

DUOMENŲ MIGRAVIMO IŠ OLTP Į OLAP DUOMENŲ BAZĘ GREITAVEIKOS TYRIMAS

Robertas Jankauskas, Dmitrij Šešok

Vilniaus Gedimino technikos universitetas

E. p.: robertas.jankauskas@vgtu.lt; dmitrij.sesok@vgtu.lt

Įvadas

Organizacijoms dažnai tenka saugoti reikalingus duomenis, kuriais naudojantis būtų galima priimti tam tikrus sprendimus. Tam yra naudojamos specifinės duomenų bazės, vadinamos duomenų saugyklomis. Duomenų saugyklos verslo analitikams teikia galimybių šiuos duomenis organizuoti, analizuoti ir panaudoti sprendimams priimti. Duomenys konkrečioje organizacijoje yra surenkami iš skirtingų duomenų šaltinių, vėliau pertvarkomi, koreguojami ir tik tada pradamas duomenų įkėlimo į saugyklą procesas. Naudojantis šiais duomenimis sukuriama ataskaitos galutiniams vartotojams. Taip galima atlikti duomenų analizę be jokių techninių žinių. Jei šie duomenys užima daug vietos atmintyje, įkelti juos į duomenų saugyklą iš skirtingų šaltinių užtrunka daug laiko (ypač tais atvejais, kai duomenų saugykla nėra tinkamai suprojektuota).

Tiek duomenų saugyklose, tiek ir kitose duomenų bazėse pasitelkiamas duomenų išgavimo, pertvarkymo ir įkėlimo (ETL) procesas, kurio metu iš skirtingų duomenų šaltinių išgaunami duomenys, jie išvalomi, filtruojami, tikrinami, susiejami, pritaikomos verslo taisyklės ir įkeliami į duomenų saugyklą. Šis procesas yra sudėtingas, reikalauja nemažai laiko ir pastangų, todėl organizacijoms svarbu nuspręsti, kokių būdu visa tai bus atliekama, pasirinkti įrankius, kurie leistų duomenis įkelti per kuo trumpesnę laiką.

Tikslas. Šio straipsnio tikslas yra ištirti duomenų migravimo greitaveiką iš OLTP bazės į duomenų

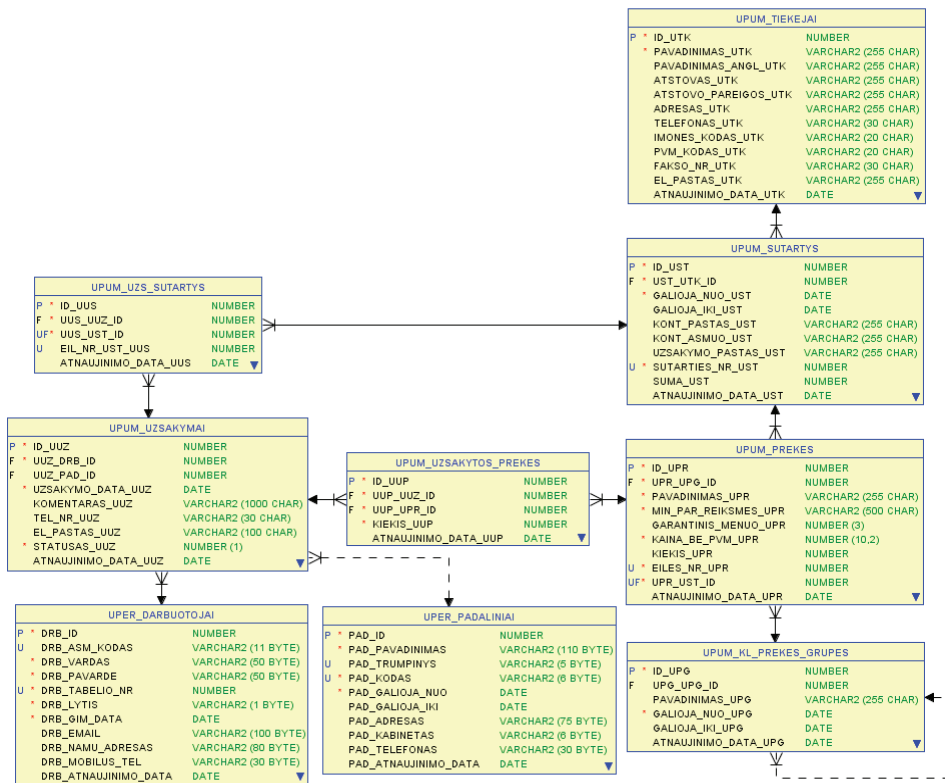
saugyklą ir pateikti rekomendacijų, padėsiančių pasirinkti optimalų ETL procesą.

Uždaviniai:

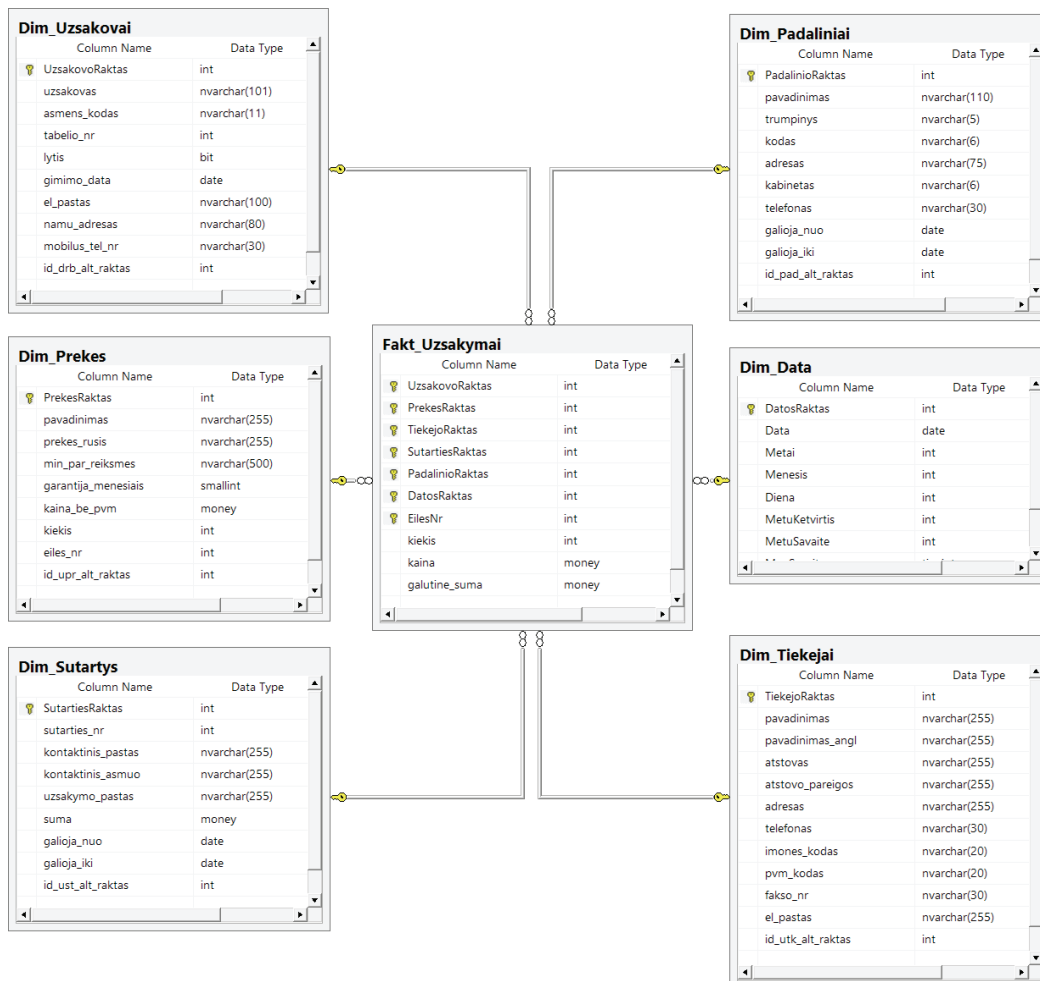
1. Išanalizuoti duomenų įkėlimo į duomenų saugyklą metodologijas ir technologijas.
2. Palyginti duomenų įkėlimo spartą taikant skirtingas metodologijas ir technologijas.
3. Suformuluoti išvadas ir rekomendacijas, padėsiančias pasirinkti optimalų ETL procesą.

Tyrimo metodai

Tyrimo metu analizuojama literatūra, dokumentacijos, elektroniniai leidiniai nagrinėta tema. Duomenims į duomenų saugyklą įkelti panaudotos dvi duomenų bazės: „Oracle 12c Enterprise Edition“ DBVS, iš kurios gaunami esami duomenys, ir „Microsoft SQL Server 2014 Enterprise Edition“. Šioje bazėje sukurta duomenų saugykla, į kurią įkeliami duomenys, ETL procesui realizuoti pasitelktas „Microsoft SQL Server Integration Services“ paketas. „Oracle“ duomenų bazėje sukurta 9 lentelės, kuriose saugoma 100 000 įrašų, „Microsoft SQL Server“ duomenų bazėje sukurta 6 dimensijų ir 1 fakto lentelės, į kurias įkeliami duomenys iš „Oracle“ duomenų bazės. 1 pav. pavaizduota su „Oracle“ DBVS sukurta reliacinės duomenų bazės schema, 2 pav. – su „Microsoft SQL Server“ DBVS sukurta duomenų saugyklos schema.



1 pav. Reliacinė užsakymų duomenų bazės schema



2 pav. Užsakymų duomenų saugyklos schema

Tiriamosios problemos aktualumas

Tiriamoji problema yra aktuali, šį klausimą tiria kiti mokslininkai. Pavyzdžiui, Tu Hitao ir Zhu Lanyuan'as nagrinėja ETL proceso vykdymą skirtingais duomenų įkėlimo būdais, apžvelgia galimus klaidų registravimo būdus, jei dėl tam tikrų priežasčių duomenų įkelti nepavyksta [1]. Alkis Simitsis, Kevinas Castelannos'as ir kt. aptaria ETL proceso vykdymo seką, kuri būtų optimali tiek vykdymo, tiek proceso patikimumo, tiek atkūrimo metu [8].

1. ETL procesas

ETL procesas susideda iš trijų etapų: duomenų išgavimas, pertvarkymas ir įkėlimas. Duomenų išgavimo metu iš skirtingų šaltinių gaunami ir surenkami duomenys. Vykstant pertvarkymui atliekamos ne tik pačios paprasčiausios duomenų transformacijos, bet ir pasitelkiamos sudėtingiausios duomenų pertvarkymo technikos. Šio proceso metu vykdomos taisyklės ar funkcijos, leidžiančios konvertuoti duomenis iš skirtingų duomenų šaltinių į standartinį formatą, tinkantį duomenų saugykloje ar kitoje duomenų bazėje. Šiame etape įrašai konvertuojami, tikrinami, atmunami, rūšiuojami, filtruojami, pasikartojantys įrašai šalinami pagal atitinkamas verslo taisykles. Įkėlimo metu išgauti ir pertvarkyti duomenys yra įkeliami į duomenų saugyklą ar kitą duomenų bazę. Duomenų įkėlimo procesas dažniausiai apima kiekvieno įrašo įterpimą kaip naują eilutę duomenų saugyklos lentelėje naudojant SQL įterpimo komandą ar ETL įrankį.

Duomenų saugyklose duomenys yra įkeliami į dimensijų ir faktų lenteles. Pirmiausia ETL procesas atliekamas dimensijų lentelėse, nes faktų lentelėse yra saugomos nuorodos (raktai) į dimensijų lentelių įrašus. Toliau iš dimensijų lentelių turi būti generuojami nauji raktai. Dažniausiai tai būna pirminis šių lentelių raktas, tačiau gali būti naudojamas ir kitas unikalus raktas. Įkėlus duomenis į skirtingas dimensijas, jie iškart keliami į faktų lenteles. Bendras ETL į duomenų saugyklą laikas skaičiuojamas taip:

$$T_i = \sum_{Dim_E} T(i, j)_{Dim_E} + T(i, j)_{Dim_{TL}} + \sum T(i, j)_{Fakt_E} + T(i, j)_{Fakt_{TL}} + T(i)_{kitos\ operacijos} \quad (1)$$

kur $T(i, j)_{Dim_E}$ yra i -osios iteracijos duomenų, reikalingų įkelti į j -ąją dimensijų lentelę, išgavimo laikas, $T(i, j)_{Dim_{TL}}$ yra i -osios iteracijos duomenų pertvarkymo ir įkėlimo laikas į j -ąją dimensijų lentelę. Atitinkamai $T(i, j)_{Fakt_E}$ ir $T(i, j)_{Fakt_{TL}}$ yra i -osios iteracijos duomenų išgavimo, pertvarkymo ir įkėlimo į j -ąją faktų lentelę laikas. $T(i)_{kitos\ operacijos}$ – i -osios operacijos laikas, skirtas kitoms operacijoms, susijusioms su ETL procesu, vykdyti [1].

1.1. Duomenų įkėlimo metodai

ETL procese išskiriami du duomenų įkėlimo metodai:

- Visiškas duomenų įkėlimo procesas.
- Pavienis duomenų įkėlimo procesas [1].

Visišku ETL proceso metu visi esami duomenys įkeliami į tarpinę duomenų bazę. Iš šios bazės duomenys įrašomi, pertvarkomi ir įkeliami į duomenų saugyklą. Dažniausiai šis metodas yra panaudojamas pirmą kartą įkėlus duomenis į duomenų saugyklą, tačiau gali būti pasitelkiamas viso ETL proceso metu. Tokiu atveju iš skirtingų duomenų šaltinių atrenkami visi (net ir neatnaujinti) duomenys, jie perkeliama į tarpinį sluoksnį, jame pertvarkomi ir įkeliami į duomenų saugyklą.

Pavienis duomenų įkėlimo procesas atliekamas tik su tais duomenimis, kurie buvo keičiami tam tikru laikotarpiu. Tai yra operacija, apimanti senų duomenų atnaujinimą ir naujų įrašų įterpimą [1].

1.2. Sistemų technologijos ETL procesui realizuoti

Norint realizuoti ETL procesą, galima pasinaudoti sistemų teikiama technologijomis, kurios leistų vienokiu ar kitokiu būdu paspartinti duomenų įkėlimo procesą. Tokios technologijos yra įvairios: pakitusių duomenų fiksavimas, kai gaunami tik pakitę duomenys; laikinų lentelių, kurios saugomos atskiroje „tempdb“ „Microsoft SQL Server“ duomenų bazėje, panaudojimas [6]; atmintį taupančios lentelės (angl. *Memory-optimized tables*), saugomos operatyvinėje atmintyje [5]; vartotojo apibrėžtos lentelių tipo struktūros, kai užklausų rezultatai perduodami per procedūros parametras, apibrėžiantį lentelės struktūrą [8]; padalytos lentelės, kuriose duomenys įkeliami į smulkesnes vienetines lentelės dalis [25]; indeksuoti vaizdiniai, saugantys užklausos rezultatus [3]; duomenų suspaudimo metodika, leidžianti sumažinti saugomų objektų dydžius [2]; indeksų panaudojimo galimybės, sudarančios galimybę paspartinti vykdomas užklausas [4]; lygiagretus proceso vykdymas, kurio metu duomenys vienu metu įkeliami į skirtingas lenteles; masinis duomenų įkėlimas, kai apdorojamas skirtingas įrašų kiekis, ir mažinamas transakcijų registravimas [7].

2. Eksperimentiniai tyrimai

Tyrimo metu ETL procesas testuotas tiek įrašant, tiek įrašant ir atnaujinant duomenis. Iš viso duomenų šaltinyje buvo panaudota 100 000 įrašų paties įrašymo proceso metu, o duomenų įrašymo ir atnaujinimo proceso metu – 50 000 naujų ir 50 000 atnaujintų duomenų. Duomenys įkeliami ir pasitelkiant tarpines lenteles, kurios yra sukuriama „Microsoft SQL Server“ DBVS duomenims saugoti ir paruošti

įkelti į dimensijų bei faktų lenteles, ir lentelių nenaudojant – pertvarkymai atliekami „Microsoft SSIS“ pakete. Technologijos, taikomos sistemai realizuoti,

yra pateiktos 1 lentelėje. Grafikuose pavaizduotos tos technologijos ir metodologijos, kurias pasitelkus buvo gauta geriausių rezultatų.

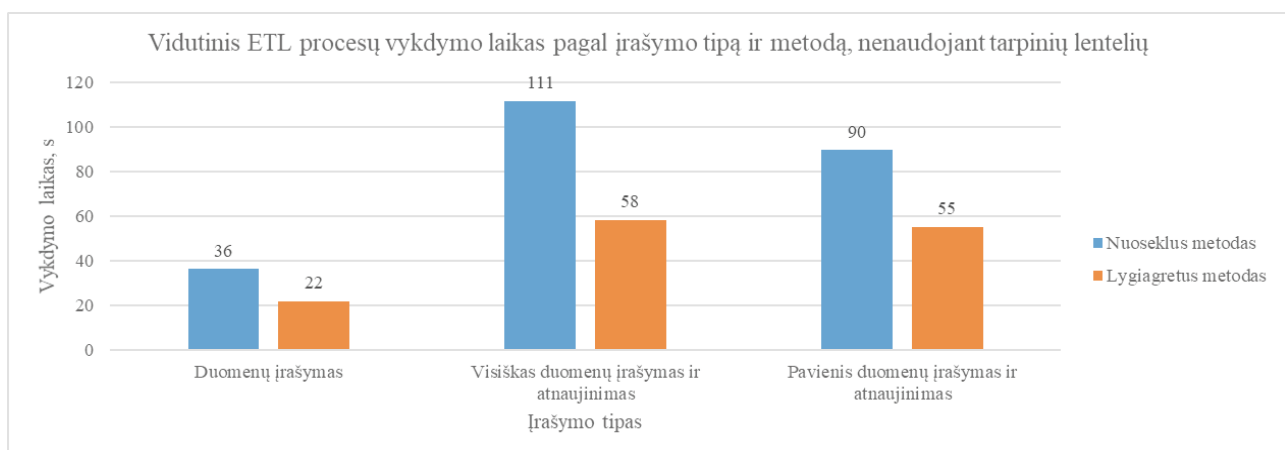
1 lentelė. *Tyrimui pasirinktos technologijos*

Kategorija	Programinė įranga
Operacinė sistema	„Windows 10 64bit“
Techninė įranga	„Intel(R) Core(TM) i5-2410M CPU @ 2.30Ghz“ procesorius, „4,00 Gb RAM“ operatyvinė atmintis
Duomenų šaltinio DBVS	„Oracle 12c Enterprise Edition 12.1.0“
Duomenų saugyklos kūrimo ir įgyvendinimo DBVS	„Microsoft SQL Server 2014 Enterprise Edition“
ETL įrankis	„Microsoft SQL Server Integration Services“ paketas
ETL vartotojo aplinka	„Microsoft Visual Studio 2015 Community Edition“
Duomenų pakitimų registravimo realizacijai pasirinkti įrankiai	Auditinis datos laukas, nurodantis, kada paskutinį kartą buvo keičiamas įrašas, ir trigeriai, kurie nustato paskutinio įrašo keitimo datą

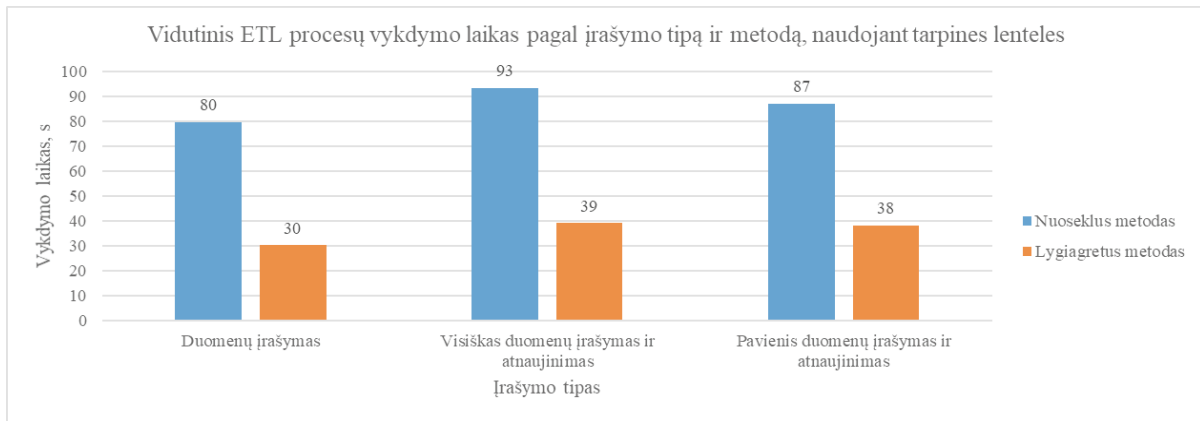
Vykstant duomenų migravimui, pasitelkus konkrečias metodologijas ir technologijas, iš viso buvo atliekami 5 bandymai esant toms pačioms sąlygoms (duomenų kiekis, struktūros ir pan.). Kiekvieno bandymo rezultatai skyrėsi nežymiai, tačiau kai kurių rezultatų skirtumas buvo kelios sekundės dėl galimai apkraunamo kompiuterio CPU.

3 pav. pateikti skirtingi duomenų išgavimo ir įrašymo į dimensijų bei faktų lenteles laikai pasitelkiant skirtingus duomenų įkėlimo būdus, kai nenaudojamos tarpinės lentelės. 4 pav. pateikti identiški procesai, juose tarpinės lentelės jau pritaikomos. Pagal pateiktus grafikus duomenys geriausiai buvo įrašomi vykdant ETL procesą lygiagrečiai: 3 pav. duomenys

įkeliami vidutiniškai 33 proc. sparčiau nei įkeliant duomenis nuosekliai, 4 pav. – 62 proc. sparčiau. Geriausias duomenų įrašymo laikas, jei lyginsime nuoseklų ir lygiagretų ETL procesą, nenaudojant tarpinių lentelių yra 19,671 sek., naudojant tarpines lenteles – 26,516 sek. Siekiant palyginti duomenų įrašymo ir atnaujinimo procesą duomenys buvo įkeliami nežymiai sparčiau naudojant pavienį duomenų įkėlimo būdą: 3 pav. pateikti duomenys buvo įkeliami vidutiniškai 5 proc. greičiau nei naudojant visišką duomenų įkėlimo metodą, 4 pav. – 6 proc. sparčiau. Geriausias duomenų įkėlimo laikas atitinkamai nenaudojant tarpinių lentelių ir jas pasitelkiant – 50,078 sek. ir 34,578 sek.



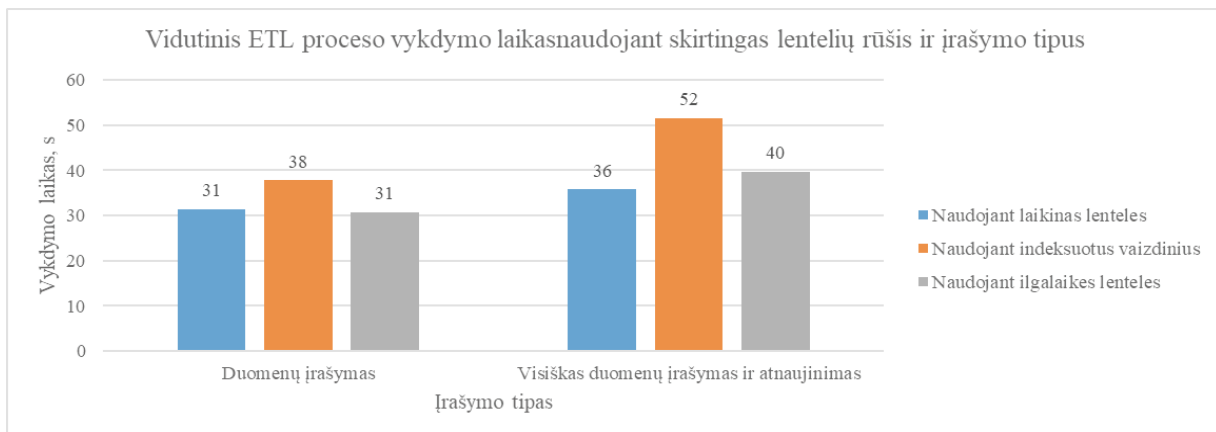
3 pav. Nuoseklus ir lygiagretus ETL procesų rezultatai nenaudojant tarpinių lentelių



4 pav. Nuoseklus ir lygiagretus ETL procesų rezultatai naudojant tarpines lenteles

5 pav. pateikti duomenų įkėlimo laikai naudojant laikinas, ilgalaikes lenteles ir indeksuotus vaizdinius. Šiose lentelėse buvo naudojamas vienas klasterinis indeksas lauke, kuris unikaliai identifikuoja įrašą lentelėje, taip pat neklasteriniai indeksai išorinio rakto apribojimus turinčiuose laukuose. Pagrindiniai skirtumai tarp klasterinio ir neklasterinio indeksų yra tokie: kadangi indeksai yra organizuojami B-medžio pavidalu, klasterinio indekso žemiausiame medžio lygmenyje yra saugomi surūšiuoti duomenys, o neklasteriniame indekse duomenys nėra surūšiuoti. Že-

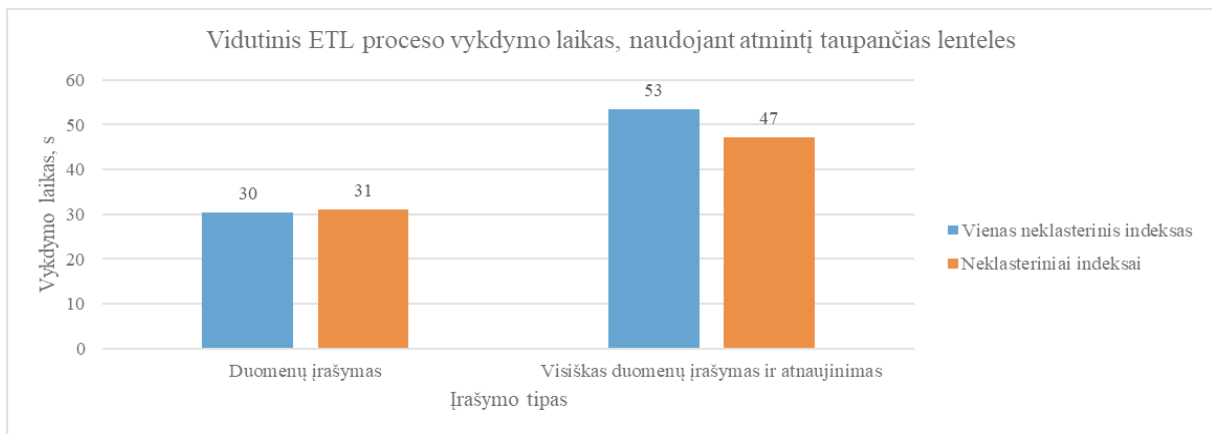
miausiame medžio lygmenyje saugomi ne duomenys, o nuorodos į šiuos duomenis. Pagal atliktus tyrimus duomenų įrašymo procesas buvo vykdomas sparčiausiai naudojant ilgalaikes lenteles: duomenys buvo įkeliami vidutiniškai per 30,716 sek., geriausias duomenų įkėlimo laikas – 26,516 sek. Duomenų įrašymo ir atnaujinimo procesas buvo vykdomas sparčiausiai naudojant laikinas lenteles: duomenys buvo įkeliami vidutiniškai per 35,750 sek., geriausias duomenų įkėlimo laikas – 32,688 sek.



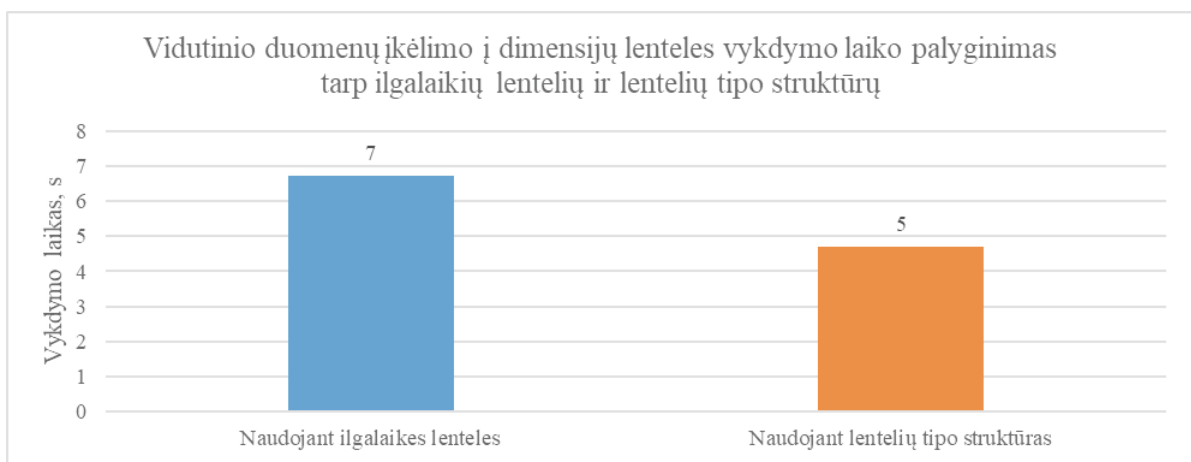
5 pav. ETL proceso rezultatai naudojant skirtingas lentelių rūšis

6 pav. pateikti duomenų įkėlimo rezultatai naudojant atmintį taupančias lenteles. Šiam procesui pasitelkiamos dvi indeksų kombinacijos: vienas neklasterinis indeksas tuose laukuose, kurie apibrėžia išorinio rakto apribojimus, ir neklasteriniai indeksai atskirai kiekviename iš šių laukų. Naudojant neklasterinius indeksus duomenų įrašymo ir duomenų įrašymo bei atnaujinimo procesai buvo vykdomi sparčiausiai: pirmas procesas vidutiniškai truko 31,037 sek., antras procesas – 47,188 sek. Geriausias užfiksuotas duomenų įrašymo laikas – 24,328 sek., duomenų įrašymo ir atnaujinimo – 42,766 sek. Tokius skirtumus tarp geriausio ir blogiausio rezultatų pateiktuose grafikuose lemia operatyviosios atminties panaudojimas: pasitelkiant didesnę kiekį operatyviosios atminties apkraunamas ir duomenų įkėlimo procesas į šias lente-

les. Atliekant eksperimentus pastebėta, kad naudojant lentelių tipo struktūras duomenys į dimensijų lenteles buvo įrašomi sparčiau nei naudojant ilgalaikes lenteles, tačiau tiek įrašant, tiek atnaujinant duomenis jie buvo įkeliami sparčiau, jei naudotos ilgalaikės lentelės, o ne lentelių tipo struktūros. 7 pav. pateiktas grafikas, kuriame pavaizduoti duomenų įrašymo ir dimensijų lenteles skirtumai naudojant ilgalaikes lenteles ir lentelių tipo struktūras. Atliekant visus penkis bandymus nustatyta, kad duomenys pasitelkiant lentelių tipo struktūras yra įkeliami sparčiau negu naudojant ilgalaikes lenteles – vidutiniškai 40 proc. greičiau. Geriausias duomenų įkėlimo laikas į dimensijų lenteles naudojant lentelių tipo struktūras – 4,578 sek., naudojant ilgalaikes lenteles – 6,578 sek.



6 pav. ETL proceso rezultatai naudojant atmintį taupančias lenteles



7 pav. Duomenų įrašymo į dimensijų lenteles rezultatai

8 pav. pateikti skirtingi duomenų įkėlimo į padalytą ir nepadalytą užsakymų faktų lentelę laikai. Iš grafiko galima spręsti, kad, jei užsakymų faktų lentelė yra padalyta, duomenys visais atvejais yra įkelia-

mi sparčiau nei įkeliant į nepadalytą lentelę – vidutiniškai 21 proc. greičiau negu į nepadalytą lentelę. Geriausias duomenų įkėlimo laikas į padalytą faktų lentelę – 11,894 sek., į nepadalytą – 14,852 sek.



8 pav. Duomenų įkėlimo rezultatai į faktų lentelę naudojant padalijimo metodologiją

Bendri jau pateiktų diagramų rezultatai su minimaliomis ir maksimaliomis bandymų metu gautomis

vykdymo laikų reikšmėmis yra aprašyti 2 lentelėje.

2 lentelė. *Tyrimų rezultatų apibendrinimas*

Pav. nr.	Įrašymo tipas	Naudojamas metodas	Vykdymo laikas, s		
			Minimali reikšmė	Maksimali reikšmė	Vidurkis
3	Duomenų įrašymas	Nuoseklus	33,487	39,082	36,217
		Lygiagretus	19,671	24,591	21,687
	Visiškas duomenų įrašymas ir atnaujinimas	Nuoseklus	99,698	119,540	111,440
		Lygiagretus	50,594	66,841	58,094
	Pavienis duomenų įrašymas ir atnaujinimas	Nuoseklus	85,254	92,684	89,593
		Lygiagretus	50,078	62,326	55,189
4	Duomenų įrašymas	Nuoseklus	75,236	88,659	79,655
		Lygiagretus	26,516	33,591	30,269
	Visiškas duomenų įrašymas ir atnaujinimas	Nuoseklus	91,136	94,623	93,272
		Lygiagretus	35,688	40,023	39,133
	Pavienis duomenų įrašymas ir atnaujinimas	Nuoseklus	84,358	90,757	87,269
		Lygiagretus	34,578	39,967	38,132
5	Duomenų įrašymas	Naudojant laikinas lenteles	30,421	32,568	31,337
		Naudojant indeksuotus vaizdinius	31,784	44,542	37,749
		Naudojant ilgalaikes lenteles	26,516	33,645	30,716
	Visiškas duomenų įrašymas ir atnaujinimas	Naudojant laikinas lenteles	32,688	37,984	35,750
		Naudojant indeksuotus vaizdinius	49,631	56,231	51,624
		Naudojant ilgalaikes lenteles	34,784	41,857	39,693
6	Duomenų įrašymas	Vienas neklasterinis indeksas	29,487	32,641	30,369
		Neklasteriniai indeksai	24,328	43,912	31,037
	Visiškas duomenų įrašymas ir atnaujinimas	Vienas neklasterinis indeksas	43,167	61,094	53,364
		Neklasteriniai indeksai	42,766	54,847	47,188
7	Duomenų įrašymas į dimensijų lenteles	Naudojant ilgalaikes lenteles	6,578	7,102	6,725
		Naudojant lentelių tipo struktūras	4,578	4,956	4,704
8	Duomenų įrašymas į faktų lentelę	Įrašymas į padalytą lentelę	11,894	13,172	12,394
		Įrašymas į nepadalytą lentelę	14,852	15,157	14,997

Išvados

1. Išanalizavus ETL metodologijas, technologijas galima teigti, kad nėra vieno metodo, leidžiančio panaudoti technologijas ETL procese – dažniausiai tenka atlikti eksperimentinius tyrimus, kuriais remiantis galima nuspręsti, kurias metodologijas ar technologijas pasitelkti.
2. Atlikti ETL proceso tyrimai naudojant skirtingas metodologijas bei technologijas, nustatyta, kad sparčiausiai duomenys buvo įrašomi nenaudojant tarpinių lentelių vykdant duomenų įrašymo procesą lygiagrečiai: duomenys įrašomi mažiausiai 33 proc. sparčiau nei taikant kitas technologijas. Pasitelkus tarpines lenteles nustatyta, kad duomenys sparčiausiai buvo įkeliami naudojant klasterinį indeksą unikaliame lauke, kuris identifikuoja įrašą, ir neklasterinius indeksus laukuose, kurie nustato išorinio raktų apribojimus tarp lentelių: duomenys buvo įkeliami mažiausiai 0,65 proc. sparčiau negu taikant kitas technologijas. Geriausias rezultatas iš visų bandymų buvo gautas pasitelkiant atmintį taupančias lenteles. Duomenų įrašymo ir atnaujinimo procesas buvo vykdomas sparčiausiai, kai buvo pasitelktos laikinosios len-

telės. Tokiu atveju duomenys įkeliami mažiausiai 10 proc. sparčiau nei taikant kitas technologijas.

Literatūra

1. Hitao Tu; Lanjuan Zhu. *An optimized ETL fault-tolerant algorithm in data warehouses*. 2013, Information Science and Technology (ICIST). ISBN 978-1-4673-2764-0 Prieiga per internetą: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6747594>.
2. *Microsoft. Data Compression*. Prieiga per internetą: <https://docs.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression>.
3. *Microsoft. Designing Indexed Views*. Prieiga per internetą: [https://technet.microsoft.com/en-us/library/ms187864\(v=sql.105\).aspx](https://technet.microsoft.com/en-us/library/ms187864(v=sql.105).aspx).
4. *Microsoft. Indexes* [interaktyvus]. 2016. Prieiga per internetą: <https://docs.microsoft.com/en-us/sql/relational-databases/indexes/indexes>.
5. *Microsoft. In-Memory OLTP (In-Memory Optimization)*. Prieiga per internetą: <https://docs.microsoft.com/en-us/sql/relational-databases/in-memory-oltp>.
6. *Microsoft. Special Table Types*. Prieiga per internetą: [https://technet.microsoft.com/en-us/library/ms186986\(v=sql.105\).aspx](https://technet.microsoft.com/en-us/library/ms186986(v=sql.105).aspx).

7. Microsoft. *The Data Loading Performance Guide*. Prieiga per internetą: [https://technet.microsoft.com/en-us/library/dd425070\(v=sql.100\).aspx](https://technet.microsoft.com/en-us/library/dd425070(v=sql.100).aspx).
8. Microsoft. *Use Table-Valued Parameters (Database Engine)*. Prieiga per internetą: <https://docs.microsoft.com/en-us/sql/relational-databases/tables/use-table-valued-parameters-database-engine>.
9. Simitsis, Alkis; Wilkinson, Kevin; Castellanos, Malu; Dayal, Umeshwar. *QoX-Driven ETL Design: Reducing the Cost of ETL Consulting Engagements*. 2009, p. 953–960. ISBN 978-1-60558-551-2. Prieiga per internetą: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.367.3287&rep=rep1&type=pdf>.

Summary

RESEARCH OF DATA MIGRATION FROM OLTP TO OLAP DATABASE LOADING SPEED

R. Jankauskas, D. Šešok

Data synchronization between several systems is one of the most frequent events when data analysis of acquired data is needed. The usage of the analysis could enable certain decisions to be made that result in conclusions about actions of an organisation. In this work data migration process between Oracle and Microsoft SQL Server database management systems is analysed using different data loading methodologies and technologies provided by systems.

Keywords: ETL, data warehouse, Oracle, Microsoft SQL Server, SSIS, DBMS.

Santrauka

DUOMENŲ MIGRAVIMO IŠ OLTP Į OLAP DUOMENŲ BAZĘ GREITAVEIKOS TYRIMAS

R. Jankauskas, D. Šešok

Duomenų sinchronizacija tarp kelių sistemų yra vienas iš dažniausiai pasitaikančių procesų. Iš esamų duomenų norima atlikti analizę, kuria naudojantis būtų galima priimti tam tikrus sprendimus, padėsiančius pateikti išvadas apie organizacijų veiklas. Šiame darbe analizuojamas duomenų migravimo procesas tarp „Oracle“ ir „Microsoft SQL Server“ duomenų bazių valdymo sistemų naudojantis skirtingomis duomenų įkėlimo metodologijomis ir sistemų teikiamomis technologijomis.

Prasminiai žodžiai: ETL, duomenų saugykla, „Oracle“, „Microsoft SQL Server“, SSIS, DBVS.

Įteikta 2017-06-09

Priimta 2017-08-29