

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 8, №4 (2016) <http://naukovedenie.ru/index.php?p=vol8-4>

URL статьи: <http://naukovedenie.ru/PDF/49TVN416.pdf>

Статья опубликована 16.08.2016.

Ссылка для цитирования этой статьи:

Сырин А.И. Алгоритм ранговой оптимизации доступа к данным в информационной системе // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 8, №4 (2016) <http://naukovedenie.ru/PDF/49TVN416.pdf> (доступ свободный).

Загл. с экрана. Яз. рус., англ.

УДК 004.623

Сырин Александр Иванович

ФГБОУ ВО «Воронежский государственный архитектурно-строительный университет», Россия, Воронеж¹

Аспирант кафедры «Управление строительством»

E-mail: vp_morozov@mail.ru; kus_vgasu@mail.ru

Алгоритм ранговой оптимизации доступа к данным в информационной системе

Аннотация. В статье представлен алгоритм ранговой оптимизации доступа к данным в информационной системе. Автором установлено, что применять разработанный алгоритм целесообразно для информационных систем в состав которых входит несколько баз данных большого объема (например, хранилище данных и витрина данных), содержащих разнородные данные. Предложена модифицированная трехслойная технология применения хранилища данных. Суть модификации заключается в введении четвертого технологического слоя под названием «Оптимизация загрузки данных», который предназначен для перемещения требуемых данных из хранилища данных в витрину данных. Его основу составляют соответствующие алгоритмы ранговой оптимизации доступа к данным. На основе вычислительных экспериментов установлено, что запросы, формируемые пользователем, представляют собой поток событий с соответствующим значением интенсивности, подчиняющийся закону Пуассона. Апробация разработанного алгоритма применительно к системе поддержки принятия инвестиционных решений позволила получить оценки интенсивности потока данных, запрашиваемых пользователем, и вероятности поступления запроса на их востребованность, представляющей по сути их ранг. Получено правило ранжирования наборов данных (в том числе и файлов) в соответствии с их рангом, позволяющее минимизировать расчетные затраты на повторное извлечение данных, содержащихся в хранилище. Установлено, что эффективность реализации данного алгоритма возрастает, если увеличивается число потенциальных кандидатов на перенос (объем базы данных). В случае, когда количество переносимых наборов данных невелико, что характерно для начала эксплуатации информационной системы, целесообразно реализовать их (данных) простой последовательный перенос.

Ключевые слова: алгоритм; витрина данных; закон Пуассона; запрос; интенсивность потока данных; информационная система; набор; оптимизация; ранг популярности; хранилище данных

¹ 394006, г. Воронеж, ул. 20-лет Октября, 84

В *информационных системах (ИС)*, содержащих объемные *базы данных (БД)*, процедура поиска и загрузки требуемых данных для исследования как правило имеет низкую оперативность. В большинстве случаев после перерыва в работе с ИС пользователю требуются массивы данных с которыми он работал в ближайшее время до прекращения работы с ней.

В некоторых ИС предусмотрена процедура запоминания последних обрабатываемых данных. Для загрузки более ранних данных, что необходимо довольно часто, требуется запустить процедуру поиска и ждать достаточно продолжительное время. Для сокращения времени ожидания предлагается при записи данных в БД использовать их ранговую популярность.

Цель работы – предложить алгоритм ранговой оптимизации доступа к данным в ИС.

Разработанный алгоритм предназначен для ИС, содержащих большие объемы разнородных данных. Одним из классов таких ИС являются *системы поддержки принятия решений (СППР)*.

СППР представляет собой интерактивную автоматизированную систему, которая помогает пользователю – лицу, принимающему решения (ЛПР) использовать данные и модели для идентификации и решения задач с целью принятия решений. Она обладает возможностью работать с интерактивными запросами с достаточно простым для изучения языком запросов [1]. СППР обладает следующими четырьмя основными характеристиками [2]:

- использует и данные, и модели;
- предназначена для помощи менеджерам в принятии решений для слабоструктурированных и неструктурированных задач;
- поддерживает, а не заменяют, выработку решений менеджерами;
- цель СППР – улучшение эффективности решений.

Их характерные особенности заключаются в изменении производительности в различные периоды времени, непредсказуемости уровня нагрузки и нерегулярности применения [3].

Для хранения разнородной информации в СППР как правило используются *хранилища данных (ХД)*. Автором концепции хранилищ данных, разработанной в конце 90-х годов, является Б. Инмон. Он определил хранилища данных, как: «предметно ориентированные, интегрированные, неизменчивые, поддерживающие хронологию наборы данных, организованные для целей поддержки управления», призванные выступать в роли «единого и единственного источника истины», обеспечивающего руководителей и аналитиков достоверной информацией, необходимой для оперативного анализа и принятия решений [4]. Отличительными свойствами ХД являются [5]:

- предметная ориентация данных (данные, содержащиеся в ХД, ориентированы на задачи поддержки принятия решений, а не на используемые приложения);
- интегрированность данных (данные очищены от индивидуальных признаков и приведены к единому формату);
- инвариантность данных во времени (сохранение данными своей истинности в любой момент проводимых с ними манипуляций);
- стабильность информации (однажды загруженные данные практически никогда не меняются).

Компанией Oracle - одним из крупнейших разработчиков ХД предложена трехслойная технология их использования [6]. Автором предлагается дополнить данную технологию четвертым слоем, который носит название «Оптимизация загрузки данных». В схематичном виде предложенная технология приведена на рисунке 1. Механизм ее функционирования заключается в следующем.

Данные поступают на хранение из различных источников. На данном этапе данные извлекаются из них (источников), преобразуются в соответствующий формат ХД и загружаются для хранения. Подобные манипуляции с данными реализуются соответствующим технологическим слоем «Преобразования и загрузки данных».

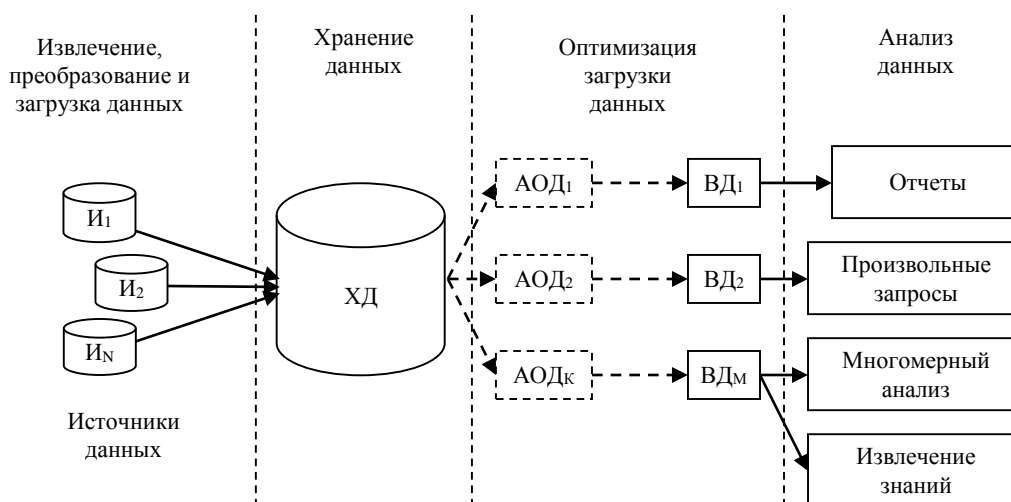


Рисунок 1. Четырехслойная технология использования ХД (разработано автором)

Второй технологический слой – «Хранение данных» обеспечивает хранение в ХД существенной, выверенной, согласованной, непротиворечивой и целостной информации упорядоченной в хронологическом порядке. ХД обладает нейтральностью по отношению к приложениям. Оно является своеобразной «емкостью» для хранения нормализованных данных.

Третий технологический слой – «Оптимизация загрузки данных» который предлагается автором предназначен для перемещения требуемых данных из ХД в витрину данных (ВД). ВД является функционально-ориентированной БД в которую загружаются оперативные данные необходимые для решения частных текущих информационно-аналитических задач. Для загрузки оперативных данных разработан соответствующий алгоритм оптимизации доступа (АОД) описание которого приведено ниже.

Технологический слой «Анализ данных» является завершающим в процессе подготовки информации для принятия решений. Он может включать формирование различных отчетов, генерацию и обработку произвольных запросов, многомерный анализ, извлечение знаний и др.

Рассмотрим более дельно АОД реализующий загрузку наиболее востребованных данных. При этом данные перемещаются в зависимости от ранга их популярности. В качестве меры популярности использована эмпирическая оценка вероятности поступления запроса: чем выше вероятность поступления запроса, тем выше ранг данных. Экспериментальным путем установлено, что запросы, формируемые пользователем, представляют собой поток событий, подчиняющийся закону Пуассона [7]. Интенсивность потока (количества событий в единицу времени) запросов на *i*-й набор данных определяется в соответствии со следующей эмпирической оценкой:

$$\lambda^i = \frac{N^i}{t_N^i - t_o^i}, \tag{1}$$

где: N^i - полное количество запросов на i - й набор данных, t_o^i - время поступления данных в архив, t_N^i - время поступления N^i -го запроса.

Вероятность поступления запроса на i - й набор данных при интенсивности потока заказов λ^i в момент времени $t > t_N^i$ определяется выражением (вероятность единичного события на интервале $t - t_N^i > 0$) [8]

$$p^i = 1 - \exp \{- \lambda^i (t - t_N^i)\}. \tag{2}$$

Полученные оценки определяют ранг набора данных. Для множества из K наборов данных ранжирование сводится к сортировке данных в соответствии с правилом:

$$p^i < p^j, \text{ при } i < j \text{ для любых } i \in 1, K; j \in 1, K. \tag{3}$$

В результате данной сортировки наборов данных набор с индексом i является искомым, который необходимо перенести в ХД. В этом случае минимизируются расчетные затраты на повторное извлечение данных из ХД.

Блок-схема алгоритма оптимизации доступа, представлена на рисунке 2.

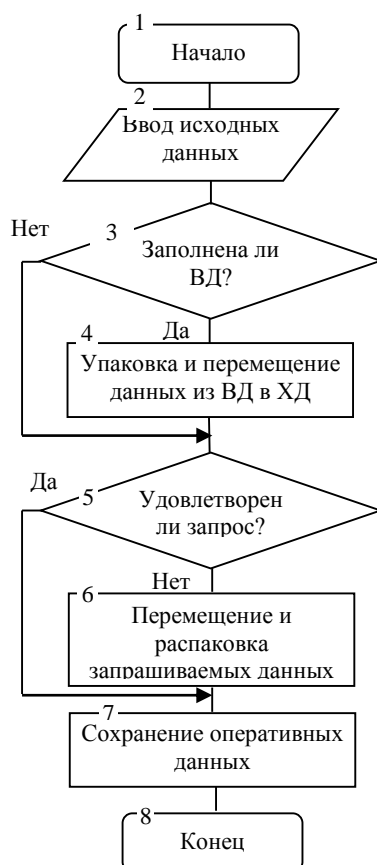


Рисунок 2. Блок-схема алгоритма оптимизации доступа (разработано автором)

Блоки 1, 8 предназначены для пуска и завершения процесса обмена данными между ХД и ВД.

Блок 2 обеспечивает реализацию ввода исходных данных. Наиболее значимыми из них являются значения верхнего и нижнего уровней объема данных ВД. Они являются критериями запуска и остановки процедуры перемещения данных. При этом используются следующие правила:

1. Система перемещения данных из ВД в ХД запускается, если общий объем хранимых в ВД данных превышает верхний допустимый уровень;
2. Система перемещения данных из ВД в ХД прекращает работу, когда общий объем хранимых в ВД данных становится меньшим, чем нижний допустимый уровень.

Блок 3 предназначен для проверки состояния загруженности ВД. Если ВД не загружено, то осуществляется переход к блоку 5. Иначе управление передается блоку 4.

В блоке 4 реализована архивация и перемещение данных из ВД в ХД.

Блок 5 предназначен для реализации поиска данных, запрашиваемых пользователем. В том случае, если они содержатся в ХД, то осуществляется их непосредственная загрузка и дальнейшая работа с ними. В случае их размещения в ВД управление передается в блок 6.

В блоке 6 реализуется поиск запрашиваемых данных в ХД. В случае их наличия, последние перемещаются из ХД в ВД, распаковываются (разархивируются) и предоставляются пользователю для дальнейшей работы.

В блоке 7 реализована запись полученных оперативных данных в ВД.

Разработанный алгоритм прошел апробацию в рамках системы поддержки принятия инвестиционных решений [9, 10]. В ходе ее проведения установлено, что эффективность использования данного алгоритма увеличивается при возрастании числа хранимых наборов данных. При незначительном их количестве, что характерно для начального этапа эксплуатации ИИС, целесообразно использовать простой последовательный перенос данных.

ЛИТЕРАТУРА

1. Bonczek, R.H. Foundation of Decision Support Systems / R.H. Bonczek, C. Holsapple, A.B. Winston. – New York: Academic Press, 1981. – 324 p.
2. Turban, E. Decision support and expert systems: management support systems. – Englewood Cliffs, N.J.: Prentice Hall, 1995. – 887 p.
3. Дейт, К.Дж. Введение в системы баз данных: Пер. с англ. / К.Дж. Дейт - М.: Издательский дом "Вильямс", 2005. - 1328 с.
4. Inmon, W.H. Building The Data Warehouse. - New York, N.Y.: John Wiley & Sons, 1992.
5. Найк, Дайлип. Системы хранения данных в Windows.:Пер. с англ. / Дайлип Найк. - М.: Издательский дом «Вильямс», 2005. – 432 с.
6. <http://oracle.com/ru/>.
7. Айвазян, С.А. Прикладная статистика. Основы моделирования и первичная обработка данных / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин – М.: Финансы и статистика, 1983. – 472 с.
8. Бронштейн, И.Н. Справочник по математике для инженеров и учащихся вузов / И.Н. Бронштейн, К.А. Семендяев. – М.: Наука, 1986. – 534 с.
9. Сырин, А.И. Система поддержки принятия инвестиционных решений малого предприятия / С.А. Баркалов, В.П. Морозов, А.В. Никитенко, А.И. Сырин. - Воронеж: Воронежский ГАСУ. - 2014. – 162 с.
10. Сырин, А.И. Модели и алгоритмы проектирования и разработки систем поддержки принятия инвестиционных решений: монография / В.П. Морозов, С.А. Баркалов, А.И. Сырин. - Воронеж: Воронежский ГАСУ. - 2015. – 232 с.

Syrin Alexander Ivanovich

Voronezh state architectural and construction university, Russia, Voronezh
E-mail: vp_morozov@mail.ru; kus_vgasu@mail.ru

Algorithm of rangovy optimization of access to data in information system

Abstract. In article the algorithm of rangovy optimization of access to data in information system is provided. By the author it is established that it is reasonable to apply developed algorithm to information systems which part several databases of large volume are (for example, storages of data and a show-window of data), containing diverse data. The modified three-layer technology of application is offered storages of data. The essence of modification consists in entering of the fourth technological layer under the name "Optimization of loading of data" which is intended for movement of the required data from storage of data in a show-window of data. Its basis is constituted by the corresponding algorithms of rangovy optimization of access to data. On the basis of computing experiments it is established that the requests created by the user represent the flow of events with the corresponding value of intensity submitting to Poisson's law. Approbation of developed algorithm in relation to system of support of adoption of investment decisions I have allowed to receive estimates of intensity of a data flow, requested by the user, and probabilities receipts of request for their demand, representing in fact their rank. The rule of ranging is received data sets (including files) according to their rank, allowing to minimize estimated costs on repeated extraction of the data containing in storage. It is established that efficiency of implementation this algorithm increases if the number of potential candidates increases by transfer (database amount). In a case when the quantity of transferable data sets is small that is characteristic to start operation information system, it is reasonable to realize their (data) simple consecutive transfer.

Keywords: algorithm; show-window of data; Poisson's law; request; intensity of a data flow; information system; set; optimization; popularity rank; storage of data

REFERENCES

1. Bonczek, R.H. Foundation of Decision Support Systems / R.H. Bonczek, C. Holsapple, A.B. Whinston. – New York: Academic Press, 1981. – 324 p.
2. Turban, E. Decision support and expert systems: management support systems. – Englewood Cliffs, N.J.: Prentice Hall, 1995. – 887 p.
3. Deyt, K.Dzh. Vvedenie v sistemy baz dannykh: Per. s angl. / K.Dzh. Deyt - M.: Izdatel'skiy dom "Vil'yams", 2005. - 1328 s.
4. Inmon, W.H. Building The Data Warehouse. - New York, N.Y.: John Wiley & Sons, 1992.
5. Nayk, Daylip. Sistemy khraneniya dannykh v Windows.:Per. s angl. / Daylip Nayk. - M.: Izdatel'skiy dom «Vil'yams», 2005. – 432 s.
6. <http://oracle.com/ru/>.
7. Ayvazyan, S.A. Prikladnaya statistika. Osnovy modelirovaniya i pervichnaya obrabotka dannykh / S.A. Ayvazyan, I.S. Enyukov, L.D. Meshalkin – M.: Finansy i statistika, 1983. – 472 s.
8. Bronshteyn, I.N. Spravochnik po matematike dlya inzhenerov i uchashchikhsya vuzov / I.N. Bronshteyn, K.A. Semendyaev. – M.: Nauka, 1986. – 534 s.
9. Syrin, A.I. Sistema podderzhki prinyatiya investitsionnykh resheniy malogo predpriyatiya / S.A. Barkalov, V.P. Morozov, A.V. Nikitenko, A.I. Syrin. - Voronezh: Voronezhskiy GASU. - 2014. – 162 s.
10. Syrin, A.I. Modeli i algoritmy proektirovaniya i razrabotki sistem podderzhki prinyatiya investitsionnykh resheniy: monografiya / V.P. Morozov, S.A. Barkalov, A.I. Syrin. - Voronezh: Voronezhskiy GASU. - 2015. – 232 s.