

На правах рукописи

Ндикумагенге Жереми

МЕТОД И МОДЕЛИ УПРАВЛЕНИЯ ТРАНЗАКЦИЯМИ В
РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ОБРАБОТКИ ИНФОРМАЦИИ

Специальность 05.13.01 - Системный анализ, управление и
обработка информации (технические системы)

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург-2010

Работа выполнена в Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В.И.Ульянова (Ленина).

Научный руководитель - кандидат технических наук, доцент
Цехановский Владислав Владимирович

Официальные оппоненты: доктор технических наук, профессор
Водяхо Александр Иванович

кандидат технических наук, доцент
Воронин Иван Викторович

Ведущая организация: Санкт-Петербургский государственный университет
аэрокосмического приборостроения

Защита диссертации состоится 28 декабря 2010 г. в 12 часов на заседании совета по защите докторских и кандидатских диссертаций Д 212.238.07 Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И.Ульянова (Ленина) по адресу: 197376, г. Санкт-Петербург, ул. Проф.Попова, 5.

С диссертацией можно ознакомиться в библиотеке университета.

Автореферат разослан 26 ноября 2010 г.

Ученый секретарь совета по защите докторских и кандидатских диссертаций Д212.238.07

кандидат технических наук, доцент

Цехановский В. В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Системы распределенных баз данных, как основной компонент распределенной системы обработки информации играют важную роль в развитии современной экономики и бизнеса. Обеспечение интенсивной динамичности выполнения процессов таких систем - одна из самых трудоемких проблем эффективной реализации бизнес процессов при эксплуатации банков данных. Создание надежных, отказоустойчивых и эффективных средств обслуживания и управления требует наличия высококвалифицированных специалистов, больших финансовых и временных затрат, как для проектирования, разработки и развертывания, так и для сопровождения, эксплуатации и администрирования. Управление выполнением распределенных и параллельных вычислительных процессов, определенных на стадии детальной разработки средств сопровождения и эксплуатации распределенной системы обработки информации в целом и их оптимизация в частности, всегда было и остается сложной и актуальной задачей.

Качественная оценка свойств распределенной системы обработки информации и прогнозирование динамического поведения вычислительных процессов, бизнес-процессов и потоков данных на этапах детальной разработки логики обслуживания и эксплуатации с учетом, по возможности их физической реализации, существенно и значительно уменьшает вероятность, долю или частоту сбоев и, соответственно, банкротство информационных и финансово-банковских инфраструктур. Полученные навыки, изобретения и достижения, нуждаются не только в усовершенствовании, но и во внедрении и создании новых методов и моделей, то есть новых технологий разработки средств сопровождения, администрирования, управления и исследования. Своевременное обнаружение, фиксация и устранение конфликтных ситуаций и противоречий данных, способных выводить системы из строя с наименьшими

финансовыми и временными затратами является одной из важных характеристик управляемости и полезности системы распределенных баз данных разного рода и назначения. Проблемы обслуживания и обеспечения работоспособности распределенных систем обработки информации (РСОИ), основанных на концепции распределенной системы баз данных (РБД), являются в настоящее время объектом все возрастающего интереса широкого круга специалистов в области обработки данных.

В связи с этим важным является создание и разработка методов и моделей управления выполнением транзакций, позволяющих системным администраторам отслеживать динамику обработки процессов, контролировать работоспособность системы и устранять сбои, которые могут возникать. Реализация процедур обслуживания должна предоставлять возможность конфигурации системы по нуждам и требованиям пользователей за счёт составления и генерации алгоритмов (протоколов) выполнения действий для решения конкретных задач в пределах поставленных целей. Поэтому, данное диссертационное исследование посвящено решению задачи разработки методов и моделей повышения управляемости обслуживанием и выполнением транзакций в РСОИ, на базе расширения результатов исследований современных ученых, таких как Кульба В.В., Ковалевский С.С., Косяченко С.А., Советов Б.Я., Чертовской В.Д., П.Чардин., G.Weikum., G. Vossen., P.A. Bernstein и др.

Объектом исследования является система распределенных баз данных, как основной компонент распределенной системе обработки данных.

Предметом исследования является средства обработки и обслуживания транзакций распределенных баз данных, предназначенные для повышения управляемости выполнением процессов в распределенной системы обработки информации.

Цель диссертационной работы заключается в разработке метода, алгоритмов и моделей управления выполнением транзакций, обеспечивающих анализ, обработку, исследование, прогнозирование

поведения и обслуживание транзакций РСОИ на базе аппарата сетей Петри и примитивов управления транзакциями СУБД.

Задачи исследования.

1. Анализ структурных характеристик распределенных систем обработки информации и проблем управления выполнением транзакций в распределенных базах данных.
2. Разработка метода и модели выявления конфликтов в распределенной базе данных.
3. Разработка методов и моделей улучшения управляемости выполнения транзакций в распределенных системах обработки информации.
4. Разработка методов и моделей исследования множества транзакций

Методы исследования.

При решении основной задачи диссертационной работы использовались методы балансировки древовидных структур, конечные автоматы; функции управления объектами СУБД SQL Server 2000; элементы теории графов сетей Петри с их методами анализа и исследования.

Основные положения, выносимые на защиту:

1. Метод обнаружения потенциальных конфликтных ситуаций по функциональной структуре РБД в РСОИ.
2. Модель выбора допустимых путей выполнения транзакций на базе основных типов операций над данными для разрешения несогласованности реализаций транзакций.
3. Модели исследования множества транзакций, вовлеченных в процесс обработки и улучшения управляемости выполнения множества транзакций.

Научная новизна работы.

1. Разработанный метод обнаружения конфликтных ситуаций с учетом структуры РБД на основе методологии имитационного моделирования, отличается от существующих тем, что позволяет априори выявить и предсказать поведение транзакций.

2. Разработанная модель выбора допустимых путей выполнения транзакций отличается от известных использованием методов и математических моделей аппарата сетей Петри.
3. Предложенные модели исследования множества транзакций и улучшения управляемости выполнения множества транзакций, отличаются от известных использованием математической теории отношений над бинарным множеством примитивов операций.

Достоверность научных результатов и выводов, полученных автором диссертации, подтверждается корректностью использования математических аппаратов теории сетей Петри и графов, теории конечных автоматов, теории систем массового обслуживания, функций управления объектами СУРБД.

Научная и практическая ценность диссертационной работы

Научная значимость заключается:

- в разработке и обосновании теоретической основы прогнозирования функционального поведения вычислительных процессов РСОИ за счёт использования современного математического аппарата;
- в развитии теории слияния и декомпозиции транзакций в виде деревьев по слоевой архитектуре серверов транзакций и теоретических методов оптимизации исследования транзакций, вовлеченных в обработку в транзакционных серверах.

Практическая ценность работы заключается в том, что результаты экспериментальных и теоретических исследований существования противоречий хранимых данных, функциональных конфликтов и обслуживания транзакций и запросов в распределенных системах обработки информации могут быть использованы для:

- своевременного предсказания системными администраторами конфликтных ситуаций при эксплуатации и обслуживании системы;

- декомпозиции, разбиения, слияния и отслеживания транзакций с целью оптимизации их выполнения и обработки;
- многоуровневого отображения сложных транзакций с целью упрощения их функциональной и структурной сложности для детального анализа и исследования;
- обслуживания множества транзакций и запросов в РБД.

Теоретические и практические результаты исследований внедрены в курсе моделирования процессов систем реального времени и реализованы в виде ряда программных схем на различных продуктах инструментальных средств разработки, таких как, Visual Paradigm. Данные программные схемы используются для развертывания информационных управленческих систем и в качестве модулей системы анализа и исследования в системах поддержки принятия решений.

Апробация результатов работы. Основные положения и результаты докладывались и обсуждались на:

- Ежегодных научно-технических конференциях профессорско-преподавательского состава СПбГЭТУ «ЛЭТИ (2008- 2010 гг);
- XV и XVI Международных научно-методических конференциях «Современные технологии обучения» (2009- 2010 гг).

Публикации: По теме диссертации опубликовано 4 научных работы, из них - 2 статьи, среди которых 1 публикация в ведущих рецензируемых изданиях, рекомендованных в действующем перечне ВАК, 2 работы в материалах международных и всероссийских научно-технических конференций.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав с выводами и заключения. Она изложена на 131 страницах машинописного текста, включает 37 рисунков, 4 таблицы и содержит список литератур из 67 наименований, среди которых 51 отечественных и 16 иностранных авторов.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы, приведены и выделены текущие положения и роль средств управления вычислительными процессами баз данных, поставлена цель исследований, сформулированы и обоснованы задачи исследования. Отмечена научная новизна, теоретическая и практическая ценность, приведены сведения об апробации работы.

В первой главе проведен обзор исследуемой предметной области. Приводится описание и анализ структуры распределенной системы обработки информации и распределенной базы данных, как основного компонента распределенной обработки информации. Приведены основные характеристики распределенных систем обработки информации и принципы РБД, сформулированные К. Дж. Дейтом. Описаны существующие методы, алгоритмы и модели управления выполнением и обработкой распределенных и параллельных транзакций, СУБД и СУРБД, выделены их достоинства и недостатки, исследованы средства СУРБД управления объектами.

Во второй главе исследован аппарат сетей Петри, методы анализа, которые позволяют создать методы и алгоритмы выявления и предсказания конфликтов по архитектурной структуре РСОИ и разработаны алгоритмы построения и анализа дерева достижимости сетей Петри, диаграммы, которые приведены на рисунках 1.а, б.

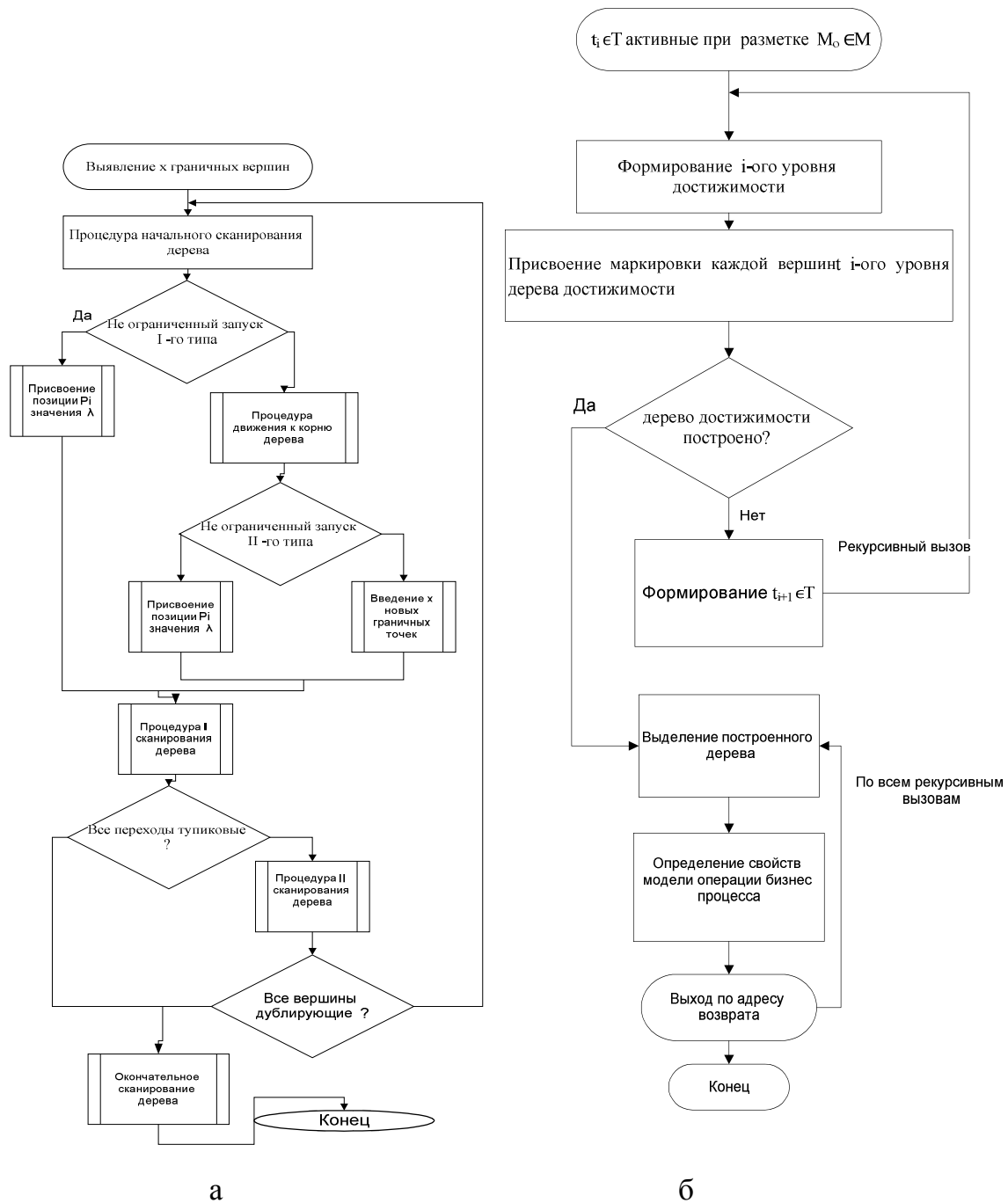


Рисунок 1.а. Итеративный алгоритм построения дерева достижимости,

б. Рекурсивный алгоритм построения дерево достижимости

В третьей главе разработаны модели реализации запросов и транзакций, на базе основных типов операций, предложен метод анализа процессов обслуживания и выполнения транзакций и запросов на основе дерева достижимости сетей Петри и выявлено необходимое условие завершения выполнения запросов и транзакций, при котором

исключаются системные тупики и конфликты при использовании разделяемых ресурсов.

Для завершения транзакции, представленной на рисунке 2, требуются операции последовательной выборки из БД записей R_1^{DB} (операция f_1), R_2^{DB} (операция f_2) и R_3^{DB} (операция f_3). Результатом операции f_1 является формирование в рабочей области памяти транзакции искомым записей R_1^{OP} , результатом операции f_2 - формирование искомым записей R_2^{OP} , с учетом результатов операции f_1 (т.е. R_1^{OP}) и R_2^{DB} , результатом операции f_3 - формирование искомым записей R_{RE}^{OP} , с учетом результатов операции f_2 (т.е. R_2^{OP}) и R_3^{DB} . Сами операции выборки при этом включают определенную последовательность процедур и операторов языка манипулирования данными СУРБД (рис.3).

Аналогичным образом, путем этапного последовательного подхода представления операции, начиная с модели операции в виде двухдольного графа, затем, построения по ней модели сетей Петри и, наконец, последовательной композиции взаимосвязанных фрагментов обработки, подвергающей только одну запись для реализации несколько операций того или иного типа, создаются остальные модели основных типов операций (вставки, обновления и удаления), выполняемых над БД. Таким образом, рассмотренные модели типов операции над БД в виде фрагментов сетей Петри являются основой для формирования моделей транзакций и запросов.

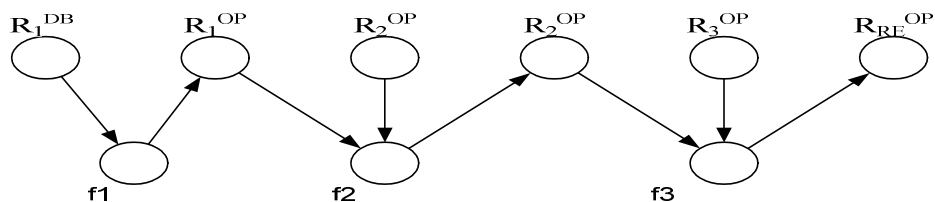


Рисунок 2. Модель в виде двухдольного графа основных типов операций, не изменяющих состав или структуру БД

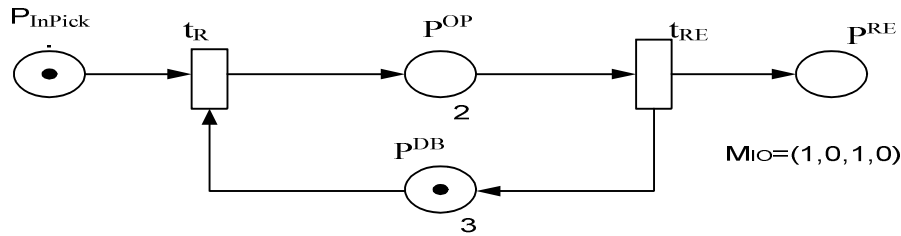


Рисунок 3. Модель операции выборки в виде маркированной сети Петри

Исследование динамики выполнения транзакций реализуется путем перехода представления операций в виде двудольных графа к представлению в виде моделей сетей Петри. На рисунке 4 представлена модель последовательной реализации несколько операций выборки в виде маркированной сетей Петри.

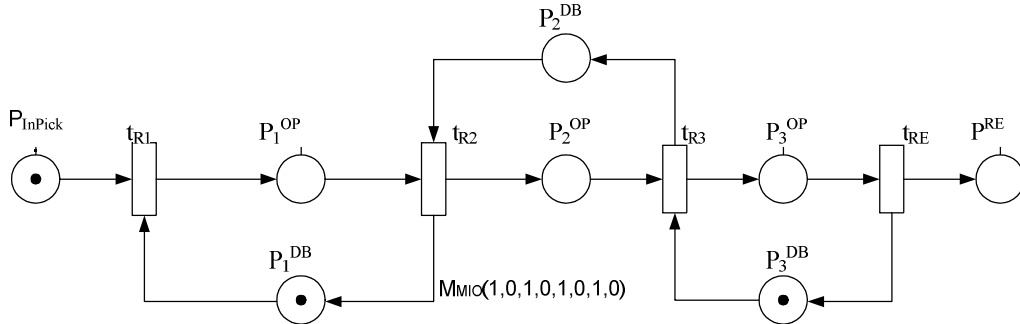


Рисунок 4. Модель последовательной реализации операции выборки в виде сети Петри.

Выявлением и устранением системных тупиков и конфликтов при использовании разделенных информационных ресурсов БД осуществляются методом анализа и исследования дерева достижимости сетей Петри.

Пусть $G = \langle V, T \rangle$ – граф дерева достижимости для заданной сети Петри S , моделирующей процесс обслуживания множества транзакций; где $V = \{V_i; i = \overline{1, I}\}$ – множество вершин дерева, V_i – вершина i -ой маркировки сети; $T = \{t_j; j = \overline{1, J}\}$ – множество дуг, помеченных активными в данный момент времени переходами t_k . Граф $G = \langle V, T \rangle$ описывается следующими известными характеристиками: $F = \{f_n; n = \overline{1, N}\}$, – множество путей, ведущих из корневой вершины V_0 в терминальные вершины V_i , где f_n – n -й путь, состоящий из последовательности дуг $\{t_j, t_k, , t_l, \dots, t_r\}$ активных переходов;

$L = \{l_n; n = \overline{1, N}\}$ – множество длин путей, определяемых суммированием количества входящих в них дуг, где l_n – длина n -ого пути. Каждая транзакция, моделируемая в виде сети Петри, описывается и определяется следующими характеристиками: γ_k – число переходов k -ой транзакции; ω_k – частота реализации k -ой транзакции. Тогда, необходимым условием завершенности выполнения множества транзакций (запросов), при котором исключаются системные тупики и конфликты в использовании ими разделяемых ресурсов РСОИ, является наличие на графе $G = \langle V, T \rangle$ таких путей f_n , длины которых равны величине:

$$l_n = \sum \gamma_k \cdot \omega_k, \quad \forall f_n \in F, \text{ где } k\text{-индекс транзакции.}$$

Данное утверждение вытекает из свойства ω - безопасности моделей множества транзакций(запроса). Из данного свойства следует, что при последовательной реализации переходов отдельного k -ой транзакции, переход $t_j \in T_k \subseteq T$ будет срабатывать ровно ω раз, т.е. $l_n = \omega_k \cdot \gamma_k$. Следовательно, при последовательной реализации K транзакций (запросов) условием завершенности их выполнения есть наличие на графе достижимости $G = \langle V, T \rangle$ путей длины:

$$l_n = \omega_1 \cdot \gamma_1 + \omega_2 \cdot \gamma_2 + \dots + \dots + \omega_{k-1} \cdot \gamma_{k-1} + \omega_k \cdot \gamma_k = \sum \omega_k \cdot \gamma_k.$$

Сформулированное утверждение может быть использовано для выявления и устранения непротиворечивости и несогласованности хранимых данных, за счёт исследования моделей смен маркировок для каждой основной операции.

В четвертой главе приведена стратегия описания структуры построения моделей транзакционных серверов, упрощающих доступ к страницам данных, за счёт абстрагирования базовых операций над данными, на основе концепции транзакции. Разработаны две

иерархически зависимые модели транзакционных серверов обслуживания и исследования транзакций.

Стратегия описания структуры и онтологического построения моделей транзакционных серверов включает:

1. Определение элементарных операций являющихся неделимыми, атомарными и изолированными от других операций, выполняемых над данными;

2. Формирование последовательности элементарных операций над данными, обеспечивающих свойства атомарности, согласованности, изолированности и долговечности;

3. Образование расписаний транзакций, как абстрактного понятия параллельного выполнения набора частичного порядка элементарных операций, за счёт формирования ряда познаваемых индивидуальных транзакций;

4. Выбор из синтаксических корректных расписаний тех, которые удовлетворяют атомарности, согласованности, изолированности и долговечности свойства;

5. Разработка протоколов создания корректных расписаний, которые по вызовам прикладных программ запускают и выполняют операции динамически.

Сервер данных представим в виде конечного множества $D = \{x, y, z, \dots\}$ неделимых и идентифицируемых элементов с неделимыми операциями чтения и записи. Отношение полного порядка определяется моделью страницы транзакции в виде конечной последовательности действий типа $r(x)$ или $w(x)$: $t = p_1 p_2 \dots p_n$, где $n < \infty$; $p_i \in \{r(x), w(x)\}$; $1 \leq i \leq n$; $x \in D$; x - элемент (с индексом и без него) из D ; r : - операция чтения, а w : - операция записи. Данная формализация определяет некую семантику для отдельной транзакции, такой, что две разные транзакции имеют разные шаги но, возможные одни и те же типы шагов. Семантическая интерпретация транзакции на уровне модели страницы базируется на $r(x)$ и $w(x)$ операции.

Если $p_j = r(x)$, то j -й шаг данной транзакции является чтением, текущее значение элемента x получает значение v_j : $V_j := x$. Если $p_j = w(x)$, то j -й шаг данной транзакции является записью, текущее значение элемента x зависит от сочетания значений v_{j_r} , прочитанных до шаг j : $x := f_j(v_{j_1}, v_{j_2}, \dots, v_{j_k})$, где x : возвращаемое значение такое, что $\{v_{j_1}, v_{j_2}, \dots, v_{j_k}\} = \{j_r \mid p_{j_r}, \text{ шаг чтение и } j_r \rho j\}$. Все значения v_{j_r} , $1 \leq i \leq k$, считанные до j -го шага транзакции t используются, как параметры в функции f_j .

Определим модель страницы транзакционных серверов: $t = (\Omega, \rho)$, где Ω конечное множество шагов вида $r(x)$ или $w(x)$, $x \in D$, и $\rho \subseteq \Omega \times \Omega$ есть отношение частичного порядка на множестве Ω , такое что, если $\{p, q\} \subseteq \Omega$, где p и q имеют доступ к одному и тому же элементу данных и хотя бы одна из этих двух операций является операцией записи, то $p \rho q \vee q \rho p$.

Формализация модели объекта транзакции проводится следующем образом: транзакция t - конечное дерево помеченных вершин со следующими характеристиками :

- идентификацией транзакции, как метки корневого вершины дерева;
- именами и параметрами операций, как метки внутренних вершин;
- моделью страницы, как метки конечных вершин (листья);
- отношением частичного порядка " ρ " на множестве Ω конечных вершин, такое что, для всех конечных вершин операции p и q , где p имеет вид $w(x)$ и q имеет вид $r(x)$ или $w(x)$, или наоборот, т.е. $p \rho q \vee q \rho p$.

На рисунке 5 представлены структуры декомпозиции двух транзакций в виде дерева по модели объекта. Для изучения и исследования множественных транзакций, выполняемых параллельно, постановка задачи должна формироваться объединением вовлеченных операционных деревьев транзакций в «лес» транзакций - и просмотра частичного порядка выполнения операций конечных вершин и неявно полученного частичного порядка выполнения относительно высокоуровневых операций. Частичный

порядок выполнения операций определяется объединением конечных вершин всех вызванных деревьев.

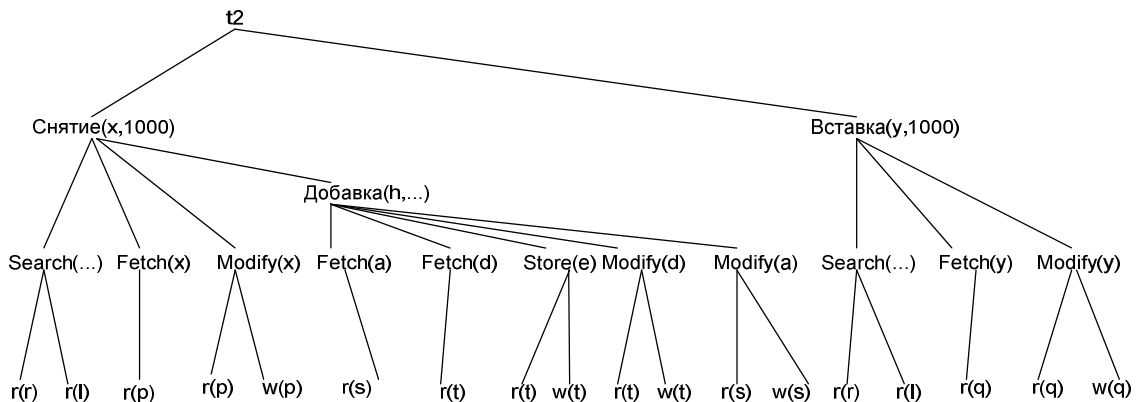


Рисунок 5. Пример слияния и декомпозиции двух транзакций по модели объекта транзакционного сервера

Абстрактное основание данных моделей может использоваться для получения конкретных алгоритмов исследования параллельной обработки сложных конкурентных транзакций и фактически жизнеспособных, ориентированных системных решений.

В заключении формулируются основные и практические результаты, выполненные в диссертационной работе исследований.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В рамках диссертационной работы проведено исследование и получено решение технической задачи разработки методов, алгоритмов и моделей улучшения управляемости выполнения транзакций в РСОИ.

Основные результаты работы:

1. Метод обнаружения потенциальных конфликтных ситуаций по функциональной структуре РБД в РСОИ.
2. Модель выбора допустимых путей выполнения транзакций на базе основных типов операций над данными для разрешения несогласованности реализаций транзакций.
3. Модели исследования множества транзакций, вовлеченных в процесс обработки и улучшения управляемости выполнения множества транзакций.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ:

Публикации в изданиях, рекомендованные ВАК России:

1. Ндикумагенге Ж. Вычислительные модели параллельных транзакционных серверов [Текст] / Ж. Ндикумагенге // Информационно-управляющие системы; № 1.(44). 2010. – С. 25-29

Другие публикации, статьи и материалы конференции:

2. Ндикумагенге Ж. Разработка методов и моделей управления транзакциями и параллельными вычислителями в распределенных системах баз данных [Текст] / Ж. Ндикумагенге // Сборник докладов, студентов, аспирантов и молодых ученых. Санкт Петербург, 27 января – 8 февраля 2009 г. 290 с.. С. 138 –143
3. Ндикумагенге Ж. Методика проектирования многопользовательских баз данных для задач учебного процесса[Текст] // Современное образование: содержание, технологии, качество. ХУ Международная конференция, Санкт-Петербург, 2 апреля 2009. – СПб, 2009 . – Том 1. - С.229 – 230
4. Ндикумагенге Ж. Метод балансировки данных в распределенных системах баз данных для задач учебного процесса [Текст] / Ж. Ндикумагенге // Современное образование: содержание, технологии, качество. ХУІ Международная конференция, Санкт-Петербург, 21-22 Апреля 2010 г. . – Том 2, - С. 208