააკმაყოფილებს, ანუ შესაძლებელი ხდება ამ კლიენტების შენარჩუნება და შემოსავლების გაზრდა, ამ კლიენტებისათვის პროდუქციის და მომსახურების გაყიდვით.

პრობლემა მდგომარეობს იმაში, თუ როგორ გამოვიყენოთ და სათანადოდ დავამუშაოთ ეს მონაცემები? როგორ განვსაზღვროთ რელევანტური მონაცემები, რომელსაც აქვს დიდი მნიშვნელობა? რა ფაქტორები განსაზღვრავს კლიენტების კმაყოფილებას? ხშირად მონაცემების ძირითადი ფასეულობა დევს დეტალებში, რომლებიც ეხება მხოლოდ ცალკეულ კლიენტებს. მაგრამ ასეთი სახის ინფორმაციით არსებობს მონაცემებში ჩაძირვის რისკი, თუ ვერ შეძლებთ მათთან “მორევას”.

თანამედროვე კონკურენტუნარიანი ეკონომიკის პირობებში წარმოებებისათვის კრიტიკულ მნიშვნელობას წარმოადგენს მათი აქტივების მართვა- კლიენტებით და მათზე ინფორმაციით. აქ Data Mining-ს აქვს დახმარების საშუალება. Data Mining-ი შეძლებს გაანალიზოს უზარმაზარი მასივები და გამოიძიოს დაფარული ფასეული ინფორმაცია, რომელსაც შეუძლია დაგვეხმაროს უკეთ შევისწავლოთ კლიენტები და ვიწინასწარმეთველოთ მათი ქცევები.

უფრო კონკრეტულად კი Oracle Data Mining Suite დაგეგმარება ყოველივე ამაში.

ამ საჭირო ინფორმაციით შეიარაღებულებს შეგვიძლია მოვახერხოთ კლიენტებთან უფრო მჭიდრო კავშირების დამყარება, რაც მოგვცემს საშუალებას, მათ უკეთ გავუგოთ და:

- კლიენტების შენარჩუნება და არასასურველი მოქმედებების თავიდან აცილება;
- კლიენტების პროფილების შედგენა, მათი ქმედებების უკეთ აღქმა;
- მოგების შენარჩუნება და მისი დონის ამაღლება;
- კლიენტების დანახარჯების შემცირება სხვადასხვა შესყიდვების დროს;
- კლიენტებისათვის საინტერესო წინადადებებით მიმართვა;

Oracle Data Mining Suite კომპლექტი საშუალებას მოგვცემს გავიდეს ტრადიციული აპლიკაციების გენერატორების (ანუ რომლებიც მოგვაწოდებენ ინფორმაციას, თუ რა მოხდა ეხლა) საზღვრებს გარეთ. ეს კომპლექტი გაანალიზებს წარსულს და გამოიყენებს მას, როგორც საფუძველს მომავლის საწინასწარმეტყველოდ.

ჩვენი საუკეთესო კლიენტების პროფილების გაანალიზებით Oracle Data Mining Suite მოგვცემს საშუალებას სხვა ადამიანების იდენტიფიცირებისათვის, ვინც ჯერ-ჯერობით არ წარმოადგენს საუკეთესო კლიენტებს თქვენს სიაში, მაგრამ მათი პროფილები ემთხვევა იმ კლიენტების პროფილებს, ვინც უკვე არიან სიაში.

ცოდნა კლიენტების “სტრატეგიული ფასეულობის” შესახებ, მომავალში შესაძლო ფასეული კლიენტების შესახებ და კლიენტების შესახებ, რომლებიც შესაძლოა საერთოდ არ გახდნენ ფასეულნი - წარმოადგენს ბიზნესის მართვის ერთ-ერთ საკვანძო პირობას.

კიდევ ერთხელ განვიხილოთ, თუ რა არის Data Mining და რით განსხვავდება იგი რეპორტების ფორმირებისა და მოთხოვნების გაცემის OLAP და სხვა ანალიტიკური საშუალებებისაგან. *“უბრალოდ რომ ვთქვათ, data mining ხშირად გამოიყენება დაფარული კანონზომიერებების*

გამოსავლენად ჩვენ მონაცემებში იმისათვის, რომ ჩვენ მოგვეცეს საუკეთესო საქმინი გადაწყვეტილებების მიღების საშუალება.”²⁸

Data mining- გვეხმარება ადრე დაფარული კანონზომიერებების და კავშირების გამოვლენაში იმისათვის, რომ შესაძლებლობა მოგვეცეს მივიღოთ უფრო საფუძვლიანი გადაწყვეტილება. ახლა კი განვიხილოთ ზოგიერთი ცნობილი განსაზღვრება.

მოთხოვნებისა და აპლიკაციების ფორმირების საშუალებები გვეხმარება მოვიძიოთ ინფორმაცია მონაცემთა ბაზებიდან და მონაცემთა საცავებიდან. ისინი საუკეთესონი არიან მაგალითად შემდეგი კითხვების პასუხის დროს: “ვინ ყიდულობდა ბოლო სამი წლის განმავლობაში განსაზღვრული ტიპის ფასიან ქაღალდებს?”

OLAP საშუალებები მიდის ამ ყველაფერზე გაცილებით შორს. შედეგებს, შედარებებს იგი ღებულობს ექსტრაპოლაციის გზით. OLAP-ი შესანიშნავია დეტალიზაციის ჩატარების დროს (drill-down) უფრო ვრცელი ინფორმაციის მისაღებად მონაცემების შესახებ, მაგალითად ისეთის როგორც: “როგორი განაწილებაა მოგებაში ფასიანი ქაღალდების მყიდველებში (mutual funds)?”

მოთხოვნებისა და აპლიკაციების ფორმირების საშუალებები, აგრეთვე OLAP, საუკეთესონი არიან მაშინ, როდესაც მომხმარებელს სჭირდება დეტალიზაცია და გაგება იმისი თუ რა მოხდა წარსულში. Data Mining-ი მიდის უფრო შორს, აანალიზებს წარსულს იმისათვის, რომ იწინასწარმეტყველოს მომავალი.

Data mining “იჭრება” მონაცემებში იმისათვის, რომ გამოიკვლიოს სასარგებლო კანონზომიერებები და ნიუანსები, რომლებიც “ჩაფლულები არიან” მონაცემებში ამ მონაცემების უზარმაზარი ზომის და სირთულეების გამო. Data mining-ი შესანიშნავია ნიუანსების მოძიების დროს (detailed insights) და ინდივიდუალური ტიპის წინასწარმეტყველებისას, მაგალითად: “ვინ იყიდის ფასიან ქაღალდებს ამა თუ იმ ტიპის მომდევნო 6 თვის განმავლობაში და რატომ?” SQL-მოთხოვნების და რეპორტების საშუალებების ფორმირება სასარგებლოა მონაცემთა დეტალიზაციის დროს. თუ თქვენ იცით, რას ეძებთ თვენს

²⁸ P. Смолл, Two Crows Corporation источник: <http://www.management-magazine.ru> Апрель 2003

უზარმაზარ კორპორატიულ მონაცემთა ბაზაში, გაგაჩნიათ კარგი ინსტრუმენტები, კარგი ანალიტიკოსი და გაქვთ უამრავი დრო, თქვენ აუცილებლად ნელ-ნელა მიხვალთ პასუხამდე. OLAP-საშუალებები და სტატისტიკური ანალიზის საშუალებები შესანიშნავი არიან ინფორმაციის მოსაძიებლად ანალიზის ჩატარებისა და აპლიკაციების ჩამოყალიბების მიზნით, მაგრამ ხშირად ისინი გამოუდგარნი ხდებიან როდესაც მონაცემთა რაოდენობა დიდ რაოდენობას აღწევს. ამას გარდა, როდესაც ჩვენ საქმე გვაქვს მაგალითად 25-ზე მეტ შემავალ ცვლადთან, სტატისტიკური რეგრესიული ანალიზის საშუალებების ტექნიკა შეუძლებლად თვლის რაიმე ანალიზის ჩატარებას.

როგორც ვთქვით დიდი რაოდენობის მონაცემის გადასამუშავებლად, სასარგებლო კანონზომიერებებისა და ნიუანსების გამოსაკვლევად და პროგნოზირების ჩასატარებლად შესანიშნავია Data Mining-ი. კერძოდ ჩვენს შემთხვევაში მას ვიყენებთ საგადასახადო სტრუქტურაში. თუ როგორ ხდება მისი გამოყენება ამ სფეროში, განვიხილავთ მომდევნო თავში.

III თავი. შედეგების პრაქტიკული რეალიზაცია და ანალიზი

ჩვენს ქვეყანაში მიმდინარე ეკონომიკური პროცესები, საბაზრო ეკონომიკის დამყარება და განვითარება, რათქმა უნდა გულისხმობს საგადასახადო სტრუქტურაში დასამუშავებელი ინფორმაციის მოცულობის ზრდას, რადგანაც იზრდება გადამხდელთა რაოდენობა. ინფორმაციის ზრდას კი მოჰყვება თანამედროვე და ეფექტური გადაწყვეტილებების დანერგვა. ინფორმაციული ტექნოლოგიების განვითარება ამ განხრით გადასახადების მოგროვების პროცესის რეგულირების, აგრეთვე ინფორმაციის დიდი ნაკადის უფრო ოპერატიულად და ზუსტად დამუშავების საშუალებას იძლევა. პროდუქტ Oracle Data Mining –ის გამოყენება ჩემი თვალსაზრისით მოგვცემს საშუალებას შევამციროთ დროითი და შრომითი დანახარჯები გადამხდელთა პირადი აღრიცხვის ბარათის ინფორმაციის დამუშავებისას, რაც გამოიწვევს სამუშაოს ოპტიმიზაციას და საგადასახადო სტრუქტურის მუშაობის ეფექტიანობას.

იმისათვის, რომ თვალნათლივ დავინახოთ, თუ რა უპირატესობას მოგვცემს ჩვენს შემთვევაში Oracle Data Miner-ის დანერგვა პირადი აღრიცხვის ბარათის ინფორმაციის დამუშავებისას, წარმოვადგენ თვით გადამხდელის პირადი ბარათის სტრუქტურას და მისი წარმოების პრინციპებს.

3.1 გადამხდელის პირადი ბარათი

პირადი აღრიცხვის ბარათების წარმოების წესი წარმოდგენილია საქართველოს საგადასახადო კოდექსის მე-3 მუხლის პირველი ნაწილის და 73-ე მუხლის მე-2 ნაწილის შესაბამისად. იგი გვიჩვენებს გადასახადის გადამხდელის/საგადასახადო აგენტის ან სხვა ვალდებული პირის პირადი აღრიცხვის ბარათებში ბიუჯეტთან ანგარიშსწორების აღრიცხვისა და ასახვის წესებს.

1) საერთო ცნობები ბარათის შესახებ

1. გადამხდელთა პირადი აღრიცხვის ბარათი (გპაბ) წარმოადგენს მკაცრი აღრიცხვის დოკუმენტს, რომელშიც აისახება გადამხდელის/საგადასახადო აგენტების ან სხვა ვალდებული პირების

მიერ დეკლარირებული და გადახდილი თანხები, აგრეთვე საგადასახადო და სხვა უფლებამოსილი ორგანოების მიერ საქართველოს კანონმდებლობით გათვალისწინებული გადასახდელად დარიცხული ან შემცირებული თანხები, ვადაგადაცილებული დავალიანებები, ზედმეტად გადახდილი თანხები, სანქციები, ჯარიმები და საურავები.

2. გპაბ არის საგადასახადო ინსპექციების პირველადი აღრიცხვის საბუთი, რომელშიც აღირიცხება გადამხდელის მიერ ბიუჯეტში ჩარიცხული საგადასახადო შემოსავლები და აისახება საგადასახადო ვალდებულებების დარიცხული და გადახდილი გადასახადის თანხები.

3. გადასახადის გადამხდელის/საგადასახადო აგენტის ან სხვა ვალდებული პირის მიერ საგადასახადო ვალდებულების შესრულების აღრიცხვის წარმოების მიზნით საგადასახადო ორგანოებში იხსნება გპაბ.

4. გპაბ – ში ყველა ჩანაწერი წარმოებს თავისდროულად და ქრონოლოგიური წესით, ჩანაწერის გაკეთების თარიღისა და დამადასტურებელი დოკუმენტაციის სისწორის დადგენით. პირადი აღრიცხვის ბარათებში ჩანაწერები კეთდება შესაბამისად „დარიცხულია“, „შემცირებულია“, „გადახდილია“ და „ბიუჯეტიდან უკან დაბრუნებულია ან გადატანილია“ გრაფების მიხედვით.

2) დარიცხვა-ჩარიცხვა-გადასახადები

1. ამ ინსტრუქციის თანახმად გადასახადის თანხების დარიცხვა გულისხმობს საგადასახადო ორგანოში კონკრეტულ საგადასახადო პერიოდისათვის გადასახადის გადამხდელის/საგადასახადო აგენტის, სხვა ვალდებული პირის კუთვნილი გადასახადის თანხის აღრიცხვას. დარიცხვა მოიცავს შესწორებით დარიცხვას და სავარაუდო დარიცხვას. ამასთან, შესწორებითი დარიცხვის შემთხვევაში უქმდება სავარაუდო დარიცხვა.

2. სახელმწიფო ხაზინიდან საგადასახადო ინსპექციაში მიღებული ყოველდღიური ამონაწერები საგადასახადო შემოსავლების შესახებ (მათ შორის გადამხდელთა საგადახდო დავალებები) შეიძლება წარმოდგენილი იქნას დამოწმებული ფორმით როგორც ქაღალდზე, ისე ელექტრონული ფორმით.

3. სახელმწიფო საზინიდან საგადასახადო ინსპექციაში მიღებული ყოველდღიური ამონაწერები საგადასახადო შემოსავლების შესახებ (მათ შორის გადამხდელთა საგადახდო დავალებები) შეიძლება წარმოდგენილი იქნას დამოწმებული ფორმით როგორც ქაღალდზე, ისე ელექტრონული ფორმით.

4. საგადასახადო დეკლარაციის წარმოდგენის ვალდებულება შეიძლება შესრულდეს არა უგვიანეს დადგენილი ვადის ბოლო სამუშაო დღის დამთავრებამდე. ამასთან, საგადასახადო დეკლარაციის დაზღვეული საფოსტო გზავნილით გაგზავნის შემთხვევაში დეკლარაციის გადაცემის თარიღად ითვლება საგადასახადო ორგანოს მიერ საფოსტო გზავნილის მიღების თარიღი, ხოლო თუ დოკუმენტის გადაცემისათვის დადგენილია ვადა, ამ შემთხვევაში ვალდებულების შესრულების თარიღად ითვლება დაზღვეული საფოსტო გზავნილის გაგზავნის თარიღი. ამასთან, გადამხდელი, რომელიც საგადასახადო დეკლარაციების წარმოდგენის სავადო დღის სამუშაო დროის დამთავრებამდე მივა საგადასახადო ინსპექციაში, მას მიეცემა საშუალება საგადასახადო დეკლარაციის ჩაბარებისა.

5. თუ გადასახადის თანხა არ არის გადახდილი საგადასახადო კანონმდებლობით დადგენილ ვადაში, ამ შემთხვევაში გადასახადის გადამხდელი/საგადასახდო აგენტი, სხვა ვალდებული პირი ვალდებულია გადაიხადოს შესაბამისი საურავი. საურავი ერიცხება საგადასახადო დავალიანების წარმოქმნის მომდევნო დღიდან თითოეული ვადაგადაცილებული დღისათვის და მისი გადახდა ხორციელდება საგადასახადო დავალიანების თანხაზე დანამატის სახით. ამასთან, ერთ რომელიმე გადასახადის თანხაზე საურავის დარიცხვის დროს გაითვალისწინება გადამხდელის სხვა გადასახადებში ზედმეტად გადახდილი თანხები. იმ შემთხვევაში, თუ ზედმეტად გადახდილი თანხების ჯამი აღემატება ვადაგადაცილებულ დავალიანებებს, ავტომატურად (პროგრამულად) შეწყდება საურავების დარიცხვა იმ გადასახადების მიხედვით, რომლებზედაც ირიცხება ვადაგადაცილებული დავალიანებები. თუ ვადაგადაცილებული დავალიანების საერთო თანხა აღემატება ზედმეტად გადახდილი გადასახადების საერთო თანხას, მაშინ საურავის ცალკეულ

6. გადასახადის გადამხდელთა პირადი აღრიცხვის ბარათებზე გადასახადების მიხედვით დავალიანებების არსებობის შემთხვევაში ბიუჯეტის დავალიანებათა დაფარვა უნდა განხორციელდეს შემდეგი თანმიმდევრობით:

7. ნარჩენ დავალიანებად ითვლება ნებისმიერი გადასახადის თანხა, რომელიც დარჩა გადაუხდელი გადასახადის გადამხდელის/საგადასახადო აგენტის, სხვა ვალდებული პირის მიერ კანონმდებლობით გათვალისწინებული ვადის გასვლის შემდეგ და განისაზღვრება როგორც სხვაობა გადასახდელად დარიცხულ გადასახადის თანხასა და ბიუჯეტსა და სახელმწიფო სპეციალურ ფონდებში ჩარიცხულ თანხას შორის.

1. გვაბ (დანართი 1) შედგება 26 სვეტისაგან (ნახ. 11);

ნახ. 11 გადამხდელის პირადი ბარათი

2. პირველ სვეტში აღინიშნება ოპერაციის ჩაწერის თარიღი – დღე, თვე, წელი. ჩანაწერები ამ სვეტში კეთდება ქრონოლოგიურად;

3. მე-2 სვეტში იწერება ოპერაციის შინაარსი, ინფორმაცია გადასახადის დარიცხვის, გადახდის, შემცირების, დაბრუნების, სხვა გადასახადებში გადატანის, გადასახადის გადახდის ვადის გაგრძელების, უიმედო ვალების ჩამოწერის და სხვათა შესახებ, შესაბამისი პერიოდის მითითებით. ამავე სვეტში მიეთითება დოკუმენტები, რომელთა საფუძველზეც განხორციელდა ეს ოპერაციები: საგადახდო დავალება, საგადასახადო დეკლარაცია, დაზუსტებული (დამატებითი) საგადასახადო დეკლარაცია, დასკვნა ზედმეტად გადახდილი თანხის დაბრუნების ან გადატანის თაობაზე, გადაწყვეტილებები სავარაუდო დარიცხვის, საგადასახადო შემოწმების აქტით დარიცხული ან შემცირებული თანხების, დავალიანების გადახდის ვადის გაგრძელების შესახებ, კანონიერ ძალაში შესული სასამართლოს/არბიტრაჟის დადგენილება ან განჩინება. ამ სვეტში შესატანი ჩანაწერების ტიპური ჩამონათვალი მოცემულია ამ ინსტრუქციის მე-4 დანართში.

4. მე-3 სვეტში „გადახდის ვადა და ჩარიცხვის თარიღი“ იწერება დარიცხული გადასახადის გადახდის ვადა, ხოლო ბიუჯეტის ანგარიშზე გადასახადის გადარიცხვის შემთხვევაში – ფაქტობრივად გადახდის თარიღი.

5. მე-4 სვეტში იწერება როგორც მიმდინარე თვეში გადასახდელად გამოანგარიშებული, ისე გასული პერიოდის მიხედვით დამატებით დარიცხული თანხები.

6. მე-5 სვეტი განკუთვნილია როგორც მიმდინარე თვის, ისე გასულ პერიოდზე ადრე დარიცხული თანხების შემცირებისათვის.

7. მე-6 სვეტში იწერება საბანკო დაწესებულების მიერ ბიუჯეტში ჩარიცხული გადასახადის თანხები. ჩანაწერები ამ სვეტში წარმოებს მხოლოდ სახელმწიფო ხაზინიდან მიღებული დოკუმენტების საფუძველზე.

8. მე-7 სვეტში – „ბიუჯეტიდან უკან დაბრუნებულია ან გადატანილია“ – იწერება ამ გადამხდელის მიერ ბიუჯეტის შემოსულობების ერთიანი ანგარიშის საგადასახადო შემოსავლის

ზედმეტად გადახდილი თანხის დაბრუნების ქვეანგარიშიდან უკან დაბრუნებული ან ბიუჯეტის მიმდინარე შემოსავლებიდან გადატანილი თანხის ოდენობა.

9. მე-8 სვეტში იწერება გადასახადის ნარჩენი დავალიანების თანხა, რომელიც დარჩა გადაუხდელი კანონმდებლობით გათვალისწინებული გადახდის ვადის გასვლის შემდეგ.

10. მე-9 სვეტში გამოიყოფა დავალიანების თანხა, რომელზედაც საურავების დარიცხვა არ ხდება სხვა გადასახად(ებ)ში არსებული საღდირებული ზედმეტად რიცხული თანხის ფარგლებში, აგრეთვე გაკოტრების ან რეაბილიტაციის რეჟიმში მყოფ გადამხდელთა ვადაგადაცილებული საგადასახადო დავალიანებების თანხა, რომლებსაც საგადასახადო კანონმდებლობის შესაბამისად საურავი არ ერიცხება.

11. მე-10 სვეტი გამოყოფილია გადამხდელის მიერ ამ გადასახადზე ზედმეტად გადახდილი ან წარმოქმნილი ზედმეტობის თანხის ოდენობის აღსარიცხავად.

12. სვეტები მე-11-16 განკუთვნილია ფინანსური სანქციებით დარიცხული ჯარიმების აღრიცხვისათვის: მე-11 – „დარიცხულია“, მე-12 – „შემცირებულია“, მე-13 – „გადახდილია ბიუჯეტში“, მე-14 – „უკან დაბრუნებულია“, მე-15 – „ნარჩენი“, მე-16 – „ზედმეტად გადახდილი“.

13. მე-17-20 სვეტები განკუთვნილია ვადაგადაცილებული დავალიანების გამო წარმოშობილი საურავების აღრიცხვისათვის.

14. მე-17 სვეტში მიეთითება დროის პერიოდი, რომლის განმავლობაშიც ხდება საურავების დარიცხვა. მე-18-ე სვეტში ჩაიწერება დარიცხული საურავის თანხის ოდენობა, მე-19 სვეტში იწერება გადახდილი ან დაბრუნებული (ნიშნით „-“) საურავი, მე-20 სვეტში აღირიცხება გადაუხდელი საურავის ნაშთი;

15. ადრე დარიცხული საურავის შემცირების თანხები (სვეტი 18) ან შემოსული (ჩარიცხული) საურავები (სვეტი 19) ძირითადი გადასახდელების დასაფარავად ჩაიწერება შესაბამისად ამავე სვეტებში (მე-19 სვეტში ნიშნით „-“). ამავდროულად მე-19 სვეტში „-“ ნიშნით ჩაწერილი თანხა იწერება მე-6 სვეტში როგორც ძირითადი

გადასახდელების შემოსავლები და მე-2 სვეტში აღინიშნება ასე: „გადატანილია ძირითადი გადასახდელი თანხის დაფარვის ანგარიშზე საურავის ანგარიშიდან“.

16. 21-ე – 24-ე სვეტები განკუთვნილია გადამხდელის ჯამური ანგარიშსწორების საღღოსათვის ძირითადი გადასახადის, ფინანსური სანქციის სახით ჯარიმისა და საურავების ჩათვლით.

17. საურავის დარიცხვა საგადასახადო ინსპექციებში წარმოებს ყოველი ოპერაციის დროს, როცა იცვლება ნარჩენის ოდენობა, აგრეთვე ყოველი თვის ბოლოს გადასახადის გადახდის ვადის დადგენილი თარიღიდან, ან საურავის უკანასკნელი დარიცხვის დღიდან თვის ბოლო რიცხვების ჩათვლით

18. 25-ე სვეტში აისახება გადახდილი თანხები ნაზარდი ჯამით წლის დასაწყისიდან.

3) რეზიუმე გადამხდელის ბარათის შესახებ

იმისათვის, რომ გადავიდეთ შემდეგ ნაბიჯზე, ანუ ჩვენი კვლევის ამოცანაზე და მიზანზე, საბოლოოდ ჩამოვაყალიბოდ ანუ მოვახდინოთ რეზიუმირება, თუ რას გვაძლევს გადამხდელის პირადი აღრიცხვის ბარათი.

1. გვაქს ინფორმაცია თვითოულ გადამხდელზე მისი ბიუჯეტთან ანგარიშსწორების შესახებ.

2. გვაქვს ინფორმაცია რაიონების რეგიონების და საქართველოს მასშტაბით.

3. შეგვიძლია გავაკეთოდ პროგნოზი ბიუჯეტში შემოსავლების შესახებ.

3.2. საგადასახადო სისტ. ინფ. უზრუნველყოფის სისტემის და წინააღმდეგობითობის მოდელების აგების პრინციპები

განვიხილოთ კონკრეტული ბარათის წარმოების პროცესი: 2005 წლის იანვრის თვის პირველ რიცხვში კონკრეტულ გადამხდელს ათვლის სათავეში ჰქონდა შემდეგი ნაშთი: ძირითადი ნარჩენი 100 ლარი (წარმოქმნილი კუთვნილი გადასახადის (დარიცხული) თანხის გადაუხდელობით), საურავი 20 ლარი (ყოველ ვადაგადაცილებულ

დღეზე ძირითადი თანხის 0.07%, აისახება ნაზარდი ჯამით), ამის შედეგად ვლინდება, რომ საერთო დავალიანება ამ თარიღისათვის წარმოადგენს 170 ლარს.

2005 წლის 25 იანვრისათვის გადამხდელს ჰქონდა წარმოდგენილი დეკლარაცია თანხაზე 350 ლარი, რის შედეგადაც ძირითადი ნარჩენის თანხა გახდა 450 ლარი, სანქციის ნარჩენი 50 ლარი – რადგან სანქციის ნარჩენი წარმოებს ნაზარდი ჯამით და რადგან 1 იანვარს გადაუხდელი საურავი ჰქონდა 20 ლარი და 25 იანვრამდე დავალიანება იყო 100 ლარი, ამიტომ დარიცხული საურავი გამოიანგარიშება შემდეგი ფორმულით $100 \times 0.07\% \times 25$ (ნახ.12).

ამ ოპერაციების მიხედვით შეგვიძლია ვივარაუდოთ, თუ გადამხდელს გააჩნია დავალიანება და არის აქტიური გადამხდელი, მოსალოდნელია, რომ ბიუჯეტი მიიღებს მისგან ნავარაუდებ თანხას. ყოველი გადამხდელისაგან მიღებული ინფორმაციით ვაჯგუფებთ საქმიანობის სახეებს, რის შედეგადაც ვღებულობთ მონაცემებს სხვადასხვა საქმიანობაზე²⁹.

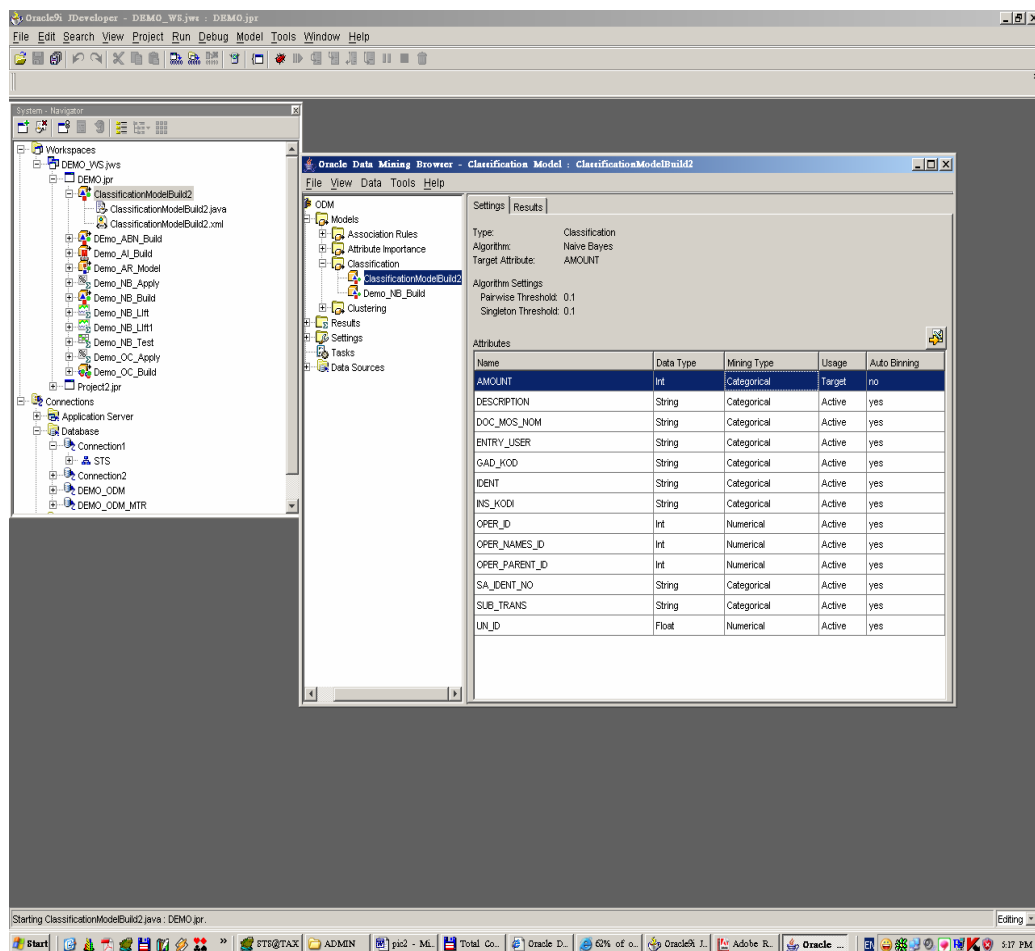
ულის პირადი აღრიცხვის ბარათი														დანართი 1											
ფილიალის კოდი														პერიოდი											
ულის დასახელება														გადასახადის დასახელება											
ოპერაციის მონარსი	გადასახადის დასახელება	ანგარიშსწორება ძირითადი გადასახადის მიხედვით								ანგარიშსწორება ფინანსური საშუალების სახით გარიგების თანხაზე						ანგარიშსწორება საურავის დარიცხვაზე და გადახდაზე				სულ ანგარიშსწორების სალდო				წლით დათვნილია სალდო	წლით დათვნილია სალდო
		დარიცხულია	შეტანილია	გადახდილია	უკმაღალოდ დარჩენილია	ანგარიშსწორების სალდო				დარიცხულია	შეტანილია	გადახდილია	უკმაღალოდ დარჩენილია	ანგარიშსწორების სალდო				ანგარიშსწორების სალდო							
						ნარკვეთი	8.3 ნარკვეთი თანხა	როგორც სავაჭრო	უკმაღალოდ დარჩენილია					დარიცხულია	შეტანილია	გადახდილია	უკმაღალოდ დარჩენილია	დარიცხულია	შეტანილია	გადახდილია	უკმაღალოდ დარჩენილია				
2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
ნაშთი წლის დასაწყისი						100							50					20.00	170.00	150	20.00				
ფარდობა XII	15.01.06	350				450							50		01.01-25.01	4.13		24.13	524.13	500	24.13				
გადახდილია	30.01.06			200		250							50		25.01-30.01	1.58		25.71	325.71	300	25.71		200		
გადახდილია 320 ლარი	20.02.06			250		0					50		0		30.01-20.02	1.90	20	7.61	0.00	0	7.61		520		
ფარდობა I სანქცია	01.03.06					0			130				130				0	7.61	137.61	130	7.61		520		
ფარდობა საურავი	01.03.06					0							130				100	0	107.61	237.61	130	107.61		520	
ფარდობა აქტი	01.03.06	230				230							130				0	0	107.61	467.61	360	107.61		520	
გამოფასებულია ნაშთი გასაღებების გარეშე																									
გამოფასებულია ნაშთი გასაღებების შიგნით																									
ჯამით პერიოდის ბოლისათვის	580		450			230			130		50		130		01.03-01.05	14.49	20	122.10	467.61	360	122.1		520		
საურავი რომელიც არ დარიცხა გასაღებების შედეგად																									
გადახდილია მომსახურების სამმართველოს უფროსი																									
(ბუღალტრ-თარიღი)																									

ნახ.12 კონკრეტული ბარათის წარმოების პროცესი

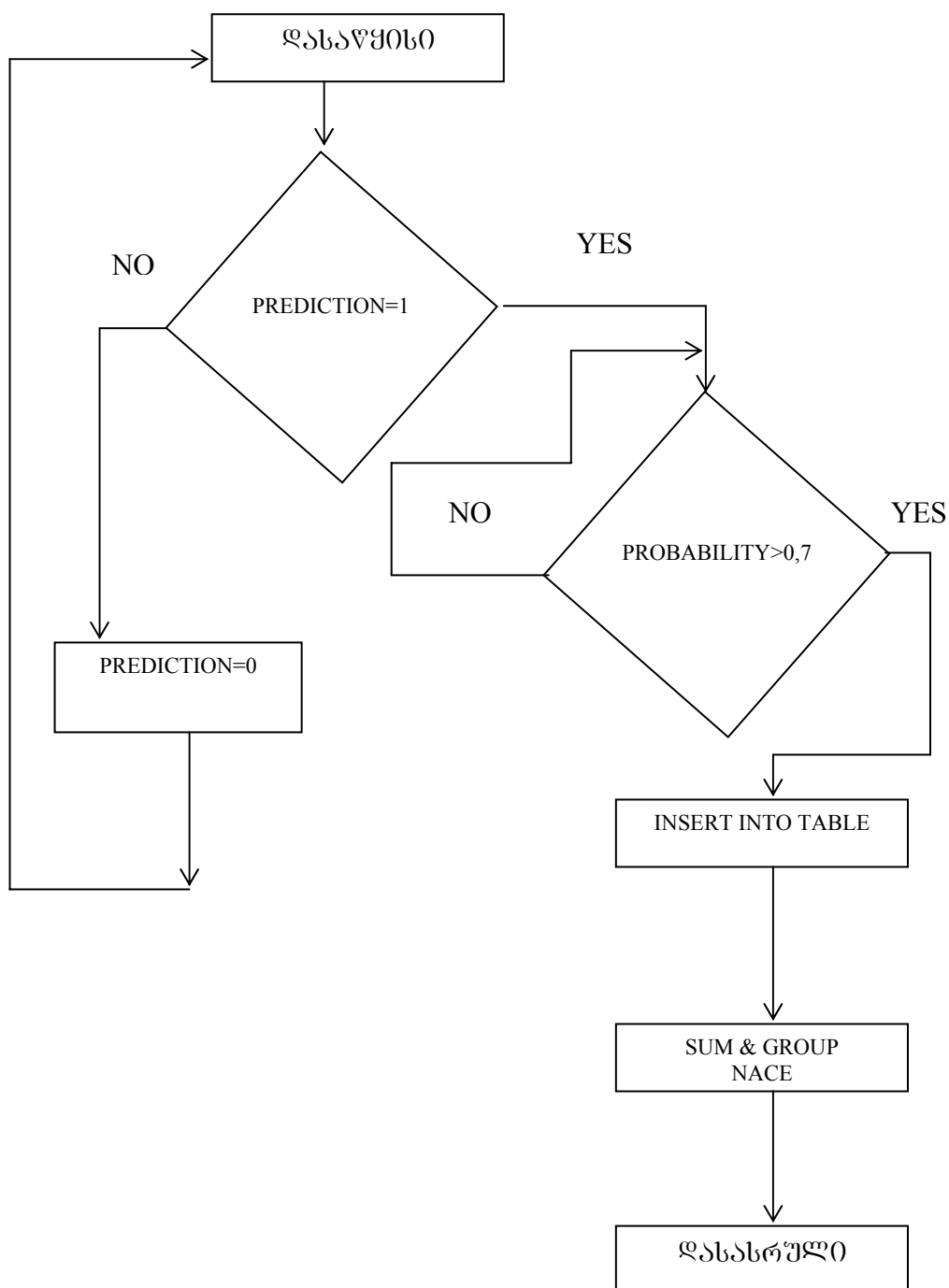
²⁹ ხ. აბულაშვილი, ე. მაღრაძე: ოპერაციების კვლევა Data Mining ტექნოლოგიის საფუძველზე, -2006 | No.4(11) [2006.12.30], (http://gesj.internet-academy.org.ge/gesj_articles/1234.pdf)

3.3. დამუშავება და კვლევა

ჩვენს შემთხვევაში ვიყენებთ Data Mining –ის კლასიფიკაციის მეთოდს, რომელიც იყენებს ჩვენს მიერ შექმნილ ალგორითმს (ნახ.13, ნახ.14). PREDICTION სვეტში მიღებული 1-ანი ნიშნავს, რომ პროგნოზი არსებობს, ხოლო მიღებული 0 გულისხმობს, რომ გადამხდელის მონაცემები პროგნოზში არ ღებულობს მონაწილეობას.



ნახ.13 წარმოდგენილია კლასიფიკაციის მოდელის შექმნის პროცესის ფრაგმენტი. ინფორმაციის მომდევნო მომზადება ჩვენს მიერ შექმნილი ალგორითმისათვის



ნახ.14 ჩვენს მიერ შექმნილი ალგორითმი

სვეტი **PROBABILITY** გვიჩვენებს პროგნოზის შესრულების ალბათობას შესაბამისად ჩვენ ვირჩევთ იმ გადამხდელებს, რომელთა **PREDICTION** არის 1 და **PROBABILITY** –ის მნიშვნელობა მაქსიმალური.

პროგნოზირებისათვის გამოვიყენებთ ცხრილ **T_PROB**-ს, სადაც წარმოდგენილია შემდეგი ველები **ins_kodi** (ინსპექციის კოდი),

T_PROB				
ins_kodi	sa_ident_no	nace_code	gad_kod	prediction
002	202161098	33103	01	1
002	202234005	29564	07	0
007	109546564	33103	01	1

ცხრილი 5. ცხრილი T_PROB

(გადამხდელის საიდენტიფიკაციო ნომერი), nace_code (საქმიანობის სახე - კოდი), gad_kod (გადასახადის სახე - კოდი), prediction (პროგნოზი) (ცხრილი5).

ინფორმაციას ვწერთ მოცემულ ცხრილში, რის შემდეგაც ვახდენთ ინფორმაციის დაჯგუფებას ეკონომიკური საქმიანობის მიხედვით, ინპექციებისა და გადასახადების მიხედვით. ეს გვაძლევს საშუალებას ვივარაუდოთ თუ რომელ სფეროს (nace_code) უნდა მიეკცეს გაცილებით მეტი ყურადღება, თუ როგორ განაწილდება ეს მონაცემები საგადასახადო ინსპექციებს შორის, ანუ სად უფრო მოსალოდნელია საგადასახადო ჩარიცხვების ზრდა და ამის შედეგადად, თუ რომელ გადასახადზე უნდა გამახვილდეს ყურადღება.

რეზულტატების მისაღებად სრულდება შემდეგი select-ი.

Select sum(sa_ident_no), ins_kodi , nace code, prediction, gad_kod, amount

From t_prob

Group by nace code.

ჩვენს მიერ დაწერილი select-ის შესაბამისად ვღებულობთ ინფორმაციას ყოველი საქმიანობის რაოდენობრივი მაჩვენებლების შესახებ.

კერძო შემთხვევისათვის ვწერთ შემდეგ select-ს:

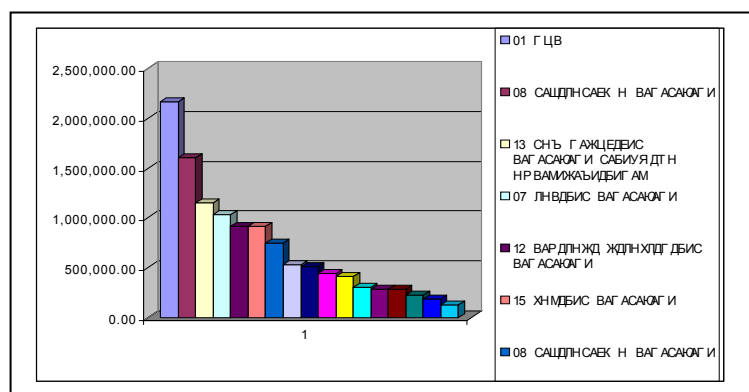
```
select
un_id,ins_kodi, SA_IDENT_NO, sts.utl_unid.get_nace(un_id) ,
sts.utl_unid.Get_Activity(sts.utl_unid.get_nace(un_id)) ,gad_kod ,
sts.utl_unid.get_gad_name(gad_kod), sum(amount)from tax_operations where sa_ident_no =
'202161098'
and
OPER_PARENT_ID = 8group by un_id,ins_kodi, SA_IDENT_NO,
sts.utl_unid.get_nace(un_id) , sts.utl_unid.Get_Activity(sts.utl_unid.get_nace(un_id))
,gad_kod , sts.utl_unid.get_gad_name(gad_kod)
```

რის შედეგადაც ვღებულობთ შემდეგ მონაცემებს (ნახ.15):

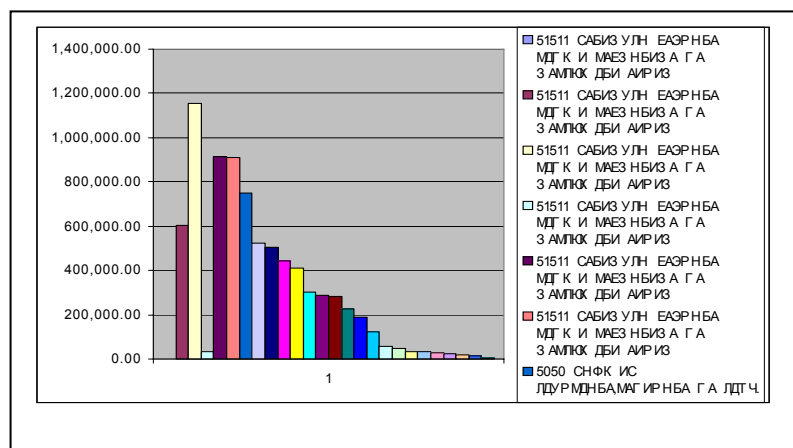
[illegible]

ნახ.15. ინფორმაცია ყოველი საქმიანობის რაოდენობრივი მაჩვენებლების შესახებ.
კერძო შემთხვევა

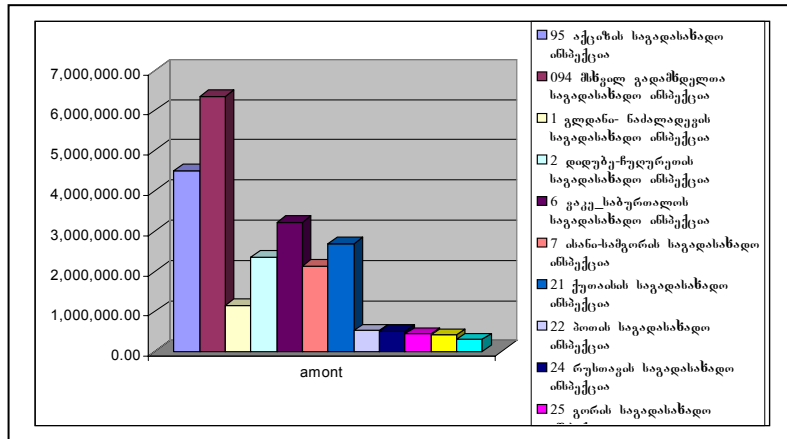
მიღებული მონაცემები შეგვიძლია აგრეთვე განვიხილოთ შემდეგი chart-ების მიხედვით (ნახ.16, ნახ.17, ნახ.18).



ნახ.16 დიაგრამა მონაცემების განხილვა გადასახადების მიხედვით

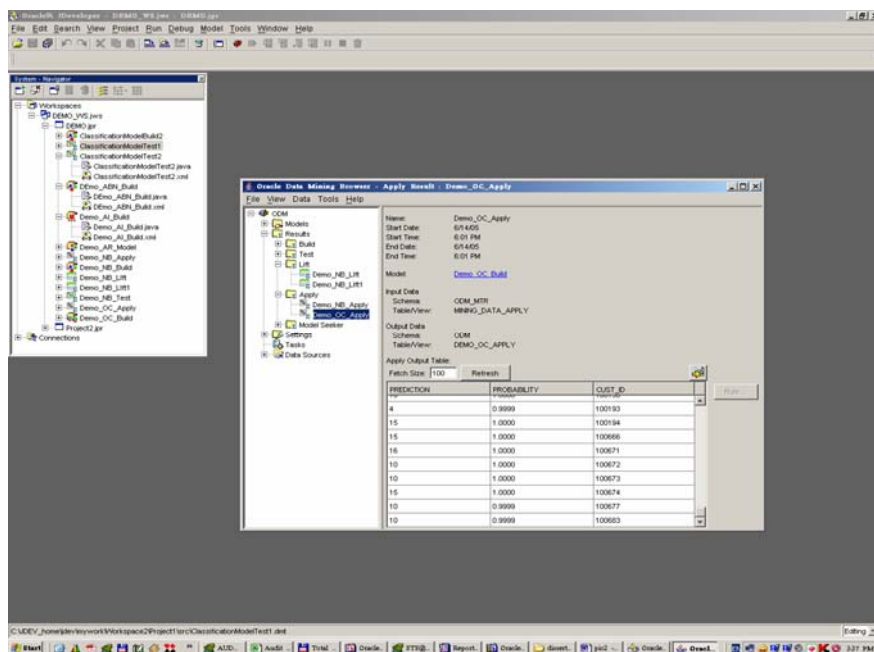


ნახ.17 დიაგრამა მონაცემების განხილვა საქმიანობის სახის მიხედვით



ნახ.18 დიაგრამა მონაცემების განხილვა ინპექციების მიხედვით

აგრეთვე კრიტერიუმების შერჩევის საფუძველზე ვღებულობთ პროგნოზს, რომლის შედეგადაც შეგვიძლია გამოვაგლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი (ნახ.19).



ნახ.19 ალგორითმის კრიტერიუმების შერჩევის საფუძველზე მიღებული პროგნოზი

ამ ინფორმაციის საფუძველზე შეგვიძლია დავადგინოთ, თუ რომელ დარგს შეუძლია მოუტანოს ბიუჯეტს მეტი შემოსავალი და რომელ დარგს უნდა მიექცეს ყურადღება.

დასკვნა

საბოლოოდ შეიძლება ითქვას, რომ ერთ-ერთი ყველაზე ძირითადი სახელმწიფო ორგანოს ფუნქციონირების ასპექტი, რომელიც არეგულირებს გადასახადების მოგროვების პროცესს, წარმოადგენს ინფორმაციული ტექნოლოგიების განვითარება, რომელიც იძლევა საშუალებას უფრო ოპერატიულად და ზუსტად დავამუშაოთ ინფორმაციის დიდი ნაკადები. სადისერტაციო ნაშრომში დასმული ამოცანების გადაწყვეტისას გამოვლინდა შემდეგი საკვანძო დასკვნები:

1. ჩამოყალიბდა ინფორმაციულ ანალიტიკური სისტემის არქიტექტურისა და მისი რეალიზაციისათვის არსებული პროდუქტების, ინსტრუმენტების კომპლექსური ანალიზი და საფეხური, რომელიც Data Mining ტექნოლოგიას უჭირავს ანალიტიკური სისტემების არქიტექტურაში;
2. მკვეთრად აღიწერა Data Mining ტექნოლოგიის არსის და მასთან არსებული სხვა საკვანძო ტერმინოლოგიის განსაზღვრებები;
3. განისაზღვრა თუ რა ადგილი უჭირავს Data Mining ტექნოლოგიას მისი გამოყენებისას ამა თუ იმ სფეროებში;
4. შევისწავლეთ, თუ რა სარგებელი მოუტანა ამ ტექნოლოგიის გამოყენებამ საბაზრო ეკონომიკას;
5. გამოიკვეთა Data Mining ტექნოლოგიის გამოყენების პერსპექტიულობა საქართველოს რესპუბლიკის საგადასახადო სისტემაში;
6. შევიმუშავეთ ალგორითმი, რომლის მეშვეობითაც იოლად ვახდენთ ინფორმაციის მომზადებას, კლასიფიკაციის მოდელის შექმნის პროცესისათვის.
7. შესაძლებელი გახდა კრიტერიუმების შერჩევის საფუძველზე პროგნოზის ჩატარება, რომლის შედეგადაც შეგვიძლია გამოვავლინოთ ის დარგები, რომლებიც არიან ცუდად განვითარებულნი. რაც მოგვცემს შესაძლებლობას დავადგინოთ, თუ რომელ დარგს შეუძლია მოუტანოს ბიუჯეტს მეტი შემოსავალი და რომელ დარგს უნდა მიექცეს ყურადღება.
8. გამოიკვეთა პრობლემები მოდელის შექმნისას: მონაცემთა ბაზა Oracle9-ზე გართულდა ODM10-ის დაყენება, რის გამოც ჩვენ დაგვჭირდა სპეციალური აპლიკაციების დაყენება.
9. დაინერგა შემუშავებული კვლევის რეზულტატები საქართველოს შემოსავლების სამინისტროს ინფორმაციული ტექნოლოგიების სამმართველოში.

ბამოყენებული ლიტერატურა:

1. საქართველოს რესპუბლიკის საგადასახადო კოდექსი
2. საქართველოს ფინანსთა სამინისტროს ბრძანება №228 “გადამხდელთა პირადი აღრიცხვის ბარათების წარმოების წესის შესახებ”//2005 წ. 7 აპრილი, ქ. თბილისი
3. Артемьев В. Что такое BUSINESS INTELLIGENCE? //Открытые системы, 2003, №4.
4. Архипенков С., Голубев Д., Максименко О. Хранилища Данных. От концепции до внедрения/ Под общ. Ред. Архипенкова С.Я. // М.: ДИАЛОГ-МИФИ, 2002 г. 4-82 გვ.
5. Архипенков С. Хранилища данных. От концепции до внедрения. 2002 г.
6. Банерджи Д. Витрины данных и системы генерации отчетов //Tate Bramald Consultancy Co.
7. Бергер Ч. Data Mining от Oracle: настоящее и будущее/ Oracle’s Data Mining Solutions – Oracle OpenWorld White Paper// Источник: Конференция OOW2000, San Francisco, доклад 134 /20 января 2001 г.
8. Большаков П.С. Уникальные возможности STATISTICA Data Miner
9. Галахов И. В., Волков И. Ю. Архитектура современной информационно-аналитической системы // Директор ИС №3, 2002г. 15-24 გვ.
10. Гик Дж., ван. Прикладная общая теория систем. - М.: Мир, 1981
11. Дюк В. Data Mining – состояние, проблемы, новые решения// 1999г., (http://on.wplus.net/sparm/science/Data_mining.html). უკანასკნელად იქნა გადამოწმებული 01.04.2008
12. Дживинаццо В. Построение Хранилища данных для web-среды. Октябрь 2002 год
13. Джукич Н. Разработка одношагового Хранилища данных для поставщика услуг финансовых приложений. Январь 2002 г.
14. Дюк В., Самойленко А., Data mining: учебный курс. - СПб: Питер, 2001г.
15. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах. - Открытые системы, № 4, 1997.
14. Кнут Д. Искусство программирования, том 1. Основные алгоритмы, 3-е изд.: - М.: "Вильямс", 2000г. 203-215 გვ.
16. Комафорд К. Корпоративная отчетность: Серверная архитектура для распределенного доступа к информации. // Открытые системы, 1999, 2.
17. Корнеев В.В., Гареев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. // М.: Нолидж, 2001 г.
18. Коровкин С. Д., Левенец И. А., Ратманова И. Д., Старых В. А., Щавелёв Л. В. Решение проблемы комплексного оперативного анализа информации хранилищ данных // СУБД. - 1997. - № 5-6.
19. Кречетов Н. Продукты для интеллектуального анализа данных// Рынок программных средств, N14-15_97.
20. Кречетов Н., Иванов П. Продукты для интеллектуального анализа данных // ComputerWeek-Москва. - 1997. - № 14-15.
21. Львов В. Создание систем поддержки принятия решений на основе хранилищ данных, СУБД//1997, №3.
22. Островский Е. В. Порядок разработки ETL- процессов//7 октября 2004 .

23. Пржиялковский В. В. Сложный анализ данных большого объема: новые перспективы компьютеризации // СУБД. - 1996. - № 4.
24. Пройдаков Э. Что такое Data Mining?// [PC Week/RE 99/26](#). უკანასკნელად იქნა გადამოწმებული 01.04.2008
25. Раден Н. Данные, данные и только данные // ComputerWeek-Москва. - 1996. - № 8.
26. Саймон А. Р. Стратегические технологии баз данных: менеджмент на 2000 год. //Пер с англ. — М: Финансы и статистика, 1999г.
27. Салливан Т. Это надо рисовать: Программное обеспечение анализа данных становится более выразительным. // ComputerWorld Россия, 2000, 42.
28. Самойленко А. Data mining: учебный курс. // СПб: Питер, 2001г.
29. Самохвалов Р. Системы поддержки принятия решений Oracle (Часть 1) // Oracle СНГ 2003г.
30. Сахаров А.А. Концепции построения и реализации информационных систем, ориентированных на анализ данных //(статья опубликована на [www.olap.ru](#)), 1996г. უკანასკნელად იქნა გადამოწმებული 01.04.2008
31. Сахаров А.А., Принципы проектирования и использования многомерных баз данных //СУБД №3, 1996
32. Семененко А.В. Витрины данных – новая технология обработки, анализа и хранения информации для систем поддержки принятия решений//(статья с сайта [www.nit7.artdesign.ru](#)). უკანასკნელად იქნა გადამოწმებული 10.04.2008
33. Смирнов Н. Система для «Манчестера» 18.08.2003// [Еженедельник "Computerworld", #30, 2003 год](#) // Издательство "Открытые системы" უკანასკნელად იქნა გადამოწმებული 10.04.2008
34. Смол Р., Two Crows Corporation// (источник: <http://www.management-magazine.ru>), Апрель, 2003 г. უკანასკნელად იქნა გადამოწმებული 10.04.2008
35. Спирли Э. Корпоративные хранилища данных. Планирование, разработка, реализация. ..Том.1: Пер. с англ. М. Вильямс, 2001г. 102-230 გვ.
36. Туманов В. Data Warehouse: с чего начать?// PC Week/RE 1998 г.
37. Тью Дж. Каждому пользователю - свое представление данных // ComputerWeek-Москва, 1996, № 38
38. Тью Дж. Инструменты для анализа информации на настольных ПК // ComputerWeek-Москва, 1996, № 38.
39. Федоров А., Елманова Н. Введение в OLAP: часть1. Основы OLAP// КомпьютерПресс 4' 2001
40. Чаудхури С., Дайа У., Ганти В. Технология баз данных в системах поддержки принятия решений //22 января 2002г,(статья с сайта [www.citforum.ru](#)). უკანასკნელად იქნა გადამოწმებული 10.04.2008
41. Alalouf C. Hybrid OLAP// St. Laurent, Canada: Speedware Corporation Inc., 1997.
42. Almeida M. S., Ishikawa M., Reinschmidt J., Roeber T. Getting Started with Data Warehouse and Business Intelligence// IBM Red Books
43. An Introduction to Multidimensional Database Technology// Kenan Systems Corporation, 1995.
44. Boulding K. E. General Systems Theory. The Skeleton of Science//Management Science, 2, 1956.

45. Caserta J., Kimball R. The Data Warehouse ETL Toolkit//Practical Techniques for Extracting, Cleaning September 13, 2004.
46. Celko Trees in SQL// Intelligent Enterprise, October 20, 2000.
47. Codd E. F., Codd S.B. Providing OLAP. On-line Analytical Processing to User-Analysts: An IT Mandate// C. T. Salley, E. F. Codd & Associates, 1993.
48. Data Warehouse Issues// Butler Group Co., UK 2004.
49. Demarest M. Building the Data Mart // DBMS. , July,1994.
50. Electronic Textbook Copyright StatSoft, Inc., 1984-2003.
51. Gartner Research, "BI Magic Quadrants: A 'Recession-Proof' Market Challenged", 17.07.2001
52. Han J. OLAP Mining: An Integration of OLAP with Data Mining// IFIP, 1997, (<ftp://ftp.fas.sfu.ca/pub/cs/han/kdd/olapm.ps.gz>). უკანასკნელად იქნა გადამოწმებული 10.04.2008
53. Harinarayan V., Rajaraman A., Ullman J. D. Implementing Data Cubes Efficiently // SIGMOD Conference. - Montreal, CA.,1996.
54. Inmon W. H. Building the Data Warehouse. New York// John Wiley & Sons, Inc. 1993. 95-210 გვ.
55. Inmon W. H., Richard D. Hackathorn Using the Data Warehouse Hardcover - July 13, 1994.
56. Inmon W.H. What is Data Warehouse 230-240 გვ.
57. Inmon W.H., Caplan J. H. Information Systems Architecture//Development in the 90's (Hardcover) Wiley September 24, 1993.
58. Kimball R. Help for Hierarchies// DBMS, September,1998.
59. Kimball R., Slowly Changing Dimensions// DBMS, April ,1996.
60. Kimball R. The Data Warehouse Toolkit// Practical Techniques for Building Dimensional Data Warehouses John Wiley & Sons; Bk&CD Rom edition February 2, 1996.
61. Knowledge Discovery Through Data Mining: What Is Knowledge Discovery? //Tandem Computers Inc., 1996.
62. Liautaud B., Hammond M. e-Business Intelligence: Turning Information into Knowledge into Profit. McGraw-Hill, 2001.
63. Mumick I. S., Quass D., Mumick B. S. Maintenance of Data Cubes and Summary Tables in a Warehouse// Stanford University, Database Group, 1996 (<http://www-db.stanford.edu/pub/papers/cube-maint.ps>). უკანასკნელად იქნა გადამოწმებული 10.04.2008
64. Newquist H. P. Data Mining: The AI Metamorphosis // Database Programming and Design, 1996, № 9.
65. OLAP for the Masses// Business Objects S. A., 1996 (<http://www.businessobjects.com/product/olap/olap.htm>). უკანასკნელად იქნა გადამოწმებული 10.04.2008
66. ORACLE. Data Mining. Administrator's Guide//10g release2(10.2) B1433B-01, June, 2002.
67. Parsaye K. Characterization of Data Mining Technologies and Processes // The Journal of Data Warehousing, 1998.-№ 1.
68. Parsaye K. OLAP and Data Mining: Bridging the Gap // Database Programming and Design, - 1997, - № 2.

69. Parsaye K. Surveying Decision Support: New Realms of Analysis // Database Programming and Design. - 1996. - № 4.
70. Pendse N. OLAP Architectures// The OLAP Report, (<http://www.olapreport.com/Architectures.htm#top>). უკანასკნელად იქნა გადამოწმებული 10.04.2008
71. Poly Analyst Решения для лучших в бизнесе// Компания Мегапьютер Интеллидженс, Аналитические программы и системы, Москва 2001.
72. Reinschmidt J., Francoise A. Business Intelligence Certification Guide// IBM Red Books.
73. Russom P. Demand Is Growing For Traditional And New Uses Of ETL//IT View Market Overview, April 25, 2005.
74. Thomsen E. OLAP Solutions: Building Multidimensional Information Systems// Wiley Computer Publishing, 1997.
75. Tukey J. Exploratory Data Analysis// NY: McMillan, 1973.
76. ხ. აბულაშვილი, ე. მაღრაძე: ოპერაციების კვლევა Data Mining ტექნოლოგიის საფუძველზე, -2006 | No.4(11) [2006.12.30], (http://gesj.internet-academy.org.ge/gesj_articles/1234.pdf)- უკანასკნელად იქნა გადამოწმებული 11.04.2008
77. Абулашвили Х., Маградзе Е.: Построение и реализация Хранилищ Данных,- “INTELECT”/ -2006| №1(24)
78. www.taxdep.ge- უკანასკნელად იქნა გადამოწმებული 11.04.2008
79. www.mof.ge- უკანასკნელად იქნა გადამოწმებული 11.04.2008
80. www.sql.ru- უკანასკნელად იქნა გადამოწმებული 05.04.2008
81. www.olap.ru- უკანასკნელად იქნა გადამოწმებული 11.04.2008
82. <ftp://ftp.fas.sfu.ca/pub/cs/han/kdd/olapm.ps.gz>- უკანასკნელად იქნა გადამოწმებული 01.07.2008
83. www.olapreport.com - უკანასკნელად იქნა გადამოწმებული 01.04.2008
84. www.cittforum.ru - უკანასკნელად იქნა გადამოწმებული 11.04.2008
85. www.interface.ru- უკანასკნელად იქნა გადამოწმებული 01.04.2008
86. www.iso.ru- უკანასკნელად იქნა გადამოწმებული 01.04.2008
87. www.oracle.ukrsat.com- უკანასკნელად იქნა გადამოწმებული 01.04.2008
88. www.garshin.ru- უკანასკნელად იქნა გადამოწმებული 01.04.2008
89. www.olapcouncil.org- უკანასკნელად იქნა გადამოწმებული 01.04.2008
90. www.rdtex.ru- უკანასკნელად იქნა გადამოწმებული 11.04.2008

91. www.datahousecorp.com- უკანასკნელად იქნა გადამოწმებული 05.04.2008
92. www.glossary.basegroup.ru- უკანასკნელად იქნა გადამოწმებული 05.04.2008
93. www.itstan.ru- უკანასკნელად იქნა გადამოწმებული 01.04.2008
94. www.db.stanford.edu- უკანასკნელად იქნა გადამოწმებული 01.04.2008
95. www.pc.week.ru- უკანასკნელად იქნა გადამოწმებული 01.04.2008
96. www.forrester.com- უკანასკნელად იქნა გადამოწმებული 11.04.2008
97. www.businessobjects.com - უკანასკნელად იქნა გადამოწმებული 01.04.2008
98. www.olapreport.com- უკანასკნელად იქნა გადამოწმებული 01.04.2008
99. www.oracle.com- უკანასკნელად იქნა გადამოწმებული 01.04.2008
100. www.osp.ru- უკანასკნელად იქნა გადამოწმებული 01.04.2008
101. www.statsoft.com- უკანასკნელად იქნა გადამოწმებული 01.04.2008
102. www.management-magazine.ru- უკანასკნელად იქნა გადამოწმებული 11.04.2008
103. www.bipartner.ru- უკანასკნელად იქნა გადამოწმებული 11.04.2008
104. www.rnkc.ru- უკანასკნელად იქნა გადამოწმებული 11.04.2008
105. www.twocrows.com- უკანასკნელად იქნა გადამოწმებული 11.04.2008