

## **ინფორმაციის კონვერტაცია მონაცემთა საცავში**

ლია პეტრიაშვილი, მარინა კაშიბაძე, მაია ოხანაშვილი  
საქართველოს ტექნიკური უნივერსიტეტი

### **რეზიუმე**

ინფორმაციის კონვერტაცია დიდი კორპორაციული სისტემების მონაცემთა საცავებისათვის საკმაოდ რთული პროცესია და ხასიათდება მაღალი შრომატევადობით. ნაშრომში წარმოდგენილია მონაცემთა საცავის კონვერტაციის პროცესის სტრუქტურა და გამოკვლეულია მისი კავშირი ორგანიზაციულ სისტემებში გადაწყვეტილების მიღების პროცესის დროით მახასიათებლებთან.

**გასაღებური სიტყვები:** მონაცემთა საცავი. კორპორაციული სისტემა. მონაცემთა კონვერტაცია. გადაწყვეტილების მიღების სისტემა.

### **1. შესავალი**

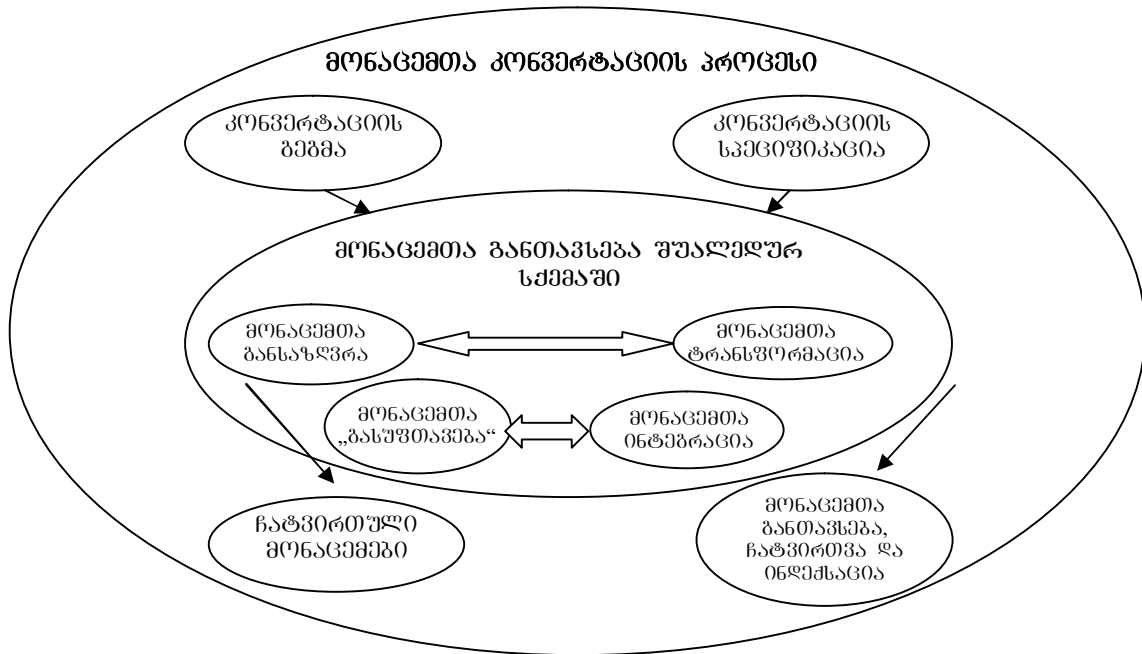
მონაცემთა საცავისათვის კონვერტაციის პროცესი საკმაოდ რთული და შრომატევადი პროცესია. მონაცემთა ხარისხიან კონვერტაციაზე დამოკიდებულია კორპორატიული სისტემებში ოპტიმალური გადაწყვეტილების მიღება.

კონვერტაციის პროცესის დაწყების წინ საკონვერტაციო ჯგუფი ქმნის პროექტს და წარმოადგენს მონაცემთა ფიზიკურ მოდელს, რაც შემდგომ აისახება მიზნობრივ სქემაზე. მონაცემთა კონვერტაციის ჯგუფი უზრუნველყოფს ბიზნესისა და ტექნიკის სფეროს წარმომადგენლებთან შეთანხმებას და საცავის სპროექტო სტრუქტურის განსაზღვრას, სადაც უნდა მოხდეს საწყისი მონაცემების ანალიზი და ბოლოს განსაზღვრავს მონაცემთა კონვერტაციის ლოგიკას. ამას გარდა ეს ჯგუფი იყენებს მონაცემთა მიგრაციის, ტრანსფორმაციის და გაწმენდის ინსტრუმენტებს. მონაცემთა კონვერტაციის წარმატების მთავარი და კრიტიკული ფაქტორი მდგომარეობს იმაში, რომ კონვერტაციის ჯგუფის ყველა წევრისათვის გასაგები უნდა იყოს სისტემის მიმართ დაყენებული მოთხოვნა. კონვერტაციის პროცესის დროს თავს იყრის რამდენიმე ალტერნატიული ვარიანტი, როდესაც მოთხოვნა სწორად არის გაგებული, მინიმალური დროის განმავლობაში შესაძლებელი ხდება კონვერტაციის სწორი ფორმის შერჩევა.

კონვერტაციის წინასწარ შედგენილი გეგმა, საცავში განსაზღვრავს საწყის მონაცემთა მიგრაციის ოპტიმალურ გზას, წინასწარი გეგმა ასევე ფლობს რესურსებს რომლის დახმარებითაც შესაძლებელი ხდება საწყის მონაცემთა მოცულობის განსაზღვრა, ერთმანეთისგან განსხვავებულ პლატფორმათა და მონაცემთა მოპოვების მეთოდების რაოდენობისა და ტიპის განსაზღვრა, მონაცემთა საცავის სტრუქტურის და აუცილებელ მონაცემთა აგრეგაციის მოცულობის განსაზღვრა.

გეგმა ყოველი საწყისი სისტემისათვის ასახულებს პლატფორმას, რომელიც უზრუნველყოფს მონაცემთა მოპოვების მეთოდსა და პროგრამირების ენას. თუ ჩვენი ჯგუფი,

რამდენიმე საწყისი სისტემიდან შეარჩევს საჭირო ინფორმაციას, მაშინ მონაცემთა კონვერტაციის გეგმა აღწერს მონაცემთა შერჩევის შესაბამის სტრატეგიას, რომელიც შემდგომში სისტემას ამზადებს მონაცემთა ტრანსფორმაციისა და ინტეგრაციისათვის. (ნახ.1)



ნახ.1 მონაცემთა კონვერტაციის პროცესი

მონაცემთა გადანაცვლების საერთო შუალედური სტრატეგიის შერჩევისას გათვალისწინებული უნდა იყოს კომპიუტერის რესურსები, საკონვერტაციო ჯგუფის პროფესიონალიზმი და საწყის მონაცემთა მოცულობა.

მონაცემთა კონვერტაციის გეგმაში გათვალისწინებული უნდა იყოს მონაცემთა საცავის სტრუქტურა და მონაცემთა ბაზის მთლიანი სქემა. ჩვეულებრივად პროცესი ასეთია: მონაცემები თავის საწყისი სისტემის სქემიდან გადაინაცვლებენ შუალედურ სქემაში. შუალედური სქემა არის სქემა, რომელიც ჩატვირთავს მონაცემებს საწყისი სისტემებიდან. შუალედური სქემა სიზუსტით არ შეესაბამება საწყის ან მთლიან სქემებს. იგი როგორც წესი არის შუალედური და შეიცავს დამატებით ველს, რომელიც შედგება ციფრისგან ან პროცენტისგან და გამოიყენება გამოსაანგარიშებლად. ან ეს ველი შესაძლოა იყოს გასაღებური ველი რომლის დახმარებითაც შესაძლებელია ცხრილების დათვალიერება, საჭირო ტექნიკური პირობების გაუმჯობესება, მონაცემთა გასუფთავება და ტრანსფორმაცია. საბოლოოდ მონაცემთა კონვერტაციის გეგმა უნდა ითვალისწინებდეს მონაცემთა ნაკადის ხასიათს კონვერტაციის მთელ გზაზე, სადც ყურადღება უნდა მიექცეს შემდეგ საკითხებს:

- საწყისი სისტემიდან როგორ ხდება მონაცემთა მიგრაცია შუალედურ სისტემაში;
- ჯგუფი როგორ ახორციელებს ტექნიკურ პირობებს: მონაცემთა გასუფთავება, ტრანსფორმაცია და ინტეგრაცია შუალედურ სისტემაში;

- ჯგუფი როგორ უზრუნველყოფს მონაცემთა საცავისათვის პირველად და შიგა გასაღებს;
- მონაცემები როგორ მიგრირდება შუალედური სისტემიდან საცავის სერვერზე;
- როგორ ინახება კონვერტირებული მონაცემები და როგორ ხდება მათი ექსპორტი რეპოზიტორებში;
- ჯგუფი როგორ ჩატვირთავს და აინდექსირებს მონაცემებს საცავის სერვერზე მონაცემებს;
- ჯგუფი როგორ უზრუნველყოფს დამატებით მონაცემთა წყაროს, რომლის ყოფნაც შესაძლოა აუცილებელი გახდეს კონვერტაციის პროცესის ბოლო სტადიაზე;
- ჯგუფი როგორ უზრუნველყოფს კონვერტაციის პროცესი მონაცემთა ხარისხს.

გეგმის შედგენის შემდგომ უნდა განისაზღვროს მონაცემთა ატრიბუტები, თუ როგორი სახის მონაცემები მანიპულირება და როგორი სახის ლოგიკა გამოიყენება მონაცემთა მიგრაციისათვის.

მონაცემთა მიგრაციის ინსტრუმენტის გამოყენებით შესაძლებელია დავასაბუთოთ კონვერტაციის სპეციფიკა და წარმოვადგინოთ ანგარიში მომხმარებლის მოთხოვნის შესაბამისად. მონაცემთა მიგრაციის ინსტრუმენტის გამოყენება, მეტამონაცემთა კონვერტაციის შენახვის პროცესს ცენტრალურ დონეზე ხდის ავტომატიზებულს, საიდანაც რეპოზიტორში დროის ნებისმიერი შუალედში მეტამონაცემები ექსპორტირდება. იმ შემთხვევაში თუ კონვერტაციის პროცედურის გენერირებას ვახდენთ ხელით, ასეთივე წესით ვახდენთ მეტამონაცემთა კონვერტირების გენერაციასაც.

იმისდამიუხედავად კონვერტაციის სპეციფიკა პროგრამულ კოდში ჩატარებულია ხელით თუ ავტომატურად ის უნდა ემსახურებოდეს მონმხმარებლისთვის მონაცემთა მიწოდების სიღრმისეული გაგებასა და ლოგიკას. პროგრამული კოდი შედგება ექვსი პროცედურის ტიპისგან, რომელიც ასრულებს შემდეგ ფუნქციებს:

- საწყისი სისტემიდან მონაცემთა გადატანა შუალედურ სქემაში;
- მონაცემთა ჩატვირთისათვის შუალედური სქემის კონვერტაცია;
- ჩასატვირთი მონაცემების აგრეგაცია;
- ჩასატვირთი მონაცემების მიგრაცია შუალედური დონიდან მონაცემთა საცავის სერვერზე;
- დიდი მოცულობის მონაცემთა ბაზებიდან საცავში მონაცემთა ჩატვირთვა;
- მონაცემთა სისწორის შემოწმება.

ზოგიერთი პროგრამული კოდი რომელიც უზრუნველყოფს საწყისი სისტემიდან მონაცემთა გადატანას შუალედურ სქემაში გამოირჩევა უნიკალურობით თავისი მეთოდით პლატფორმითა და საწყისი სისტემის ენით. ამას გარდა ეს კოდი საწყისი მონაცემებიდან ამზადებს, წმენდს, ახდენს

მონაცემთა ტრანსფორმაციასა და ინტეგრირებას საცავში. ამ კოდის მუშაობას დიდი მნიშვნელობა აქვს სცავისათვის, იგი ამცირებს საცავში მიგრირებულ მონაცემთა მოცულობას, რაც ამავე დროს ეკონომიურად იყენებს კომპიუტერის და ქსელის რესურსებს.

საწყისი მონაცემები, როგორც კი თავს მოიყრის შუალედურ დონეში, დეგბა კონვერტაციის პროცედური შესრულების დრო, რომელსაც შეუძლია გაასუფთავოს მონაცემები. მონაცემთა გასუფთავება სპეციალური პროგრამის გამოყენებით უზრუნველყოფს მონაცემთა მთლიანობის დაცვას, გასწორებას და მათ სიზუსტეს ფართო გამოყენებისათვის. ასევე შესაძლებელია გამოვიყენოთ პროგრამები, რომელსაც შეაქვს შესწორებები და აუმჯობესებს ჩანაწერის ფორმატს.

მონაცემთა „გასუფთავების“ ძირითად კომპონენტებია:

- მონაცემთა კვლევა, მონაცემთა განმსაზღვრელი ხარისხი, მოდელის ხარისხი და ველების რაოდენობა;
- მონაცემთა სინტაქსის შემოწმება, შინაარსი და ყოველი კომპონენტის დანიშნულების ადგილი ყოველი ველისათვის;
- მონაცემთა კორექცია, ამ დროს კონვერტაციის ჯგუფი უნდა თანამშრომლობდეს დამკვეთებთან და კორექცია მოხდეს ინფორმაციის წყაროს ფარგლები;
- ჩანაწერების შედგენა, არის შემთხვევა, როდესაც ორი ჩანაწერი წარმოადგენს ერთსა და იმავე ობიექტს. ასეთ შემთხვევაში გამოიყენება რთული პროგრამული ინსტრუმენტი, რომელიც ახდენს მონაცემთა იდენტიფიკაციას.

კონვერტაციის გეგმში და კორექციის პროცედურაში გათვალისწინებული უნდა იყოს დამკვეთის შეხედულება, რომელიც აუცილებლად შესაბამისობაში უნდა მოდიოდეს საწყის მონაცემებთან. ეფექტური კონვერტაციის სპეციფიკისათვის, საჭიროა რაც შეიძლება მეტი ინფორმაცია გვქონდეს საწყისი მონაცემების სტრუქტურასა და თითოეული ველის მნიშვნელობაზე. პირველი ნაბიჯის გადადგმა კონვერტაციის პროცესის ნებართვაზე იწყება მას შემდეგ, რაც საწყის მონაცემებს განვითავსებთ შუალედურ სქემაში. შუალედურ სქემაში ხდება მონაცემთა გაფილტვრა, ტრანსფორმაცია და ინტეგრაცია. იმ შემთხვევაში თუ თავს იჩენს არასაიმედო ჩანაწერი შეტყობინება ეგზავნება დამკვეთს და ერთობლივად ხდება შეთანხმება საწყის სისტემაში მონაცემის შესწორებისა ან კონვერტაციის პროცესის გამოყენების შესახებ. იმ შემთხვევაში თუ შევასწორებთ მონაცემებს კონვერტაციის პროცესის დახმარებით, აუცილებლად უნდა მიეთითოს შესწორება მეტა მონაცემთა კონვერტაციაში.

კონვერტაციის პროცესის განხორციელება სისტემაში არსებული საბაზო ინფორმაციული ტექნოლოგიების ბაზაზე, არცთუ რთული პროცესს წარმოადგენს და აქვს ფართო გამოყენება.

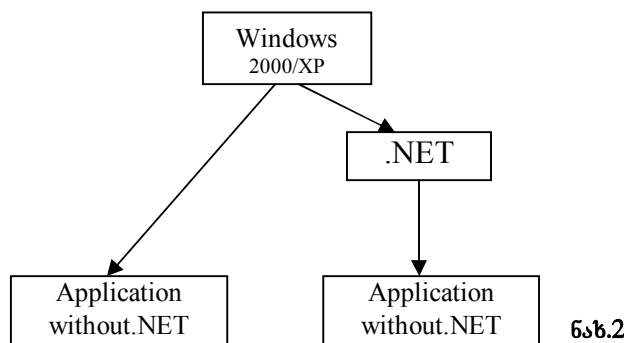
საბაზო ინფორმაციულ ტექნოლოგიას (სიბ) წარმოადგენს განაწილებული საინფორმაციო მართვის სისტემის ძირითადი პროგრამული პაკეტი. ესაა ოპერაციული სისტემა (მაგ., მაიკროსოფტის Windows 2000/XP და ლოტ-NET პლატფორმები) და მაღალი დონის პროგრამული

ენები, რომელთაც აქვს ობიექტ-ორიენტირებული მეთოდები და ინსტრუმენტული საშუალებანი (მაგ., C#.NET, Visual Basic.NET, JavaScript.NET, Visual C++.NET, Borland C++ Builder და სხვ.).

საინფორმაციო ტექნოლოგიები (სტი-ი, შეიძლება განვიხილოთ ფართო სპექტრით. კონკრეტული ინტეგრირებული საინფორმაციო მართვის სისტემის დაპროექტებისას ობიექტ-ორიენტირებული ანალიზის ეტაპზე დაზუსტდება აუცილებელი პროგრამული პაკეტებისა და მათ შორის საკომუნიკაციო ინტერფეისების შედგენილობა. მაგ., დოტ-NET საბაზო პლატფორმის შემთხვევაში ინფორმაციული ტექნოლოგიები შეიძლება იყოს C#.NET, Visual Basic.NET, MsOffice, ASP.NET (Web-აპლიკაციისთვის), ADO.NET. მონაცემთა ბაზების მართვის სისტემების სახით გამოიყენება MsSQL Server-2000, Oracle, InterBase და ა.შ. ყოველ ინფორმაციულ ტექნოლოგიას აქვს საბაზო ტექნოლოგიასთან კომუნიკაციის საინტერფეისო ენა [1].

მონაცემთა საცავეებისა და ბაზებიდან ინფორმაციის ამოსაღებად ფართოდაა მიღებული სტანდარტული სტრუქტურირებადი მოთხოვნების ენის ინსტრუმენტის გამოყენება (SQL). სხვადასხვა ინფორმაციული ტექნოლოგიების (მაგ., დაპროგრამების სხვადასხვა ენები, სხვადასხვა ბაზები) გამოყენებისას მონაცემების ან პროგრამული კოდების თავსებადობისათვის შეიძლება CORBA (Common Object Request Broker Architecture) ან DCOM (Distributed Component Object Model) ტექნოლოგიების გამოყენება [3].

დოტ-NET პლატფორმის შემთხვევაში მაკროსოფტმა ორიგინალურად გადაწყვიტა და გაამარტივა ეს პროცესი. მან შექმნა შუალედური გარდაქმნის ენა (IL- Intermediate Language). პროგრამები, რომელთა საწყისი კოდები დაწერილია, მაგალითად C#, VB. ან .NET-ის IL-ენაზე, კომპილირდება მანქანურ კოდში. .NET-ის ყველა ენა ფლობს CTS (Common Type System) მონაცემთა შესათანხმებელ საერთო ტიპებს, რათა ენების სტანდარტიზაცია იქნეს მიღწეული [19,22,23]. ამგვარად, ობიექტური კოდები IL-ენის საშუალებით ისე მიიღება, რომ მათში არაა დაფიქსირებული, თუ რომელ ენაზეა დაწერილი საწყისი კოდი. დოტ-NET პლატფორმა მდებარეობს Windows-ოპერაციულ სისტემასა და სავტომატიზაციო სისტემის აპლიკაციას შორის (იხ. ნახ.2).



**ლიტერატურა**

1. სუგულაძე გ., პეტრიაშვილი ლ., მონაცემთა საცავის აგების ტექნოლოგია ინტერნეტული ბიზნესის სისტემებისათვის. სტუ. თბილისი, 2005.
2. Albrecht J. Anfrageoptimierung in Data – Warehouse-Systemen auf Grundlage des multidimensionalen Datenmodells. Friedrich-Alexander-Universität, Erlangen-Nürnberg, 2003.
3. Albrecht J. Grundlagen der Anfrageverarbeitung in relationalen Datenbanksysteme.

**CONVERTING OF THE INFORMATION FOR DATA WAREHOUSE**

Petriashvili L., Kashibadze M., Okhanashvili M.

Georgian Technical University

**Summary**

Converting of the information for data warehouse of the big corporate systems is complex process. It is characterized by high labour input. The structure of process of converting in data warehouse of is offered. Problem of the dependence between processes of converting and decision-support are researched.

**КОНВЕРТАЦИЯ ИНФОРМАЦИИ ДЛЯ ХРАНИЛИЩА ДАННЫХ**

Петриашвили Л., Кашибадзе М., Оханашвили М.

Грузинский Технический Университет

**Резюме**

Конвертация информации для хранилища данных больших корпоративных систем является сложным процессом и характеризуется высокой трудоемкостью. Предлагается структура процесса конвертации в хранилище данных и исследуются вопросы связи этого процесса на временные характеристики процесса принятия решений.