

Эвристический алгоритм решения задачи оптимального размещения информационных ресурсов

А. Н. Скоба, Айеш Ахмед Нафеа Айеш (Ирак), В.К. Михайлов

Южно-Российский государственный политехнический университет (НПИ)

им.М. И. Платова, Новочеркасск

Аннотация: В данной статье для решения задачи оптимального распределения информационных ресурсов по узлам распределённой информационной системы по критерию минимума среднего времени реакции системы на запросы пользователей был предложен оригинальный эвристический алгоритм, использующий представление о базах данных, как о точках многомерного пространства, а об узлах, в которых эти базы размещаются как о кластерах. Представлены результаты численных экспериментов в зависимости от типа используемой архитектуры распределённой системы и от способа обеспечения целостности данных. Численные эксперименты показали достаточно высокую эффективность предложенного алгоритма

Ключевые слова: распределённая информационная система, распределённая база данных, локальная вычислительная сеть, нелинейная комбинаторная оптимизация, интенсивности формирования запросов, множество запросов на чтение информации, матрица вероятностей формирования запросов пользователями, матрица распределения отношений по узлам локальной вычислительной сети, матрица объёмов считываемой информации, средний объём информации циркулирующей между пользователем и отношением, архитектура «файл-сервер», двухуровневая архитектура «клиент-сервер», блокировка базы данных, время реакции системы.

Как было отмечено в работах [1,2], одной из важных задач, без решения которой невозможна разработка распределённой информационной системы, является рациональная организация вычислительного процесса, реализованного в среде ЛВС. В процессе работы ЛВС поддерживает распределённую базу данных (РБД) [3], которая обладает несомненными преимуществами перед централизованной (меньшее время ответа для пользователей, меньшее время блокировки записей, более простое планирование заявок). Однако при проектировании таких систем будут существовать большие общие пересекающиеся массивы данных, в которых определённая информация будет присутствовать многократно. Неоптимальное обращение с информационными ресурсами может с одной стороны увеличить время реакции системы на запросы пользователей, а с

другой стороны, может стать очень дорогим для пользователей [3]. Поэтому в одной из задач, возникающих при проектировании распределённой информационной системы на базе ЛВС, является задача оптимального размещения информационных ресурсов (частей РБД) по узлам ЛВС, включающая выбор топологии сети, критерия эффективности, конструирование математической модели, разработку алгоритма оптимизации и ее программную реализацию.

Исходными данными для моделирования являлись [4-7]: множество узлов сети – $U = \{U_1, \dots, U_s, \dots, U_n\}$; множество пользователей – $A = \{A_1, \dots, A_s, \dots, A_n\}$; множество отношений – $R = \{R_1, \dots, R_j, \dots, R_d\}$; множество интенсивностей формирования запросов – $\Lambda = \{\lambda_1, \dots, \lambda_s, \dots, \lambda_n\}$; множество запросов на чтение – $Q = \{Q_1, \dots, Q_l, \dots, Q_q\}$; множество объёмов отношений – $V = \{V_1, \dots, V_j, \dots, V_d\}$; скорость считывания в узлах – $VV = \{VV_1, \dots, VV_s, \dots, VV_n\}$; скорость записи в узлах – $VD = \{VD_1, \dots, VD_s, \dots, VD_n\}$; скорость передачи данных по каналу связи – θ ; постоянная задержка при передаче данных по каналу связи – θ_0 ; постоянная задержка при обработке в узле – α_0 ; производительность процессора U_z -го узла – $PU_z(z = \overline{1, n})$; матрица вероятностей формирования запросов пользователями – $F = \|f_{A_s Q_l}\| = \|f_{sl}\| (s = \overline{1, n}, l = \overline{1, q})$, где элемент f_{sl} представляет собой вероятность того, что s -й пользователь сформировал l -й запрос; матрица объёмов считываемой информации $B = \|b_{Q_l R_j}\| = \|b_{lj}\| (l = \overline{1, q}, j = \overline{1, d})$, где b_{lj} – объём считываемой информации по l -му запросу на чтение из j -го отношения; $\delta(B) = \|\delta(b_{Q_l R_j})\| = \|\delta_{lj}\| (l = \overline{1, q}, j = \overline{1, d})$, где

$$\delta_{lj} = \begin{cases} 1, \text{ если } l - \text{й запрос обращается к } j - \text{му отношению,} \\ 0, \text{ такого отношения нет,} \end{cases}$$

причём $\sum_{j=1}^d \delta_{lj} = 1, l = \overline{1, q}$; матрица распределения отношений по узлам ЛВС -

$$X = \|x_{R, U_k}\| = \|x_{jk}\| (j = \overline{1, d}, k = \overline{1, n}), \text{ где } x_{jk} = \begin{cases} 1, \text{ если } R_j \in U_k, \\ 0 - \text{ в противном случае,} \end{cases}$$

причём $\sum_{k=1}^n x_{jk} = 1, j = \overline{1, d}$.

Таким образом задача оптимального размещения РБД по узлам ЛВС по критерию среднего времени реакции системы сводится к задаче

$$\left. \begin{aligned} \bar{T} = \bar{T}(U, R, X, A, V, Q, VV, VD, \theta, PU, F, B) \rightarrow \min, \\ \text{при ограничении} \\ \sum_{k=1}^n x_{jk} = 1, j = \overline{1, d}. \end{aligned} \right\} \quad (1)$$

Подробный расчёт величины $\bar{T} = \bar{T}(U, R, X, A, V, Q, VV, VD, \theta, PU, F, B)$ в зависимости от особенностей используемой архитектуры распределённой системы был выполнен в работах [4-7].

Как показал проведённый анализ, задача (1) является задачей нелинейной комбинаторной оптимизации с булевыми переменными. Ввиду того, что функция \bar{T} имеет сложный вид, а так же ввиду отсутствия в настоящее время эффективных алгоритмов решения такого класса задач (кроме как метод полного перебора) [8-10], для решения задачи (1) может быть применен разработанный авторами эвристический алгоритм, основанный на численном прогнозировании поведения целевой функции и использующий представления о базах данных, как о точках многомерного пространства, а об узлах, в которых эти базы размещаются как о кластерах. Обозначим

$P_{ij} = \sum_{l=1}^n f_{ij} \delta_{ij}(l = \overline{1, n}, j = \overline{1, d})$ – вероятность того, что i -й пользователь обратится

к j -му отношению; $V_{ij} = \sum_{l=1}^n f_{ij} \delta_{ij}(l = \overline{1, n}, j = \overline{1, d})$ – средний объем информации,

циркулирующий между i -м пользователем и j -м отношением.

1. Полагаем $K_{ij} = \lambda_i P_{ij} V_{ij} (i = \overline{1, n}, j = \overline{1, d}), k = 0$.
2. Выбираем для $\forall j = \overline{1, d}: K_{sj} = \max_i \{K_{ij}\} (i = \overline{1, n})$ и отношение R_j помещаем в s -ю ПЭВМ.
3. Вычисляем значения $\{\bar{T}_r\} (r = \overline{1, n})$, по формуле (7), а значение \bar{T} по формуле (8) работы [4]. Если $k \neq 0$ – переходим на шаг 7, в противном случае – на следующий шаг.
4. Полагаем $\{\bar{T}_i^{(k)}\} = \{\bar{T}_i\} (i = \overline{1, n})$ и $\bar{T}^{(k)} = \bar{T}$.
5. Выбираем $\bar{T}_\mu = \max_i \{\bar{T}_i^{(k)}\}, i = \overline{1, n}$ и $K_{\mu f} = \max(\{K_{\mu i}\} / \{K_{\mu f}\})$, где $\{K_{\mu f}\}: f = \overline{1, d}, x_{\mu f} = 1$.
6. Если $K_{\mu f} = -1$ или $\{K_{\mu i}\} / \{K_{\mu f}\} = \{\emptyset\}$, то переходим на шаг 8, иначе отношение R_j помещаем в μ -ю ПЭВМ, полагаем $k = k + 1$ и переходим на шаг 3.
7. Если $\bar{T} < \bar{T}^{(k)}$ – переходим на шаг 4, в противном случае полагаем $K_{\mu f} = -1$ и переходим на шаг 5.
8. Конец. (Полученное распределение X и будет оптимальным).

Целесообразно пункты алгоритма 1-3 выполнять уже на этапе конструирования исходных данных, что совершенно очевидно приведёт к уменьшению количества итераций алгоритма и соответственно, к уменьшению временных затрат на его работу.

Данный алгоритм был программно реализован на языке C# при следующих исходных данных: скорость считывания в s -м узле

$VV_s \in [60000; 100000]$ КБ/сек, $s = \overline{1, n}$; скорость записи в оперативную память s -го узла $VD_s \in [1 \times 10^7; 3 \times 10^7]$ КБ/сек, $s = \overline{1, n}$; производительность процессора s -го узла $PU_s \in [2,5 \times 10^9; 3,5 \times 10^9]$ операций/сек, $s = \overline{1, n}$; скорость передачи данных по каналу связи $\theta \in [1000; 10000]$ КБ/с; постоянная задержка при передаче по каналу $\theta_0 = 3 \times 10^{-4}$ с; постоянная задержка при обработке в узле $\alpha_0 = 3 \times 10^{-6}$ с; объём j -го отношения $V_j \in [50000; 150000]$ КБ, $j = \overline{1, d}$; объём считываемой информации $b_{lj} \in [1; 1000]$ КБ ($l = \overline{1, q}, j = \overline{1, d}$), $j = \overline{1, d}$ по l -му запросу на чтение из j -го отношения, по l -му запросу на чтение из j -го отношения; $\bar{b}_{lj} \in [1; 500]$ КБ ($l = \overline{1, q}, j = \overline{1, d}$) – объём информации, получаемый после процессорной обработки по l -му запросу на чтение из j -го отношения. Расчет проводился на компьютере на базе процессора фирмы Intel, с тактовой частотой 3,0 ГГц. В таблицах 1-3 приведены некоторые результаты машинных экспериментов, полученных для решения задачи оптимального размещения информационных ресурсов по узлам распределённой информационной системы: в таблице № 1-на базе файл-серверной архитектуры при отсутствии блокировок, в таблице № 2-на базе файл-серверной архитектуры с учётом влияния блокировок на уровне всей базы данных, в таблице № 3-без учёта влияния блокировок на базе двухуровневой архитектуры «клиент-сервер», в таблице №4- с учётом влияния блокировок на базе двухуровневой архитектуры «клиент-сервер».

Таблица №1

Размерность задачи $n \times d \times q$	Начальное значение Γ	Число итераций МПП	Значение $\Gamma^{(r)}$	Время решения задачи МПП, с	Число итераций ЭА	Значение $\Gamma^{(n)}$	Время решения задачи ЭА, с
3x4x5	6,7718	81	1,8644	7,1	1	1,8787	1,02
6x8x10	1,6890	6 ⁸	0,6991	216	3	0,6298	16,35
8x13x15	2,4534	8 ¹³	-	-	5	1,5490	142,6
10x15x20	1,0437	10 ¹⁵	-	-	8	0,4417	1182,4

Таблица №2

Размерность задачи $n \times d \times q$	Начальное значение Γ	Число итераций МПП	Значение $\Gamma^{(r)}$	Время решения задачи МПП, с	Число итераций ЭА	Значение $\Gamma^{(n)}$	Время решения задачи ЭА, с
3x4x5	7,8916	81	2,9403	7,7	2	2,8697	1,37
6x8x10	3,1814	6 ⁸	0,7993	245	6	0,7303	40,35
8x13x15	14,5347	8 ¹³	-	-	10	11,7819	207,6
10x15x20	34,8739	10 ¹⁵	-	-	14	14,3618	1908,6

Таблица №3

Размерность задачи $n \times d \times q$	Начальное значение Γ	Число итераций МПП	Значение $\Gamma^{(r)}$	Время решения задачи МПП, с	Число итераций ЭА	Значение $\Gamma^{(n)}$	Время решения задачи ЭА, с
--	-----------------------------	--------------------	-------------------------	-----------------------------	-------------------	-------------------------	----------------------------

3x4x5	5,0956	81	1,4501	6,6	1	0,3379	1,00
6x8x10	1,0773	6 ⁸	0,4511	216	4	0,0658	11,35
8x13x15	9,2486	8 ¹³	-	-	7	1,5490	93,6
10x15x20	21,0437	10 ¹⁵	-	-	8	6,4048	1106,4

Таблица №4

Размерность задачи nxdxq	Начальное значение T	Число итераций МПП	Значение $T^{(r)}$	Время решения задачи МПП, с	Число итераций ЭА	Значение $T^{(a)}$	Время решения задачи ЭА, с
3x4x5	7,7018	81	1,8644	5,7	1	1,4576	1,00
6x8x10	2,3895	6 ⁸	1,6991	216	3	1,6298	16,35
8x13x15	12,4534	8 ¹³	-	-	6	7,5209	178,6
10x15x20	25,0437	10 ¹⁵	-	-	12	11,4417	1158,0

Здесь МПП –метод полного перебора; ЭА – эвристический алгоритм; $T^{(r)}$ – среднее время реакции системы для оптимального размещения РБД, полученного МПП; $T^{(a)}$ – среднее время реакции системы для оптимального размещения РБД, полученного ЭА.

Анализ полученных результатов работы алгоритма показал, что наличие в модели блокировки на уровне всей базы данных для небольших размерностей задачи размещения информационных ресурсов не оказывает существенного влияния на реактивность работы всей системы в целом, однако, с ростом размерности задачи, влияние времени блокировки на её реактивность становится более значительным и данную величину нужно учитывать при конструировании реальных вычислительных систем.

Литература

- 1.Бойко В.В., Савенков В.М. Проектирование баз данных информационных систем.М.: Финансы и статистика, 1989. 351с.
 - 2.Кузнецов Н.А., Кульба В.В., Косяченко С.А. Методы анализа и синтеза модульных информационно-управляющих систем. М.: ФИЗМАТЛИТ, 2002. 880с.
 - 3.Цегелик Г.Г. Системы распределённых баз данных. Львов: СВИТ,1990. 167с.
 - 4.Скоба А.Н., Состина Е.В. Математическая модель оптимального размещения распределенной базы данных по узлам ЛВС на базе файл-серверной архитектуры. // Инженерный вестник Дона. 2015. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2881.
 - 5.Скоба А.Н., Логанчук М.Л. Математическая модель функционирования распределённой информационной системы на базе архитектуры «файл-сервер» с учётом влияния блокировок // Инженерный вестник Дона.2015. №3. URL:ivdon.ru/ru/magazine/archive/n3y2015/3276.
 - 6.Скоба А.Н., Состина Е.В. Математическая модель оптимального размещения распределенной базы данных по узлам ЛВС на базе двухуровневой клиент-серверной архитектуры. // Инженерный вестник Дона. 2015. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2882.
 - 7.Скоба А.Н., Панфилов А.Н. Модель оптимального размещения информационных ресурсов по узлам распределенной информационной системы предприятия на базе двухуровневой архитектуры “клиент-сервер” с учетом влияния блокировок // Изв. вузов. Электромеханика. 2017. Т. 60, № 2. С. 77-84.
 8. Antunes C.H. et al. A Multiple Objective Routing Algorithm for Integrated Communication Network // Proc. ITC-16.-1999.V.3b. pp.1291-1300.
-



9. Chakka R., Harrison P.G. A Markov modulated multi-server queue with negative customers. The MM CPP/GE/c/LG-queue // Acta Informatika/-2001.-v.37. pp.785-799.
10. Круглый З.Л. Алгоритмы расчёта моделей структур вычислительных систем с различными классами заданий // Управляющие системы и машины. 1989. №4. С.22-24.

References

1. Bojko V.V. Proektirovanie baz danyh informacionnyh sistem [Design of bases of these information systems]. Moscow, Finansy i statistika, 1989, 351p.
2. Kuznecov N.A., Kul'ba V.V., Kosjachenko S.A. Metody analiza i sinteza modul'nyh informacionno-upravljajushhih sistem [Methods of the analysis and synthesis of modular management information systems]. Moscow, FIZMATLIT Publ., 2002, 880p.
3. Cegelik G.G. Sistemy raspredel'jonnyh baz dannyh [Systems of the distributed database]. Lviv, SVIT publ., 1990, 167p.
4. Skoba A.N., Sostina E.V. Inženernyj vestnik Dona (Rus), 2015, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2881.
5. Skoba A.N., Loganchuk M.L. Inženernyj vestnik Dona (Rus), 2015. №3. URL: ivdon.ru/ru/magazine/archive/n3y2015/3276.
6. Skoba A.N., Sostina E.V. Inženernyj vestnik Dona (Rus), 2015. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2882.
7. Skoba A.N., Panfilov A.N. Izv. vuzov. Jelektromehanika. 2017. T. 60, № 2. pp. 77-84.
8. Antunes C.H. et al. A Multiple Objective Routing Algorithm for Integrated Communication Network. Proc. ITC-16. 1999.V.3b. pp.1291-1300.



9. Chakka R., Harrison P.G. A Markov modulated multi-server queue with negative customers. The MM CPP/GE/c/LG-queue. Acta Informatics. 2001. v.37. pp.785-799.
10. Kruglyj Z.L. Upravljajushhie sistemy i mashiny. 1980. №4. pp.22-24.