

УДК 004.658.3

Мунзер Аль Абдо

ТЕХНОЛОГИЯ МОДЕЛИРОВАНИЯ ФРАГМЕНТОВ РЕЛЯЦИОННЫХ БАЗ ДАННЫХ

Аннотация. Проанализированы программные способы повышения производительности информационных систем и определены основные функции модели реляционной базы данных. Предложена математическая модель запроса к базе данных. Определены способы выявления наиболее часто используемых таблиц и полей для запросов различных типов. Предложена структура имитационной модели фрагмента базы данных и методика проведения исследований на модели.

Ключевые слова: база данных, запрос, модель запроса, производительность информационной системы, временной ряд запросов, структура реляционных базы данных

Munzer Al Abdo

RELATIONAL DATABASES FRAGMENTS MODELLING TECHNOLOGY

Abstract. The programmatic methods of increase of the productivity of the informative systems are analysed and the basic functions of model of relational database are certain. The mathematical model of inquiry offers database. The methods of exposure of the most often used tables and fields are certain for the queries of different types. The structure of simulation model of fragment of database and methodology of realization of researches offer on a model.

Keywords: query, model of query, productivity of the informative system, temporal row of queries, structure of DB, model of fragment of DB

Мунзер Аль Абдо

ТЕХНОЛОГІЯ МОДЕЛЮВАННЯ ФРАГМЕНТІВ РЕЛЯЦІЙНИХ БАЗ ДАНИХ

Анотація Проаналізовано програмні засоби підвищення продуктивності інформаційних систем і визначено основні функції моделі реляційної бази даних. Запропоновано математичну модель запиту до бази даних. Визначено способи виявлення найчастіше використовуваних таблиць і полів для запитів різних типів. Запропоновано структуру імітаційної моделі фрагмента бази даних і методіку проведення досліджень на моделі.

Ключові слова: база даних, запит, модель запиту, продуктивність інформаційної системи, часовий ряд запитів, структура бази даних

Постановка задачи

Существуют различные способы повышения производительности информационных систем (ИС), использующих реляционные базы данных (РБД). Условно их можно разделить на аппаратные и программные. Аппаратные часто связаны с большими материальными затратами, поэтому перед заменой оборудования полезно убедиться в том, что программные способы повышения производительности себя исчерпали. К программным способам повышения производительности можно отнести следующие:

- тиражирование данных [1];
- использование материализованных представлений [2, 3];
- реструктуризация РБД [4, 8];
- настройки индексной системы [5, 6];
- автоматизация настройки СУБД [7, 10].

Аналитической оценки эффективности указанных способов в конкретных условиях часто оказывается недостаточно для приня-

тия решения об их внедрении. В таких случаях учитывая цену ошибок нужно применять моделирование. Известные способы моделирования баз данных [9, 11] не позволяют моделировать отдельные фрагменты РБД, а также не дают рекомендаций о выборе этих фрагментов.

Определение задач моделирования

Поскольку процесс создания модели достаточно трудоёмкий, требует специальных знаний и приёмов, которые обычно не используются прикладными программистами или администраторами БД, предлагается общую модель РБД представить в виде двух частей (рис. 1).

Основные функции модели представляют собой набор инструментальных средств, которые по выбору исследователя конкретного решения могут использоваться в различных комбинациях. Специфические функции модели реализуются исследователем, дополняя или конкретизируя основные функции.

© Аль Абдо Мунзер, 2013

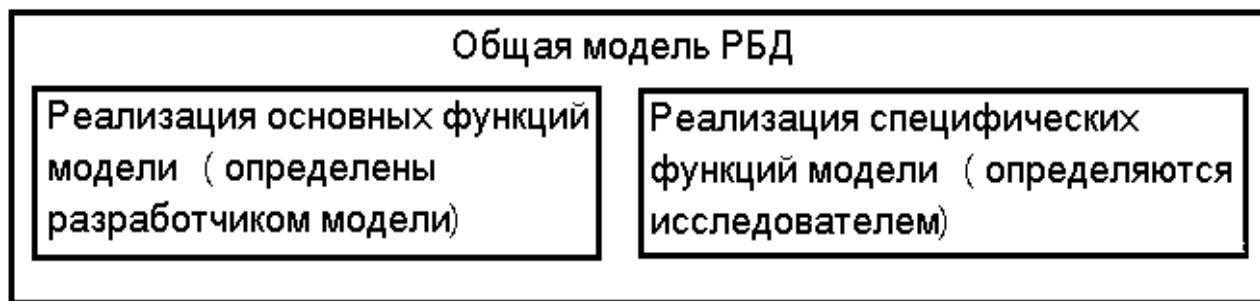


Рис. 1. Структура модели РБД

Задачей данной работы является выявление и реализация основных функций модели. В дальнейшем моделью будем называть эту часть общей модели РБД.

В результате анализа технологий внедрения перечисленных выше программных способов повышения производительности были установлены основные функции модели.

1. Получение временной последовательности всех запросов, поступающих в ИС в течение некоторого времени наблюдения (для всех способов повышения производительности).

2. Выделение из общей последовательности запросов – запросов определенного типа (для всех способов повышения производительности).

3. Определение наиболее часто используемых таблиц в запросах определенного типа (тиражирование данных, материализованные представления, реструктуризация, настройка индексов).

4. Выделение одной либо нескольких таблиц из БД и создание на их основе модели фрагмента БД (тиражирование данных, материализованные представления, реструктуризация, настройка индексов).

5. Создание резервных копий для таблиц РБД (тиражирование данных, материализованные представления, реструктуризация, настройка индексов).

6. Выделение запросов, использующих выделенные фрагменты (тиражирование данных, материализованные представления, реструктуризация, настройка индексов).

7. Многократный «прогон» тестовой последовательности запросов для фрагментов РБД (тиражирование данных, ма-

териализованные представления, реструктуризация, настройка индексов).

Математическая модель запроса

Функционирование РБД, а, следовательно, и её модели основано на выполнении запросов. Поэтому модель запроса – неотъемлемая часть модели РБД.

Пусть Q представляет временной ряд всех запросов, поступивших в ИС за время наблюдения t_0 с момента времени τ_s до момента времени τ_f . Каждый запрос q_i можно представить в виде

$$q = \langle tx, t_{st}, t_{fin}, T_q, T_y \rangle, \quad (1)$$

где t_x – текст запроса; t_{st} , t_{fin} – время начала и окончания выполнения запроса соответственно; $T_q = \{T_i\}, i = \overline{1, n_q}$ – множество таблиц, используемых в запросе q ; T_y – тип запроса.

Каждая таблица представляет собой кортеж

$$T_i = \langle NT_i, F_i \rangle,$$

где NT_i – наименование таблицы; $F_i = \{f_{ij}\}, j = \overline{1, n_{Fi}}$ – множество полей (колонок) таблицы.

Тип запроса T_y является элементом множества

$$T_y \in \{t_d, t_i, t_u, t_s, t_t\},$$

где t_d – запрос типа DELETE; t_i – запрос типа INSERT; t_u – запрос типа UPDATE; t_s – запрос типа SELECT; t_t – запрос типа изменения структуры БД – DROP TABLE, ALTER TABLE.

Использование в (1) t_{st} , t_{fin} позволяет говорить об уникальности каждого запро-

са и дает возможность ввести понятие множества для запросов, поступивших на вход ИС

$$Q = \{q_i\}, i = \overline{1, n}, \quad (2)$$

где n – количество запросов.

Определение наиболее часто используемых таблиц и полей

Существенного повышения производительности ИС можно добиться, если изменения коснуться наиболее интенсивно используемых её фрагментов. Для ИС, основанных на РБД, это отдельные таблицы или группы таблиц.

Предлагается ввести понятие множества таблиц БД

$$MT = \{ \langle NT_i, nt_i, F_i \rangle, i = \overline{1, n_T} \}, \quad (3)$$

где каждый элемент определяется: именем таблицы NT_i , количеством обращений к таблице nt_i во всех запросах за время наблюдения t_0 , множеством полей таблицы $F_i = \{ \langle Nf_{ij}, nf_{ij} \rangle, j = \overline{1, n_{F_i}} \}$, где Nf_{ij} – название поля; nf_{ij} – количество обращений к полю Nf_{ij} во всех запросах за время наблюдения t_0 .

В начале исследования все значения nt_i и nf_{ij} приравниваются нулю.

Введём операцию определения принадлежности некоторой таблицы NT_i запросу q

$$NT_i \in_T q,$$

результатом, которой будет $true$, если таблица NT_i используется в запросе q и $false$ в противном случае.

Введём операцию определения принадлежности некоторого поля f_{ij} таблице NT_i

$$f_{ij} \in_f NT_i,$$

результатом, которой будет $true$, если поле f_{ij} принадлежит таблице NT_i и $false$ в противном случае.

Специально для запросов обновляющих данные (INSERT, UPDATE, DELETE) введём операцию определения таблицы, в которой происходит обновление

$$NT_i \in_{new} q.$$

Определим функцию $fTF(q)$, подсчитывающую количество обращений к таблицам и полям в запросе q . Значение nt_i увеличивается на 1, если

$$NT_i \in_T q = true.$$

Значение nf_{ij} увеличивается на 1, если

$$NT_i \in_T q \wedge f_{ij} \in_f NT_i = true.$$

Сформируем на основе множества таблиц MT список таблиц Lmt , где элементы будут упорядочены по убыванию параметра nt_i

$$MT \Rightarrow_{sort} Lmt.$$

Определение наиболее часто используемых таблиц и полей в запросах определенного типа

Предложенный способ создания списка наиболее часто используемых таблиц не сложно использовать для запросов определенного типа. Например, если требуется найти наиболее часто используемые таблицы в запросах типа SELECT, то требуется создать на основе множества запросов Q (3) подмножество $Q_{sel} = \{q_i \mid q_i \in Q \wedge Ty_i = ts\}$.

Использование функции $fTF(q)$ для запросов $q \in Q_{sel}$ позволяет создать множество таблиц MT_{sel} и упорядоченный список $MT_{sel} \Rightarrow_{sort} Lmt_{sel}$.

Для получения информации о наиболее быстро «растущих» таблицах необходимо сформировать множество запросов типа

INSERT

$$Q_{ins} = \{q_i \mid q_i \in Q \wedge Ty_i = ti\}.$$

Затем, применив функцию $fTF(q)$ для запросов $q \in Q_{ins}$, получим множество таблиц MT_{ins} и упорядоченный список $MT_{ins} \Rightarrow_{sort} Lmt_{ins}$.

Для исследования наиболее ресурсоемких запросов типа SELECT можно получить список таблиц, которые используются в запросах, время выполнения которых превышает среднее время выполнения запросов такого типа. Для этого определяем среднее время выполнения запросов типа SELECT

$$t_a = \frac{1}{n_s} \sum_{i=1}^{n_s} (tf_i - ts_i),$$

где n_s – мощность множества $Qsel$; tf_i и ts_i – время окончания и начала выполнения запроса $q_i \in Qsel$.

Подмножество запросов, время выполнения которых превышает среднее время, определяется так:

$$Qsel_t = \{q_i \mid q_i \in Qsel \wedge (tf_i - ts_i) > t_a\}.$$

Применив функцию $fTF(q)$ для запросов $q \in Qsel_t$, получим множество таблиц $MTsel_t$ и упорядоченный список Lmt_t .

Построение моделей для фрагментов БД

Списки $LmtX$, где X указывает на способ построения списка, позволяют выделить одну или несколько таблиц, для которых необходимо провести некоторые исследования. Поскольку исследования предусматривают выполнение запросов, то, прежде всего, следует определить множество запросов, использующих выделенные таблицы.

Пусть на основании анализа $LmtX$ была выделена таблица $T0$. Определим множество запросов, использующих эту таблицу:

$$Q0 = \{q_i \mid T0 \in Tq_i\}.$$

Определим множество таблиц, которые используются в запросах из множества $Q0$

$$MT0 = \{T_{ij} \mid T_{ij} \in Tq_i \wedge q_i \in Q0\} i = 1, |Q0|.$$

Создадим для модели фрагмента режим тестирования максимально приближенный к режиму реальной эксплуатации. Для этого следует учесть, что результаты выполнения запросов из множества $Q0$ и, следовательно, время их выполнения, зависят не только от данных, находящихся в таблице $T0$, но и от данных других таблиц множества $MT0$. Поэтому определим ряд условий, которым должны соответствовать все запросы, обновляющие данные в таблицах, входящих в $MT0$.

Условие принадлежности к одному из типов, обновляющих данные – $C_1 = Ty_i \in \{t_i, t_u, t_d\}$.

Условие обновления данных в одной из таблиц модели – $C_2 = Tq_{ij} \in_{new} MT0$.

Условие использования в запросе только таблиц из множества $MT0$ – $C_3 = Tq_{ij} \in MT0$.

Создадим множество запросов, удовлетворяющих этим условиям:

$$Qu = \{q_i \mid C_1 \wedge C_2 \wedge C_3\}.$$

Тогда множество всех запросов, поступающих на вход модели

$$Qm = Q0 \cup Qu. \quad (4)$$

Исследователю предоставляется возможность создания списка таблиц по своему усмотрению, а также выделения из него подмножества таблиц для модели. На основании полученного подмножества таблиц формируется соответствующее подмножество запросов. В подмножество таблиц пользователь может вносить изменения, например, вводить или удалять индексы. Если изменения связаны с реструктуризацией, то исследователь может вносить изменения и в подмножество запросов, которые, однако, не должны отразиться на результатах их выполнения.

Поскольку при формировании множества Qu были приняты ограничения на используемые таблицы, процесс выполнения запросов в модели может несколько отличаться от их выполнения в исходной БД. Однако поскольку в запросах типа INSERT, UPDATE и DELETE редко используется более одной таблицы, то можно считать, что полученные определения для $MT0$ и Qm позволяют создать модель, адекватную фрагменту реальной РБД, не только по структуре и набору запросов, но и по используемым данным.

Процесс построения модели фрагмента БД иллюстрирует рис. 2.

Методика работы с моделью

Основной характеристикой, определяющей эффективность большинства решений, связанных с модернизацией БД,

является время выполнения запросов. При моделировании фрагментов БД рекомендуется использовать следующие абсолютные оценки.

1) Суммарное время выполнения запросов из выделенного для модели подмножества Qm , полученное в условиях исходной БД (данные из журнала транзакций),

$$S_0 = \sum_{i=1}^{n_m} (tf_i - ts_i), \quad (5)$$

где n_m мощность множества Qm .

2) Суммарное время выполнения запросов из выделенного для модели подмножества Qm , выполненных в тестовой БД (данные по результатам эксперимента), при условии, что подмножества таблиц и запросов не подвергались модернизации – S_{m0} .

3) Суммарное время выполнения запросов из выделенного для модели подмножества поданных в тестовую БД (данные по результатам эксперимента), при модернизации подмножества таблиц и запросов – S_m .

Полученное множество запросов Qm распределено по всему интервалу времени наблюдения t_0 . Во время испытания модели необходимо подавать запросы на её вход с определенным интервалом времени

$$\Delta q \geq \Delta q_{\max},$$

где $\Delta q_{\max} = tf - ts$ – наибольшее время выполнения запроса $q \in Q0$. Это будет гарантировать выполнение каждого предыдущего запроса к моменту поступления следующего.

На рис. 3 представлен алгоритм работы исследователя с моделью фрагмента БД.

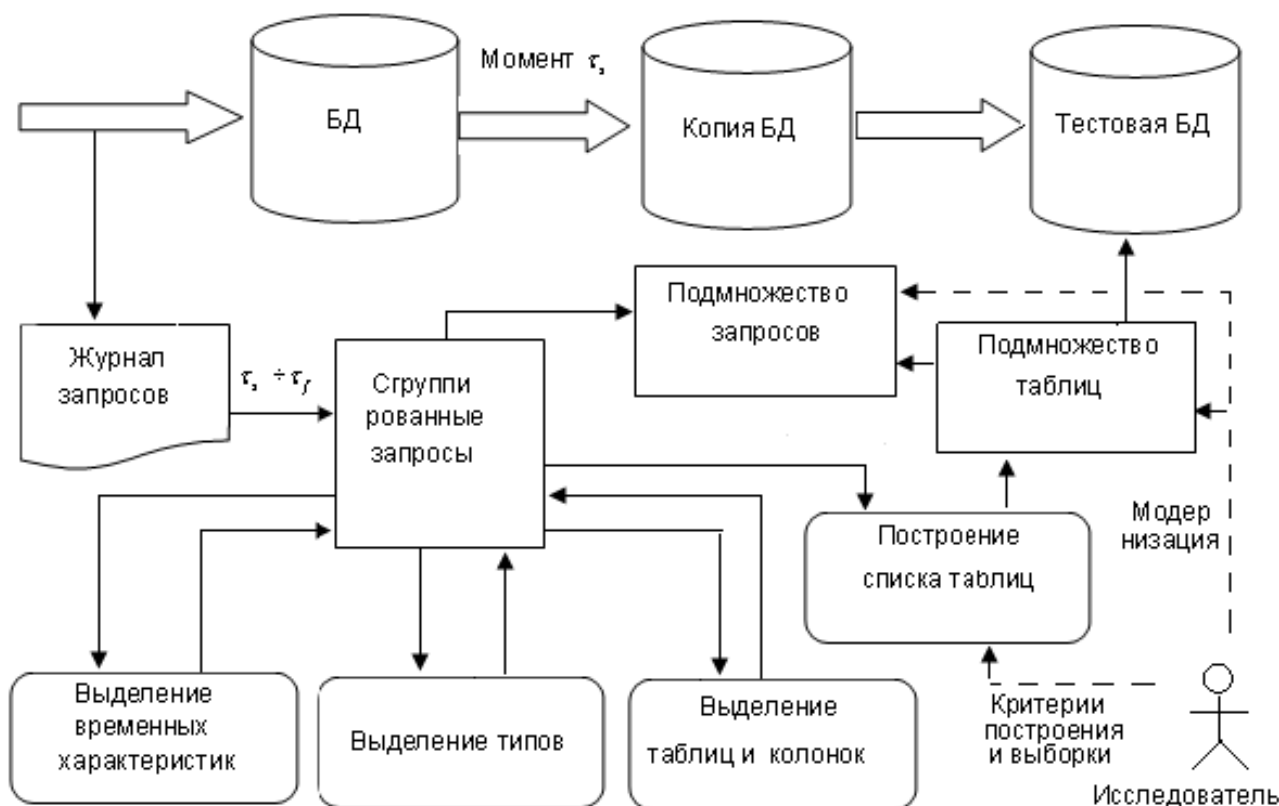


Рис. 2. Построение модели фрагмента БД

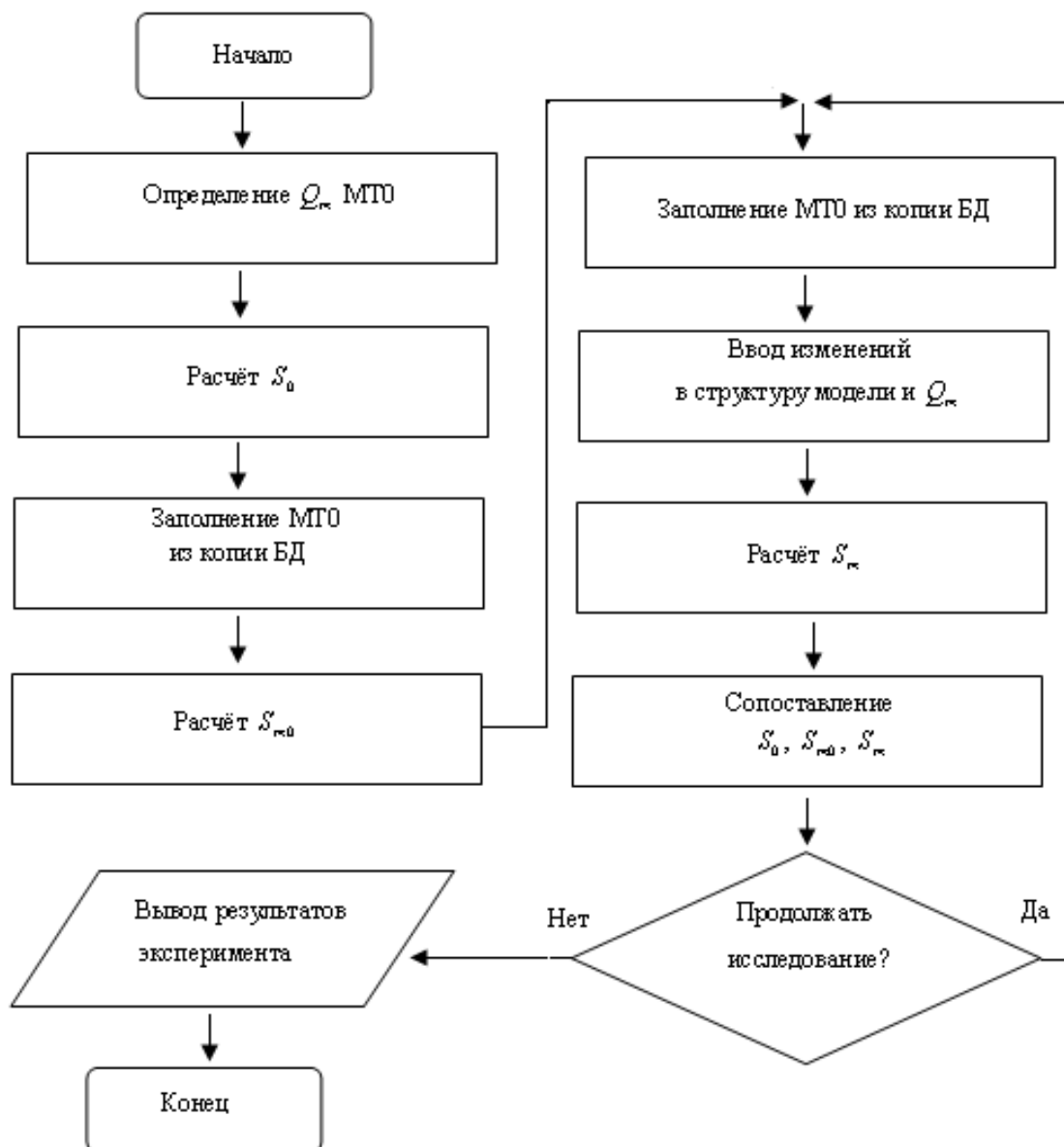


Рис. 3. Алгоритм работы исследователя с моделью

Выводы

Определены требования к моделированию фрагментов РБД для тестирования различных решений из области программных способов повышения производительности ИС. Предложена модель запроса, позволяющая формировать различные множества запросов, которые могут быть поданы на вход модели. Предложено решение по выделению фрагментов исходной БД в зависимости от поставленных задач по её модернизации. Разработана технология построения и испытания моделей для фрагментов РБД.

Список использованной литературы

1. Блажко А. А. Анализ потоков транзакций с целью перераспределения информационных ресурсов / А. А. Блажко, А. Б. Кунгурцев, Н. А. Медведь // Вестник Харьковского государственного политехнического университета. – Харьков : 1999. – Вып. 73. – С. 19 –23.
2. Кунгурцев А. Б. Анализ возможности применения материализованных представлений в ИС / А. Б. Кунгурцев, Куок Винь Нгуен Чан // Труды Одесского политехнического университета. – Одесса : Вып. 2(20). – 2003. – С. 102 – 106.

3. Возовиков Ю. Н. Имитационное моделирование материализованных представлений / Ю. Н. Возовиков, А. А. Мунзер // *Вісник Чернігівського державного технологічного університету*. Чернигов : – 2011. – Вып. 4(53). – С. 196 – 200.

4. Зиноватная С. Л. Анализ информационной системы для принятия решения о восходящей денормализации / С. Л. Зиноватная // *Математическое моделирование и информационные технологии : IV семинар*. – Одесса : 19–20 января 2005. – С.41 – 43.

5. Кунгурцев А. Анализ запросов к базе данных для определения полей-кандидатов на создание кластерного индекса /А. Кунгурцев, С. Зиноватная, С. Мусанна, А. Манчук // *Годишник на технічески університет*. – Варна. – 2010. – Том 1. – С. 127 – 130.

6. Zengfeng Huang, Ke Yi, and Yunhao Liu. Optimal Sampling Algorithms for Frequency Estimation in Distributed Data (2011). *Conference: IEEE INFOCOM-INFOCOM*, pp. 1997 – 2005, doi: 10.1109/INFOCOM.2011.5935005.

7. Левченко А. Ю. Методика тестирования производительности систем управления базами данных / А. Ю. Левченко // *Електромашинобудування та електрообладнання. Міжвідомчий науково-технічний збірник*. – Одеса : – 2011. – № 04 (80). – С. 201 – 204.

8. Hassan Artail, Hadi El Amine, and Fehmi Sakkal. SQL Query Space and Time Complexity Estimation for Multidimensional Queries, (2008), *International Journal of Intelligent Information and Database Systems*, Vol. 2, No. 4, pp. 460 – 480, doi: 10.1504/IJIDS.2008.021448.

9. Justus S., and Iyakutti K. An Empirical Validation of the Suite of Metrics for Object-relational Data Modelling, (2011), *International Journal of Intelligent Information and Database Systems*, Vol. 5, No. 1, pp. 49 – 80, doi: 10.1504/IJIDS.2011.037704.

10. Peter Aiken, Mark Gillenson, Xihui Zhang, and David Rafner. Data Management and Data Administration: Assessing 25 Years of Practice, (2011), *Journal of Database Management (JDM)*, Vol. 22, Iss. 3, pp. 24 – 45, doi: 10.4018/jdm.2011070102.

11. Keng Siau, Fiona F.H. Nah, and Qing Cao. A Meta-Analysis Comparing Relational and Semantic Models, (2011), *Journal of Database Management (JDM)*, Vol. 22, Iss. 4, pp. 57 – 72, doi: 10.4018/jdm.2011100103.

Получено 04.11.2013

References

1. Blazhko A.A., Kungurtsev A.B., and Medved N.A. Analiz potokov tranzaktsii s tsel'yu pereraspredeleniya informatsionnykh resursov [Analysis of Transaction Flows to Redistribute the Information Resources], (1999), *Bulletin of the Kharkov State Polytechnic University Publ.*, Kharkov, Ukraine, Vol.73, pp.19 – 23 (In Russian).

2. Kungurtsev A.B., and Kuok Vin Nguen Chan. Analiz vozmozhnosti primeneniya materializovannykh predstavlenii v IS [Analysis of Possibility of MV Implementation in IS], (2004), *Pratsi Odessa National Polytechnic University publishing*, Odessa, Ukraine, Vol. 2(20), pp.102 – 106 (In Russian).

3. Vozovikov Y.N., and Munzer A.A. Imitatsionnoe modelirovanie materializovannykh predstavlenii Chernigivs'kogo derzhavnogo tekhnologichnogo universitetu, [Simulation Modeling of Mmaterialized Views], (2011), *Visnik Chernigivskogo Sovereign Tehnologichnogo Universitetu Publ.*, Chernigov, Ukraine, Vol. 4(53), pp. 196 – 200 (In Russian).

4. Zinovatnaya S.L. Analiz informatsionnoi sistemy dlya prinyatiya resheniya o voskhodyashchei denormalizatsii: IV seminar, [Analysis of the Information Systems for Decision-Making on Rising Denormalization], (2005), *Mathematical Modeling and Information Technologies Publ.: IV Seminar*, Odessa, Ukraine, pp. 41 – 43 (In Russian).

5. Kungurtsev A.B., Zinovatnaya S.L., Muthanna S., and Manchuk A. Analiz zaprosov k baze dannykh dlya opredeleniya polei-kandidatov na sozdanie klasternogo indeksa, [Analysis of Database Queries to Define Fields Candidates for the Creation of a Clustered Index], (2010), *Godishnik Technically*

universitet Publ., Varna, Bulgaria, Vol. 1, pp. 127 – 130.

6. Zengfeng Huang, Ke Yi, and Yunhao Liu. Optimal Sampling algorithms for Frequency Estimation in Distributed Data, (2011), *Conference: IEEE INFOCOM-INFOCOM*, pp. 1997 – 2005, doi: 10.1109/INFOCOM.2011.5935005.

7. Levchenko A.Yu. Metodika testirovaniya proizvoditel'nosti sistem upravleniya bazami dannyakh, [Methods of Testing the Performance of Database Management Systems], (2011), *Elektromashino-buduvannya that Elektroobladnannya. Mizhvidomchy Naukovy-tehnichny Zbirnik Publ.*, Odessa, Ukraine, Iss. 04(80), pp.201 – 204 (In Russian).

8. Hassan Artail, Hadi El Amine, and Fehmi Sakkal. SQL Query Space and Time Complexity Estimation for Multidimensional Queries, (2008), *International Journal of Intelligent Information and Database Systems Publ.*, Vol. 2, No. 4, pp. 460 – 480, doi: 10.1504/IJIDS.2008.021448.

9. Justus S., and Iyakutti K. An Empirical Validation of the Suite of Metrics for Object-relational Data Modelling, (2011), *International Journal of Intelligent Information and Database Systems*, Vol. 5, No. 1, pp. 49 – 80, doi: 10.1504/IJIDS.2011.037704.

10. Peter Aiken, Mark Gillenson, Xihui Zhang, and David Rafner. Data Management and Data Administration: Assessing 25 Years of Practice, (2011), *Journal of Database Management (JDM)*, Vol. 22, Issue 3, pp. 24 – 45, doi: 10.4018/jdm.2011070102.

11. Keng Siau, Fiona F.H. Nah, and Qing Cao. A Meta-Analysis Comparing Relational and Semantic Models, (2011), *Journal of Database Management (JDM)*, Vol. 22, Issue 4, pp. 57 – 72, doi: 10.4018/jdm.2011100103.



Мунзер Аль Абдо,
аспирант каф. системно-
го программного обеспе-
чения Одесского на-
ционального политехн.
ун-та, пр. Шевченко, 1,
Одесса, Украина, 65044.
Телефон: 0931562643
E-mail:
munther1427@yahoo.com