

УДК 004.021

Разживин Н.О.

Студент магистратуры

НИЯУ МИФИ

Россия, г. Москва

Razzhivin N.O

Master's student

National Research Nuclear University МЕРФИ

Russia, Moscow

**ИССЛЕДОВАНИЕ СПОСОБА ОЦЕНКИ ВРЕМЕНИ ВЫПОЛНЕНИЯ
ЗАПРОСОВ НА ЭТАПЕ ПРОЕКТИРОВАНИЯ ХРАНИЛИЩА
ДАННЫХ**

**INVESTIGATION OF THE METHOD EVALUATING TIME
EXECUTION OF REQUESTS IN THE INITIAL DESIGNING OF THE
DATA STORAGE**

Аннотация:

В статье рассматривается вычисление производительности информационной системы. Как частный случай информационной системы, приводится вычисление времени выполнения запросов в хранилище данных, используя ориентированные графы. Рассматривается возможность применения графовой модели для оценки времени выполнения запроса. Описывается различие между оценкой времени выполнения программы и оценкой выполнения запроса в хранилище данных. Приводится алгоритм оценки времени выполнения исследуемого запроса методом исключения преобразования графов.

Ключевые слова: ИС, СУБД, хранилище данных, графовая модель, условная вероятность, дисперсия.

Abstract:

The paper deals with computing the performance of an information system. As a special case of the information system, the calculation of the execution time of queries in the data warehouse is performed using oriented graphs. The possibility of applying a graph model for estimating the execution time of a query is considered. A distinction between the estimation of the execution time of the program and the evaluation of the execution of the query in the data warehouse is made. An algorithm for estimating the execution time of the query is investigated by the exclusion graph method.

Keywords: IS, DBMS, data warehouse, graph model, conditional probability, variance.

Введение

В наше время практически во всех сферах жизни человеческого общества применяются те или иные системы обработки данных [1]. Одним из наиболее важных элементов информационных систем (ИС) является хранилище данных. При разработке любой ИС один из наиболее важных вопросов – ее производительность. Система не только должна успешно решать установленные перед ней задачи, она должна делать это с приемлемым для пользователя быстродействием. Интерес к эффективности применения вычислительных систем вполне закономерен из-за значительного числа сфер применения информационных технологий. Под оценкой качества вычислительной системы понимается получение количественных оценок, дающих представление о том, насколько хорошо система справляется с поставленными задачами [2]. Необходимость в таких оценках существует на всех этапах жизненного цикла вычислительного комплекса, начиная от начальных стадий его проектирования и вплоть до реализации и повседневной эксплуатации.

Оценка времени выполнения запросов в информационном хранилище

При создании вычислительных систем одним из ключевых факторов является оценка производительности программ, осуществляющих обработку информации.

Основной целью при оценке производительности программы является уменьшение времени таким образом, чтобы время выполнения программы не превышало заданное “директивное” время. В качестве моделей программ обычно используются ориентированные графы с конечным количеством дуг и вершин, с которыми сопоставлены соответствующие параметры, отражающие ход выполнения программы. Конфигурация графа в большинстве случаев практически совпадает с конфигурацией блок-схемы моделируемой программы [3].

Как частный случай оценки производительности программ допустимо рассмотреть оценку производительности при работе с хранилищами данных. Система управления базой данных (СУБД) может представляться как программа, работающая с информацией, и поэтому допустимо говорить об изучении ее производительности. Если в случае работы с обычной программой критерием оценки было выбрано время выполнения, то в случае с хранилищами данных таким критерием будет время обработки системой поступившего от пользователя запроса.

Для разработки методики оценки производительности хранилищ данных необходимо рассмотреть применение данного метода в новой для него области - для оценки времени выполнения запроса.

Пусть хранилище - это класс сущностей E_i ($i = 1, 2, \dots, N$). Запрос пользователя $Q = Q [E_i]$ можно представить, как совокупность подзапросов вида

$$q(i) = a(E_j), (i, j = 1, 2, \dots, N) \quad (1)$$

Время обслуживания подзапроса $q(i)$ обозначается как $t(q(i))$. Общее время обработки запроса определяется по формуле:

$$T[Q] = \sum_{i=1}^N t(q(i)) \pm \Delta t(i), \quad (2)$$

где $\Delta t(i)$ среднее квадратическое отклонение времени выполнения каждого подзапроса, так как время выполнения каждого этапа обработки запроса может раз от раза сильно меняться. Например, при осуществлении поиска, один раз введенным критериям может соответствовать немного значений, а в другой раз может получиться достаточно большое количество.

Время обработки запроса пользователя $T[Q]$ будет считаться главным критерием производительности при работе с информационным хранилищем.

Исходя из того, что обработка запроса в хранилище данных аналогична выполнению обычной программы, можно заключить, что графовые модели допустимо использовать для моделирования процесса обслуживания запроса. В соответствии с (1) делается заключение, что обслуживание практически любого запроса может быть разбито на некое конечное число этапов.

Пусть исследуемый запрос представлен в виде графовой модели. Для исследования операционных характеристик его выполнения могут быть применены методы исключаящего преобразования графов.

Одним из таких методов является метод исключаящих преобразований. Данный метод был впервые предложен Б. Байцером в [3] для оценки времени выполнения программ. Модель исследуемой системы представляется в виде ориентированного графа, в котором каждая дуга описывается тройкой

$$(P, U, D), \text{ где} \quad (3)$$

P - условная вероятность передачи управления по этой дуге;

U - среднее время выполнения дуги;

D - дисперсия этого времени;

Указанный граф строится таким образом, чтобы его дуги представляли собой операторы программы, а вершины связи между операторами. Вся

обработка в такой модели осуществляется в дугах, и каждая дуга характеризуется тройкой (3). Предполагается, что для выполнения программы среднее время выполнения дуги известно, а дисперсия это времени равна нулю. Такое предположение о дисперсии может быть сделано потому, что при исполнении машинных команд изменчивость времени их выполнения столь мала, что ею можно пренебречь.

В случае с хранилищами данных дисперсия времени выполнения каждой дуги графа не равна 0. В этом состоит основное условие применимости метода исключаящих преобразований для работы с хранилищами данных — если при исследовании программы дисперсия времени выполнения каждого ее этапа может считаться равной 0, то при обработке запросов данное предположение неприемлемо.

Оценка времени выполнения исследуемого запроса осуществляется путем исключаящих преобразований:

1. Выбирается вершину графа, которая не является ни входом, ни выходом, для ее исключения;
2. Эта вершина разделяется на копии так, чтобы ее выходящие дуги были последовательно связаны с входящими дугами;
3. Для исключения новых вершин применяются уравнения последовательного преобразования ($a \rightarrow b \rightarrow c \Rightarrow a \rightarrow c$):

$$P'_{ac} = P_{ab}P_{bc} \quad (4)$$

$$U'_{ac} = U_{ab}U_{bc} \quad (5)$$

$$D'_{ac} = D_{ab} + D_{bc} \quad (6)$$

4. Если в графе появились параллельные связи, то они исключаются с помощью уравнений параллельного преобразования ($a \overset{\circ}{\rightarrow} b \Rightarrow a \rightarrow b$):

$$P'_{ab} = P^{\circ}_{ab} + P^{\circ\circ}_{ab} \quad (7)$$

$$U'_{ab} = (P^{\circ}_{ab}U^{\circ}_{ab} + P^{\circ\circ}_{ab}U^{\circ\circ}_{ab})/P^{\circ}_{ab} + P^{\circ\circ}_{ab} \quad (8)$$

$$D'_{ab} = (P^{\circ}_{ab}D^{\circ}_{ab} + P^{\circ\circ}_{ab}D^{\circ\circ}_{ab})/(P^{\circ}_{ab}+P^{\circ\circ}_{ab}) + [(P^{\circ}_{ab}U^{\circ}_{ab}{}^2 + P^{\circ\circ}_{ab}U^{\circ\circ}_{ab}{}^2)/P^{\circ}_{ab}+P^{\circ\circ}_{ab}] - U'^2_{ab} \quad (9)$$

5. Если появились циклы, их исключение производится с помощью уравнений преобразования циклов ($a \rightarrow b \Rightarrow a \rightarrow b$):

$$P'_{ab} = P_{ab}/(1 - P_{aa}) \quad (10)$$

$$U'_{ab} = U_{ab} + P_{aa}U_{aa}/(1-P_{aa}) \quad (11)$$

$$D'_{ab} = D_{ab} + [D_{aa}P_{aa}/(1-P_{aa})] + P_{aa}U^2_{aa}/(1-P_{aa})^2 \quad (12)$$

Если дальнейшие исключения невозможны, происходит возврат на пункт 1. Исключающие преобразования необходимо повторять до тех пор, пока весь граф не сведется только к двум вершинам - входу и выходу. Эти две вершины будут соединяться единственной дугой, которая будет характеризоваться той же самой тройкой (P, U, D). Среднее время выполнения запроса U и является искомым критерием оценки производительности. Изменяя вероятности передачи управления по определенным дугам и повторяя анализ, можно определить распределение вероятностей общего времени выполнения запроса в хранилище данных.

После всех проделанных шагов граф будет состоять только из двух вершин - входа и выхода. Среднее значение времени выполнения всего запроса может быть вычислено путем подстановки в окончательную формулу всех промежуточных результатов, полученных на разных этапах преобразования графа.

Заключение

Оценка эффективности функционирования является одной из важнейших задач при создании любой вычислительной системы. Исследование производительности хранилищ данных производится аналогично исследованию производительности программ. Основным

критерием производительности в таком случае является время обработки запроса пользователя. Процесс обработки запроса представляется в виде графовой модели аналогично процессу выполнения программы. Определение динамических характеристик обработки запросов может осуществляться с использованием методов упрощающих преобразований графов, известных из математической теории марковских процессов. Наиболее удобный из таких методов — метод исключаящих преобразований, который может использоваться для оценки производительности как в случае исследования обычных программ, так и в случае моделирования обслуживания запросов в хранилищах данных на начальном этапе проектирования хранилища данных.

Список использованных источников

1. Дмитриев В. Sybase Enterprise Application Server: новый подход к построению корпоративных информационных систем. URL: http://www.sybase.ru/Syb/products/appservers/eas_article.htm (дата обращения 25.04.2018)
2. Липаев В.В. Качество программного обеспечения. - М.: Финансы и статистика, 1983. - С. 263.
3. Байцер Б. Архитектура вычислительных комплексов. - М.: Мир, 1974. - 498 с.
4. О.И. Авен, Н. Я . Турин, А. Я. Коган. Оценка качества и оптимизации вычислительных систем - М.: Наука, 1982. – 464 с.