

Оптимизация физических параметров организации баз данных

Э.П. Голенищев¹, И.В. Клименко^{2✉}

¹ Донской государственный технический университет (ДГТУ); г. Ростов-на-Дону, Россия;

² Филиал Ростовского государственного университета путей сообщения в г. Туапсе (филиал РГУПС г. Туапсе); г. Туапсе, Россия

¹ goledu@yandex.ru

² klimenko.t6@gmail.com✉

АННОТАЦИЯ

Основной задачей транспортной системы является наиболее полное удовлетворение потребностей хозяйствующих субъектов и населения в перевозках, сокращение сроков доставки грузов и перевозки пассажиров на основе качественного повышения мощности всей транспортной системы. Указанная задача решается с помощью вычислительной техники и новых информационных технологий в системах управления транспортом. Самая востребованная — технология баз данных (БД), реализованная в современных автоматизированных информационных системах управления транспортом.

С учетом экспоненциального роста объемов применяемых БД, жестких временных лимитов на принятие управленческих решений в реальном времени в системах управления транспортом, особую актуальность приобретает задача минимизации времени выполнения информационных запросов в БД.

На время осуществления информационных запросов влияние оказывают методы физического хранения БД. Для решения задачи минимизации времени выполнения информационных запросов необходимо определить физические параметры, оказывающие влияние на эффективность обработки БД.

Описывается имитационная модель внешней памяти для оптимизации физических параметров организации БД, основанная на реализованных в современных системах управления базами данных (СУБД) процедурах обмена между внешней и оперативной памятью при информационных запросах. В качестве ключевого параметра оптимизации обосновывается коэффициент резервирования блока памяти, позволяющий минимизировать количество медленных обращений к внешней памяти.

Рациональность и адекватность предложенной модели в процессе исследований подтверждены результатами имитационного моделирования.

Ключевые слова: автоматизированная информационная система; система управления базами данных; база данных; язык определения данных; язык манипулирования данными; язык запросов; блок данных; запись БД; оперативная память; программное обеспечение; ключевое поле; ключ записи

Для цитирования: Голенищев Э.П., Клименко И.В. Оптимизация физических параметров организации баз данных // Техник транспорта: образование и практика. 2022. Т. 3. Вып. 2. С. 183–188. <https://doi.org/10.46684/2687-1033.2022.2.183-188>.

Original article

Optimization of the physical parameters of the organization of databases

Eduard P. Golenishchev¹, Igor V. Klimenko^{2✉}

¹ Don State Technical University (DSTU); Rostov-on-Don, Russian Federation;

² Branch of Rostov State Transport University in Tuapse (branch of RSTU in Tuapse); Krasnodar region, Tuapse, Russian Federation

¹ goledu@yandex.ru

² klimenko.t6@gmail.com✉

ABSTRACT

The main task of the transport system is the most complete satisfaction of the needs of business entities and the population in transportation, reducing the time of delivery of goods and transportation of passengers on the basis of a qualitative increase in the capacity of the entire transport system. This problem is solved with the help of computer technology and new information technologies in transport management systems. The most demanded is the database technology implemented in modern automated transport management information systems.

© Э.П. Голенищев, И.В. Клименко, 2022

Given the exponential growth in the volume of databases used, strict time limits for real-time management decisions in transport management systems, the task of minimizing the time to execute information queries in the database is of particular relevance.

The execution time of information requests is influenced by the methods of physical storage of databases. To solve the problem of minimizing the execution time of information requests, it is necessary to determine the physical parameters that affect the efficiency of database processing.

An external memory simulation model is described for optimizing the physical parameters of database organization, based on the exchange procedures implemented in modern database management systems between external and operational memory upon information requests. As a key optimization parameter, a memory block reservation factor is substantiated, which allows minimizing the number of slow accesses to external memory.

The rationality and adequacy of the proposed model in the process of research is confirmed by the results of simulation modeling.

Keywords: automated information system; database management system; database; data definition language; data manipulation language; query language; data block; database entry; operational memory; software; key field; record key

For citation: Golenishchev E.P., Klimenko I.V. Optimization of the physical parameters of the organization of databases. *Transport technician: education and practice*. 2022;3(2):183-188. (In Russ.). <https://doi.org/10.46684/2687-1033.2022.2.183-188>.

ВВЕДЕНИЕ

В Автоматизированной информационной системе (АИС) управления транспортом широко применяется технология базы данных (БД). В БД необходимая для принятия управленческих решений информация представляется в формализованном виде, позволяющем хранить, обрабатывать и передавать ее с помощью определенных физических процессов и соответствующих технических средств. Качество организации данных в БД влияет на временные показатели обработки сведений, а также на их достоверность. Временные показатели обработки информации и ее достоверность являются критически важными для информационных систем управления транспортом в режиме реального времени [1–5]. Примерный комплекс информационных задач в системах управления транспортом представлен на рис. 1.

Задачи обеспечения полноты и достоверности хранимых данных, а также требуемой производительности обработки информационных запросов относятся к функционалу специального программного обеспечения (ПО) — системам управления базами данных (СУБД). С учетом сложности описываемых в БД предметных областей, к числу которых относятся и процессы управления транспортом, выбор и настройка СУБД для проектирования БД поручается высококвалифицированным специалистам — администраторам БД (АБД). При реализации проекта БД АБД использует средства администрирования. Так, на этапе проектирования БД обязательно применяется формализованный входной язык — язык определения данных (ЯОД), а на продолжительном этапе эксплуатации БД — язык манипулирования данными (ЯМД), включающий в свой состав язык запросов (ЯЗ) [5–11].

Эффективность БД значительно зависит от квалификации ее разработчика — АБД, который проектирует инфологическую модель, отражающую семантику предметной области. Вместе с тем результативность БД отчасти зависит от применяемой СУБД, под управлением которой находится разработанная БД, так как в СУБД реализованы алгоритмы физического размещения во внешней памяти логической структуры данных, а также алгоритмы обмена данными между внешней памятью и оперативной памятью (ОП) АИС в процессе обработки информационных запросов.

Для описания предметной области в БД следует задать определенную структуру: представить сведения в памяти вычислительной системы (ВС), а также задать соответствующие семантические связи между порциями информации с использованием ЯОД. В любом случае данные будут

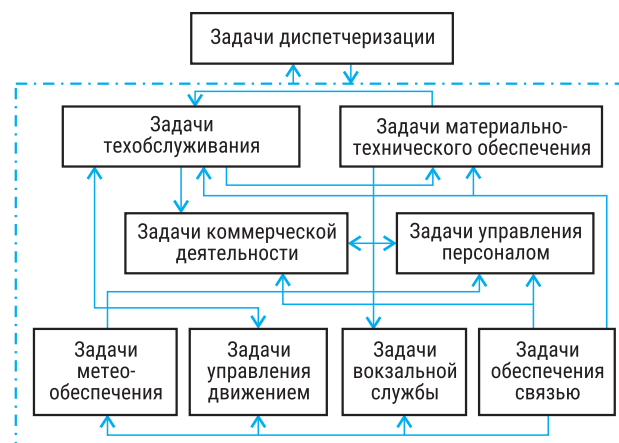


Рис. 1. Примерная структура взаимосвязей информационных задач на транспорте

храниться во внешней памяти, а их обработка будет производиться в ОП ВС. Выполнение информационных запросов в БД производится под управлением СУБД: некоторые порции данных пересылаются из внешней памяти в оперативную, происходит обработка, а результат пересылается на сохранение во внешнюю память, а также при необходимости отправляется пользователю, сформулировавшему запрос. При этом в качестве пользователей могут выступать сотрудники (персонал транспортной компании) или программные приложения БД [12].

С учетом экспоненциального роста количества хранимых данных в системах управления транспортом, для современных БД требуются значительные объемы устройств внешней памяти, а также производительные и емкие устройства ОП. Важно отметить, что объемы устройств внешней памяти на порядки превосходят объемы применяемых СУБД буферов ОП, исходя из чего, обмен между ними организуется специальными небольшими порциями — страницами или блоками (с объемом в несколько сотен байт).

ИСХОДНЫЕ ДАННЫЕ, ОГРАНИЧЕНИЯ И МЕТОДЫ

В зависимости от объема БД для физического размещения ее записей во внешней памяти распределяется соответствующее количество страниц (блоков).

Для выполнения информационного запроса определенные блоки (запись или группа записей) из внешней памяти копируются в назначенную СУБД страницу ОП. Обращение к записи производится по полю-идентификатору (ключу) или адресу [13, 14].

Как известно, устройства внешней памяти медленнее, чем устройства ОП. При этом осуществление информационных запросов проводится процессором путем обновления, изменения или удаления данных, уже размещенных в ОП. Следовательно, для повышения общей производительности АИС следует стремиться минимизировать количество медленных операций пересылки информации между ОП и устройствами внешней памяти.

Обычно записи в БД должны состоять из непосредственно данных, а также полей, выполняющих служебные функции, в частности, определяющих логическую структуру хранения или обеспечивающих ускоренный поиск по индексированным полям.

В БД могут применяться записи фиксированной или переменной длины. В любом из названных случаев для долговременного хранения записи

располагают без пропусков с целью экономии ресурсов устройств внешней памяти.

Количество k записей фиксированной длины для размещения в одном блоке внешней памяти зависит от длины записи и объема блока и рассчитывается следующим образом

$$k = \left\lfloor \frac{V_{\text{бл}}}{V_3} \right\rfloor,$$

где $V_{\text{бл}}$ — объем блока внешней памяти (в байтах); V_3 — длина записи (в байтах).

Целесообразно рассмотреть случай, когда часть страницы ОП, предназначенной для временного хранения обрабатываемой порции данных, свободна от записей. В таком случае, эту область можно наполнять, расширяя имеющиеся записи или дополняя новыми записями, которые по определенным условиям (например, по значению ключа) должны попасть в этот блок. Очевидно, в этом случае система какое-то время не будет выполнять медленные операции обмена с устройствами внешней памяти. Значит, для повышения производительности системы некоторую долю страниц ОП стоит оставлять незаполненными, с так называемыми «зарезервированными» областями.

В описанном случае требуется периодическая реорганизация БД, в частности, может потребоваться выделение дополнительного количества блоков в основной области с последующей переконфигурацией записей.

Если на страницах ОП r часть которых в начале сеанса работы с БД зарезервирована (не заполнена), то для размещения БД потребуется следующее количество страниц $N_{\text{стр}}$

$$N_{\text{стр}} = \left\lceil \frac{N_3}{\frac{V_{\text{бл}}(1-r)}{V_3}} \right\rceil,$$

где N_3 — количество записей в БД; $r < 1$ — доля зарезервированной части страницы.

Время t_0 операции пересылки блока данных из устройства внешней памяти в ОП (или наоборот) зависит от технических характеристик используемых аппаратных средств, а также от длины пересылаемого блока

$$t_0 = V_{\text{бл}} t_{\text{сч}} + t_{\text{подг}},$$

где $t_{\text{сч}}$ — время считывания одного байта (приведенное время считывания); $t_{\text{подг}}$ — время подготовки устройства внешней памяти.

Выражение для времени поиска $t_{\text{п}}$ запрашиваемых данных в БД

$$t_{\text{п}} = X_0 t_0 + X_c t_c,$$

где X_0 — необходимое для выполнения информационного запроса количество операций обмена между устройством внешней памяти и ОП; X_c —

необходимое для выполнения информационного запроса количество операций сравнения; t_c — среднее время одной операции сравнения, выполняемой процессором.

С учетом того, что $t_c \ll t_o$, время выполнения запроса на поиск данных практически определяется количеством операций обмена между устройством внешней памяти и ОП. Следовательно, при реализации в низкоуровневом функционале СУБД процедур поиска сведений в БД желательно минимизировать параметр X_o [15, 16].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

При заданных в АИС параметрах (объем БД, объем блока внешней памяти, используемая длина записи, типы используемых индексов, количество индексированных полей в записи, технические характеристики аппаратных средств, в том числе устройств внешней памяти и процессора и т.п. [8]), существенное влияние на эффективность выполнения информационных запросов оказывает и актуальное состояние БД, которое характеризуется следующими показателями:

- количество записей БД, удовлетворяющих конкретному условию Q ;
- доля r резервируемой части страницы ОП.

Предельно упрощенный пример, поясняющий возможную организацию БД при начальной загрузке, приведен на рис. 2.

Очевидно, что для начальной загрузки БД при таких характеристиках потребуется пять страниц ОП в основной области.

Ценность хранимой в БД информации заключается в возможности выдачи ответов на информационные запросы, которые могут формулироваться как людьми, эксплуатирующими эту БД, так и задачами, запрограммированными в соответствующих приложениях БД. Как правило, выполнению информационного запроса в СУБД предшествует проверка записей БД на их соответствие сформулированному условию Q . В результате такой проверки выделяется множество записей, значения полей которых удовлетворяют условию Q . В последующем выбранные записи (или часть полей найденных записей) выдаются пользователю, сформулировавшему запрос, или подвергаются модификации или удалению.

В доступных СУБД реализованы популярные условия отбора записей БД:

1. $Q_1: K = a$ (отобрать записи, значение ключевого поля которых равно заданному значению a). При организации последовательного поиска в БД, в которой ключевые поля принимают определенные значения с равной вероятностью, время $t_{п1}$ поиска по условию Q_1 можно оценить как

$$t_{п1} \approx t_o \left[\frac{n_{\text{бл.БД}}(1+r)}{2} \right],$$

где $n_{\text{бл.БД}}$ — количество блоков в БД.

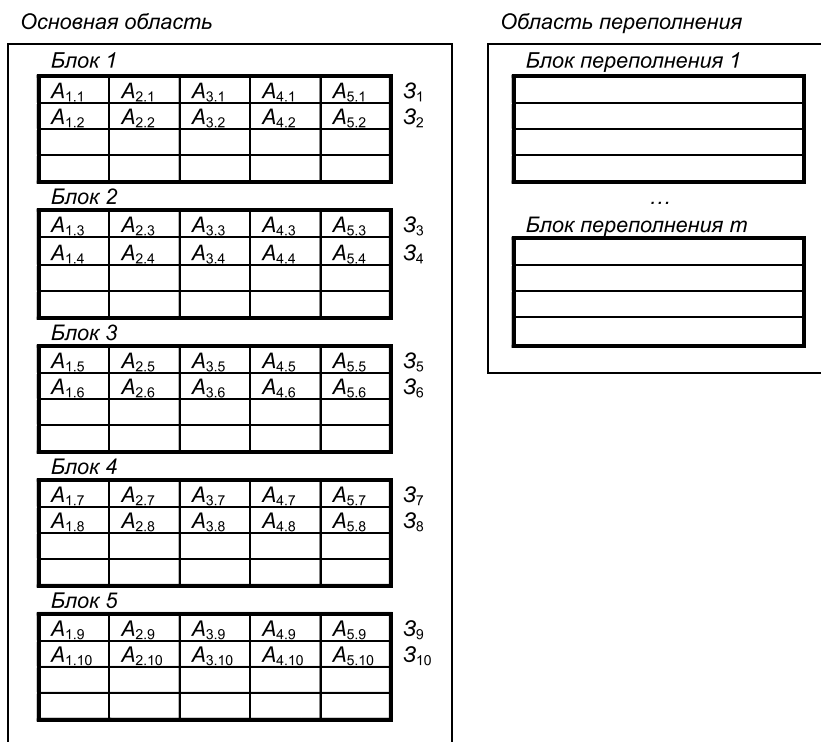


Рис. 2. Пример начальной загрузки БД со следующими характеристиками: $N_s = 10$; $r = 0,5$.

Записи БД описываются схемой: $F(A_1, A_2, A_3, A_4, A_5)$

2. $Q_2: a \leq K \leq b$ (отобрать записи, значение ключевого поля которых находится в диапазоне от a до b). При организации последовательного поиска в БД время $t_{п2}$ поиска по условию Q_2 можно оценить как

$$t_{п2} \approx t_o n_{бл.БД} (1 + r).$$

3. $Q_3: K = a_i, i = 1, 2, \dots, n$; где $a_i \in \{a_1, a_2, \dots, a_p, \dots, a_n\}$ (отобрать записи, значение ключевого поля которых равно одному из множества значений a_i). При организации последовательного поиска в БД время $t_{п3}$ поиска по условию Q_3 зависит также от n :

$$t_{п3} \approx t_o n_{бл.БД} (1 + r) + t_c \frac{N_3 n}{2}.$$

Имитационная модель для решения задачи оптимизации физических параметров организации БД должна быть основана на вышеприведенных временных оценках, а пример постановки исследовательской задачи может быть сформулирован в следующем виде.

ЗАКЛЮЧЕНИЕ

Для заданных в АИС параметров БД, а также технических характеристик аппаратных средств найти оптимальную величину $r_{опт}$:

$$r_{опт} : K_{эфф} \rightarrow \min,$$

где $K_{эфф}$ — обобщенный показатель эффективности обработки запросов в БД.

На начальном этапе в качестве обобщенного показателя эффективности рекомендуется использовать его аддитивную форму:

$$K_{эфф} = \alpha t_{обн} + \gamma t_{п1} + \delta t_{п2} + \sigma t_{п3},$$

где $\alpha, \gamma, \delta, \sigma$ — весовые коэффициенты, показывающие предпочтения разработчика в отношении наиболее важных показателей эффективности обработки информационных запросов в БД.

Для получения зависимости $K_{эфф}$ от r и нахождения гопт в процессе имитации целесообразно построить временные зависимости:

а) $t_{обн}$ от r ;

б) $t_{п1}, t_{п2}$ и $t_{п3}$ от r .

Найденное предложенным способом имитационного моделирования значение коэффициента резервирования гопт обеспечит оптимальные характеристики физической организации БД по критериям времени выполнения информационных запросов, а также позволит принимать обоснованные решения по необходимым объемам внешней и оперативной памяти в автоматизированной информационной системе управления транспортом.

ЛИТЕРАТУРА

- Алексеев О.П., Пронин С.В. Применение баз данных для оценки работы пассажирского транспорта // Автомобильный транспорт. 2009. № 24. С. 94–97.
- Нурмухамедов Т.Р., Гулямов Ж.Н. Автоматизация учета товарных ценностей на складе вагонного депо // Интеллектуальные технологии на транспорте. 2021. № 4 (28). С. 39–45. DOI: 10.24412/2413-2527-2021-428-39-45
- Похилко С.П., Иванченко И.С. Внедрение базы данных отказов приема порожних грузовых вагонов к погрузке в автоматизированные системы железнодорожного транспорта // Сборник научных трудов ДониЖТ. 2021. № 61. С. 40–48.
- Танайно Ю.А., Бессоненко С.А., Медведев В.И. Автоматизированная система управления качеством погрузки // Вестник Сибирского государственного университета путей сообщения. 2018. № 2 (45). С. 19–23.
- Фоменко Ю.С., Абу Хасан Р., Хомоненко А.Д. Логическая интеграция данных на примере организации запросов к базе данных авиаперевозок // Интеллектуальные технологии на транспорте. 2019. № 4 (20). С. 39–47.
- Дейт К. Введение в системы баз данных. М.: Наука, 1980. 463 с.
- Мартин Дж. Планирование развития автоматизированных систем. М.: Финансы и статистика, 1984. 196 с.
- Четвериков В.Н., Ревунков Г.И., Самохвалов Э.Н. Базы и банки данных: учебник для вузов. М.: Высш. шк., 1987. 245 с.
- Джексон Г. Проектирование реляционных баз данных для использования с микроЭВМ. М.: Мир, 1991. 252 с.
- Коннолли Т., Бегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика. 2-е изд. / пер. с англ. М.: Вильямс, 2000. 1120 с.
- Озкарахан Э. Машины баз данных и управление базами данных / пер. с англ. М.: Мир, 1989. 695 с.
- Голенищев Э.П., Клименко И.В. Информационное обеспечение систем управления. Ростов н/Д.: Феникс, 2010. 315 с.
- Мейер Д. Теория реляционных баз данных / пер. с англ. М.: Мир, 1987. 608 с.
- Копейкин М.В., Спиридонов В.В., Шумова Е.О. Выбор способа хранения отношений в информационных системах // Записки Горного института. 2014. Т. 208. С. 249–255.
- Дрождин В.В., Володин А.М. Методы физической организации данных, поддерживаемые существующими системами управления данных // Известия Пензенского государственного педагогического университета им. В.Г. Белинского. 2008. № 12. С. 103–106.
- Дорофеева О.С., Казаков Б.В., Казакова И.А. Критерий оптимального размещения блоков базы данных, основанный на простых цепях Маркова // Труды Международного симпозиума «Надежность и качество». 2020. Т. 1. С. 44–45.

REFERENCES

1. Alexeev O., Pronin S. The application of database for passenger transport estimation. *Automobile Transport*. 2009;24:94-97. (In Russ.).
2. Nurmukhamedov T.R., Gulyamov J.N. Automation of inventory accounting in the wagon depot warehouse. *Intellectual Technologies on Transport*. 2021;4(28):39-45. DOI: 10.24412/2413-2527-2021-428-39-45 (In Russ.).
3. Pokhilko S.P., Ivanchenko I.S. Implementation of a database of refusals to accept empty freight cars for loading into automated railway transport systems. *Collection of Scientific Works of the Donetsk Institute of Railway Transport*. 2021;61:40-48. (In Russ.).
4. Tanayno Yu. A., Bessonenko S.A., Medvedev V.I. Automated quality management system for loading. *Bulletin of the Siberian State University of Communications*. 2018;2(45):19-23. (In Russ.).
5. Fomenko Yu.S., Abou Hasan R., Khomonenko A.D. Logical data integration on the example of organizing queries to the database air transportation. *Intellectual Technologies on Transport*. 2019;4(20):39-47. (In Russ.).
6. Date C.J. *An Introduction to Database Systems*. Moscow, Nauka, 1980;463. (In Russ.).
7. Martin J. *Strategic data-planning methodologies*. Moscow, Finance and statistics, 1984;196. (In Russ.).
8. Chetverikov V.N., Revunkov G.I., Samokhvalov E.N. *Databases and data banks: textbook for universities*. Moscow, Higher school, 1987;245. (In Russ.).
9. Jackson G. *Designing relational databases for use with microcomputers*. Moscow, Mir, 1991;252. (In Russ.).
10. Connolly T., Begg K., Strachan A. *Databases: design, implementation and maintenance. Theory and practice, 2nd ed.* Moscow, Williams, 2000;1120. (In Russ.).
11. Ozkarakhan E. *Database machines and database management*. Moscow, Mir, 1989;695. (In Russ.).
12. Golenishchev E.P., Klimenko I.V. *Information support of control systems*. Rostov-on-Don, Phoenix, 2010;315. (In Russ.).
13. Meyer D. *Theory of relational databases*. Moscow, Mir, 1987;608. (In Russ.).
14. Kopeikin M.V., Spiridonov V.V., Shumova E.O. The choice of a method for storing relations in information systems. *Journal of Mining Institute*. 2014;208:249-255. (In Russ.).
15. Drozhdin V.V., Volodin A.M. Methods of physical organization of data, supported by existing data management systems. *Izvestiya of the Penza State Pedagogical University*. V.G. Belinsky. 2008;12:103-106. (In Russ.).
16. Dorofeeva O.S., Kazakov B.V., Kazakova I.A. Criteria for the optimal placement of database blocks based on simple Markov chains. *Transactions of the International Symposium on Reliability and Quality*. 2020;1:44-45. (In Russ.).

Об авторах

Эдуард Павлович Голенищев — кандидат технических наук, доцент, доцент кафедры вычислительных систем и информационной безопасности; **Донской государственный технический университет (ДГТУ)**; 344003, г. Ростов-на-Дону, пл. Гагарина, д. 1; РИНЦ ID: 611030 goledu@yandex.ru;

Игорь Валерьевич Клименко — кандидат технических наук, заведующий кафедрой гуманитарных, естественнонаучных и общепрофессиональных дисциплин; **Филиал Ростовского государственного университета путей сообщения в г. Туапсе (филиал РГУПС г. Туапсе)**; РИНЦ ID: 528034 352800, Краснодарский край, г. Туапсе, ул. Богдана Хмельницкого, д. 85; klimenko.t6@gmail.com.

Bionotes

Eduard P. Golenishchev — Cand. Sci. (Tech.), Associate Professor, Associate Professor of the Department of Computer Systems and Information Security; **Don State Technical University (DSTU)**; 1 Gagarin square, Rostov-on-Don, 344003, Russian Federation; RSCI ID: 611030 goledu@yandex.ru;

Igor V. Klimenko — Cand. Sci. (Tech.), Head of the Department of Humanities, Natural Sciences and General Professional Disciplines; **Branch of Rostov State Transport University in Tuapse (branch of RSTU in Tuapse)**; 85 Bohdan Khmelnytsky st., Krasnodar region, Tuapse, 352800, Russian Federation; RSCI ID: 528034 klimenko.t6@gmail.com.

Заявленный вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article.

The authors declare no conflicts of interests.

Автор, ответственный за переписку: Игорь Валерьевич Клименко, klimenko.t6@gmail.com.

Corresponding author: Igor V. Klimenko, klimenko.t6@gmail.com.

Статья поступила в редакцию 11.02.2022; одобрена после рецензирования 29.03.2022; принята к публикации 30.05.2022.

The article was submitted 11.02.2022; approved after reviewing 29.03.2022; accepted for publication 29.05.2022.