

УДК 519.6

СИНТЕЗ ОПТИМАЛЬНЫХ ЛОГИЧЕСКИХ СТРУКТУР РАСПРЕДЕЛЕННЫХ БАНКОВ ДАННЫХ В ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СЕТЯХ

Кухта С.В., Косяк Л.Н., Глебка А.Н.

Полоцкий государственный университет

Создание автоматизированных систем управления и информационно-вычислительных систем (ИВС) связано с широким внедрением сетей ЭВМ, распределенных банков данных и систем передачи информации.

Распределенный банк данных представляет собой систему логически интегрированных и территориально распределенных баз данных, языковых, программных, технических и организационных средств, предназначенных для создания, ведения и обработки распределенной информации. Для распределенных банков данных (РБнД) характерным является накопление, обновление и хранение данных в географически распределенных узлах ИВС, логическая интеграция территориально-распределенных данных, процессов обработки, обновления и поиска информации, обеспечение автоматического взаимодействия между локальными базами данных в процессе решения задач пользователей.

Синтез оптимальной логической структуры РБнД является одной из центральных проблем разработки ИВС. Решаемые на этом этапе задачи отличаются большой размерностью, сложностью и трудоемкостью, что определяется следующими факторами:

- распределенностью процессов хранения и обработки данных;
- необходимостью учета большого количества характеристик используемой информации и процедур ее обработки;
- многообразием критериев эффективности;

7. Технология создания информационных систем

- ограничениями систем управления РБнД, СУБД, операционных систем и технических средств.

Содержательная постановка задачи синтеза оптимальной логической структуры РБнД формулируется следующим образом: по известным характеристикам $P_U, P_Q, P_\Omega, P_3, P_K$ множеств пользователей РБнД, узлов ИВС, групповых информационных элементов, детерминированных запросов и заданий на корректировку необходимо определить логическую структуру РБнД в виде множеств типовых логических записей $H = \{ h_i \}$ и отношений между ними $A^{3П} = [a_{ij}^{3П}]$, а также размещение типов записей в ИВС $\Psi^{3П} = [\psi_{ij}^{3П}]$, которые обеспечивают оптимальное значение заданного критерия эффективности функционирования РБнД и выполнение сетевых, системных и структурных ограничений, т.е. выполнить преобразование

$$IM \langle P_U, P_Q, P_\Omega, P_3, P_K \rangle \xrightarrow{\mathcal{E}^{N\Omega} (\Phi \longrightarrow opt)} N\Omega \langle P_{DDV} \rangle,$$

где $\mathcal{E}^{N\Omega} (\Phi \longrightarrow opt)$ - оператор синтеза, обеспечивающий оптимум заданной целевой функции;

P_{DDV} - множество характеристик оптимальной логической структуры РБнД.

Формальная постановка задачи синтеза оптимальных логических структур РБнД не приводится в виду ее большого объема.

Результаты решения задачи синтеза позволяют определить состав, структуру и характеристики типов логических записей, отношения между ними, размещение типов записей в ИВС и использование их процедурами обработки данных.

В качестве системных, сетевых и структурных ограничений в задачах логического синтеза используется:

- стоимость обработки множества запросов и корректировок;
- заданное время выполнения множества запросов, корректировок;
- длина формируемой логической записи;

- количество типов формируемых записей;
- степеней дублирования формируемых записей при размещении в ИВС;
- пропускная способность каналов связи и другие.

Рассмотрим приближенный алгоритм решения задачи синтеза оптимальной логической структуры РБнД, обеспечивающей минимум суммарного времени выполнения множества корректировок записей. Для решения задачи большой размерности разработан приближенный алгоритм, позволяющий на основе ряда упрощающих допущений выполнить декомпозицию сложной задачи проектирования РБнД на множество более простых, последовательно решаемых задач оптимизации, для решения которых адаптированы известные алгоритмы. Алгоритм решения задачи состоит из следующей последовательности взаимосвязанных этапов:

Этап 1. Определение оптимального места размещения в ИВС деревьев корректировки, обеспечивающего минимум трафика в ИВС. Под деревом корректировки понимается подграф графа канонической структуры, в терминах которого пользователь формирует задание на корректировку и обновление.

Этап 2. Определение оптимального места размещения в ИВС избыточного множества групповых информационных элементов.

Этап 3. Синтез типов логических записей, обеспечивающий минимум общего времени локальной обработки задания на корректировку в узле ИВС.

Этап 4. Проверка ограничений на задачу. Если ограничения выполняются, то конец алгоритма. В противном случае выполнить классификацию типов записей. Разместить дубликаты типов записей в ИВС в пределах заданных ограничений.

Решаемая на первом этапе задача эквивалентна общей задаче о паросочетаниях в двудольном графе, минимизационный вариант которой известен как задача о назначениях или транспортная задача. Ее решение основано на применении модифицированного венгерского алгоритма.

7. Технология создания информационных систем

Решение задачи этапа 2 осуществляется методом предварительной ранжировки дублируемых групповых информационных элементов по весовым коэффициентам и определения с учетом результатов ранжировки оптимального размещения. В качестве оценки используется величина суммарного трафика, получаемого при корректировке дублируемых групповых информационных элементов.

Задача этапа 3 эквивалентна задаче разбиения графа канонической структуры на множество подграфов (типов записей), которые удовлетворяют ряду условий и обеспечивают оптимальное значение критерия разбиения - максимум внутренних связей между групповыми информационными элементами, входящими в один тип записей.

Результатом выполнения этапов 1-3 является множество типов записей, распределенных по узлам ИВС, без учета ограничения на стоимость передачи информации при реализации множества запросов и ограничения на оперативность получения заданной информации по запросу k -го пользователя. Алгоритм задачи, решаемой на этапе 4, заключается в последовательном размещении в ИВС дубликатов типов записей в пределах данных ограничений.

Предлагаемый алгоритм использовался при проектировании информационного обеспечения корпоративной информационно-вычислительной сети с распределенной обработкой данных Белорусского производственно-технического управления связи. Практическое применение приближенного алгоритма при решении задач средней размерности (количество групп порядка 70, количество узлов - 6) дало результаты, отличающиеся на 10-15% от результатов, полученных точными методами.