

ДУНАЕВ ВАЛЕРИЙ АЛЕКСАНДРОВИЧ

**МОДЕЛЬ И АЛГОРИТМЫ УПРАВЛЕНИЯ ПАРАМЕТРАМИ РЕПЛИКАЦИИ
В РАСПРЕДЕЛЕННОЙ БАЗЕ ДАННЫХ ПРЕДПРИЯТИЯ
ГОРНОПРОМЫШЛЕННОГО КОМПЛЕКСА**

Специальность 05.13.06

Автоматизация и управление технологическими процессами
и производствами (промышленность)

АВТОРЕФЕРАТ
на соискание ученой степени
кандидата технических наук

Орел 2014

Работа выполнена в Государственном казённом образовательном учреждении высшего профессионального образования Академия Федеральной службы охраны Российской Федерации.

Научный руководитель: кандидат технических наук, доцент
Тараканов Олег Викторович,

Официальные оппоненты: Корсунов Николай Иванович,
доктор технических наук, профессор,
ФГАОУ ВПО «Белгородский государственный
национальный исследовательский университет»,
профессор кафедры математического и программного
обеспечения информационных систем

Лунёв Роман Алексеевич,
кандидат технических наук, доцент,
ФГБОУ ВПО «Госуниверситет - УНПК»,
доцент кафедры «Информационные системы»

Ведущая организация: ФГБУН Институт программных систем
им. А.К. Айламазяна РАН
(ИПС им. А.К. Айламазяна РАН),

Защита состоится « 24 » июня 2014 г. в 15 часов 30 минут на заседании диссертационного совета Д 212.182.01 при федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Государственный университет — учебно-научно-производственный комплекс» по адресу: 302020, РФ, г. Орел, Наугорское шоссе, д. 29., ауд. 212.

С диссертацией можно ознакомиться в библиотеке Госуниверситета – УНПК.

Материалы по защите диссертации размещены на официальном сайте Госуниверситета – УНПК по адресу: www.gu-unpk.ru

Автореферат разослан « 29 » апреля 2014 г.

Ученый секретарь
диссертационного совета Д.212.182.01
кандидат технических наук, доцент



В.Н. Волков

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Горнопромышленные комплексы (ГПК) по добыче угля являются территориально распределенными большими организационно-техническими системами. Производственный процесс угледобычи относится к классу опасных производств, что обуславливает высокие требования к системе управления ГПК. Множество служб ГПК, отличающихся по уровню и выполняемым задачам, порождает высокую интенсивность потоков информации, передаваемых между различными объектами шахты. Эффективное управление сложным объектом обеспечивается за счет применения комплекса средств автоматизации, на основе которого построена АСУ предприятием ГПК. АСУ предприятием ГПК "ШахтИнвестКузбасс" сформирована по классическим принципам и территориально распределена по службам шахты.

Повышенные требования к безопасности в угольных шахтах делают необходимым применение оперативных и достоверных систем сбора и обработки информации. Необходима организация информационного обеспечения, при которой пользователи на различных участках ГПК могли бы оперативно получать актуальную информацию для принятия управленческих решений. В системе управления ГПК "ШахтИнвестКузбасс" информационное обеспечение построено на основе распределенной базы данных (РБД), интегрирующей на уровне информации функциональные подсистемы АСУ.

Существующие технологии РБД позволяют равномерно распределить нагрузку между удаленными серверами, собирающими и обрабатывающими данные от различных служб ГПК, повысить отказоустойчивость системы. Однако применение в рамках специализированного информационного обеспечения управления предприятием ГПК технологий РБД предполагает решение дополнительных задач по поддержанию копий данных в непротиворечивом состоянии, разрешению конфликтов при блокировках наборов записей, восстановлению при сбоях. Это достигается за счет репликации данных. Расчет значений её параметров, при которых достигается высокая эффективность функционирования системы, является сложной задачей, которая на настоящий момент, применительно к условиям функционирования ГПК, не имеет однозначного решения.

Так как основное назначение РБД состоит в удовлетворении информационных потребностей пользователей (приложений), то оценку эффективности её функционирования целесообразно производить с точки зрения эффективности выполнения запросов и транзакций, отражающих информационные потребности конечных пользователей, приложений и действий над РБД.

Вопросам управления репликацией в РБД посвящено множество работ зарубежных ученых: К. Луни, Б. Брила, С. Рига, Х. Кросинга, С. Марча, Ж. Науманна и др. Среди отечественных исследователей необходимо отметить работы, посвященные вопросам разработки моделей и алгоритмов управления репликацией в РБД: А. Апанасевича, А. Ю. Иванова, В. Н. Кухарева, В. Е. Белоусова, С. Д. Кузнецова и др. Работа Л. И. Мейкшан посвящена разработке модели двухуровневой информационной системы с репликацией данных, доработка которой выполнена в диссертационном исследовании.

Так как АСУ предприятием ГПК функционирует в условиях большого количества случайных факторов: повышенная влажность, запыленность, сейсмоактивность, низкие и высокие температуры, вибрации, то используемое оборудование оснащается дополнительными элементами, снижающими влияние данных факторов. Так, в помещениях шахты с взрывоопасной средой используются понижающие разделительные трансформаторы, оборудование взрывозащищенного и искробезопасного исполнения. С другой стороны, наличие специализированного оборудования повышает стоимость и, как следствие,

требует эффективного распределения ресурсов при настройке специализированного информационного обеспечения. При этом нерациональный выбор параметров репликации данных влечет за собой превышение допустимых временных задержек на различных этапах обработки запросов и, как следствие, снижает оперативность отклика РБД.

В результате возникает противоречие, состоящее в том, что уменьшение количества реплицируемых данных порождает увеличение числа удаленных заявок на обслуживание. Как следствие, среднее время отклика РБД на запросы увеличивается за счет удаленного обслуживания запросов. С другой стороны, увеличение количества реплицируемых данных приводит к росту числа заявок на репликацию, что создает дефицит вычислительных и сетевых ресурсов. Среднее время отклика РБД на запросы увеличивается за счет роста временных задержек на различных этапах обслуживания запросов.

Исходя из данных положений тема, посвященная модели и алгоритмам управления параметрами репликации в распределенной базе данных предприятия горнопромышленного комплекса, является актуальной и обуславливает выбор объекта и предмета исследования.

Объект исследования: система управления репликацией в РБД предприятия ГПК.

Предмет исследования: способы, алгоритмы и методы управления репликацией в процессе обработки информации в РБД предприятия ГПК.

Цель исследования: уменьшение времени отклика РБД предприятия ГПК на запросы при заданных ограничениях на временные задержки путем управления параметрами репликации.

В диссертации поставлены следующие частные исследовательские задачи:

1. Провести сравнение известных моделей функционирования РБД при репликации, способов управления параметрами репликации, используемых в АСУ предприятием ГПК.

2. Разработать математическую модель отклика РБД на запросы при репликации, обеспечивающую управление совокупностью параметров репликации на уровне физической интерпретации при ограничениях на временные задержки обработки запросов в различных её элементах.

3. Разработать алгоритм вычисления оптимальной загруженности резервного узла РБД при репликации, позволяющий определять значения параметров репликации и обеспечивающий снижение среднего времени отклика РБД на запросы.

4. Разработать алгоритм выбора фрагментов данных для немедленной репликации, позволяющий по установленным значениям параметров репликации определять наборы фрагментов данных для немедленной репликации, при которых достигается снижение среднего времени отклика РБД на запросы.

5. Разработать алгоритм автоматизации процесса конфигурирования репликации в РБД предприятия ГПК, позволяющий вычислять параметры репликации в РБД предприятия ГПК и формировать решения для администратора по её конфигурированию.

Методы исследования, использованные в процессе выполнения диссертационной работы: теория вероятностей и математической статистики, теория массового обслуживания, дифференциального исчисления, теория эффективности целенаправленных процессов, статистического планирования экспериментов.

Диссертационная работа соответствует паспорту специальности 05.13.06 – «Автоматизация и управление технологическими процессами и производствами (промышленность)» по пункту №9: «Методы эффективной организации и ведения специализированного информационного и программного обеспечения АСУТП, АСУП, АСТПП и др., включая базы и банки данных и методы их оптимизации».

Научная новизна:

1. Математическая модель отклика РБД на запросы при репликации, базирующаяся на модели двухуровневой информационной системы с репликацией данных, отличающаяся учетом совокупности параметров: интенсивности запросов на обновление ($I'u$) и интенсивности поисковых запросов ($I'q$), обрабатываемых на резервных серверах, на уровне физической интерпретации.

2. Алгоритм вычисления оптимальной загруженности резервного узла при репликации в РБД, описываемой математической моделью отклика на запросы, основанный на модифицированном методе линейных комбинаций, отличающийся формированием ограничений, обеспечивающих режим функционирования РБД предприятия ГПК без блокировки.

3. Алгоритм выбора фрагментов данных для немедленной репликации, основанный на оптимизированном методе частично-целочисленного линейного программирования с аддитивным алгоритмом для задач с двоичными переменными, отличающийся процедурой принятия решения по критерию минимума объема пересылаемых реплик.

4. Способ управления репликацией в РБД, основанный на гибридном методе репликации, отличающийся автоматизацией подготовки принятия решения по управлению репликацией, защищенный патентом на изобретение.

Положения, выносимые на защиту:

1. Математическая модель отклика РБД на запросы при репликации.

2. Алгоритм вычисления оптимальной загруженности резервного узла РБД при репликации.

3. Алгоритм выбора фрагментов данных для немедленной репликации.

4. Способ управления репликацией в РБД.

Теоретическая значимость полученных решений заключается в разработке нового гибридного метода репликации, позволяющего за счет управления параметрами репликации в РБД предприятия ГПК подстраиваться под имеющиеся вычислительные и сетевые ресурсы с целью повышения её реактивности.

Практическая значимость заключается в разработке совокупности алгоритмов и доведении их до программной реализации, что подтверждается свидетельствами о государственной регистрации программ для ЭВМ № 2013611771 от 4 февраля 2013 года и № 2013616315 от 19 июня 2013 года, патентом на полезную модель № 126161 от 20 марта 2013 года и изобретением (положительное решение от 25.10.2013 о выдаче патента на изобретение "Способ репликации информации в распределенных базах данных с конкурентным распределением потоков" по заявке № 2012116021).

Полученные результаты могут использоваться на предприятиях ГПК с целью эффективной организации специализированного информационного обеспечения, создающей условия для снижения среднего времени отклика на запросы при заданных ограничениях на временные задержки обработки запросов в различных элементах РБД.

Публикации. По теме диссертационного исследования опубликовано 16 работ, в том числе 5 в изданиях, рекомендованных ВАК при Минобрнауки РФ.

Апробация. Основные положения и результаты работы были доложены и обсуждены на 17-ой Международной открытой научной конференции "Современные проблемы информатизации в моделировании и социальных технологиях" (г. Воронеж, 2012 г.), Международной молодежной научно-практической конференции СКФ МТУСИ "ИН-ФОКОМ-2012" (г. Ростов-на-Дону, 2012 г.), Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых "Научная сессия ТУСУР-2013" (г. Томск, 2013 г.), Международной молодежной научно-практической конференции СКФ

МТУСИ "ИНФОКОМ-2013" (г. Ростов-на-Дону, 2013 г.), Всероссийской научно-практической конференции "Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов" (г. Барнаул, Алтайский государственный университет, 2013 г.); 19-ой Международной открытой научной конференции "Современные проблемы информатизации" (г. Воронеж, 2014 г.)

Структура и объем работы. Диссертация состоит из введения, четырех глав и заключения. Диссертация содержит 142 страницы, 31 рисунок, 20 таблиц, 1 приложение. Список литературы содержит 111 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во **ВВЕДЕНИИ** обоснована актуальность работы, описан объект и предмет исследования, сформулированы цель и задачи диссертационной работы, перечислены использованные в работе методы исследования, обоснована научная новизна, теоретическая и практическая значимость результатов работы.

В **ПЕРВОЙ ГЛАВЕ** выделены особенности РБД и технологии репликации данных, приведено общее описание информационного обеспечения управления предприятием ГПК, представлена структура РБД предприятия ГПК "ШахтИнвестКузбасс", приведены особенности репликации в ней, рассмотрены подходы к моделированию процессов, протекающих в РБД при репликации.

Информационное обеспечение управления предприятием ГПК представлено РБД с репликацией резервных серверов по расписанию, рисунок 1. Их взаимодействие строится на основе общего информационного фонда, обеспечивающего единый общесистемный подход на всех этапах сбора, обработки и выдачи информации. При этом к подсистеме информационного обеспечения выдвигается ряд требований со стороны других подсистем по обеспечению оптимальным объемом информации в требуемые сроки.

Информационное обеспечение управления предприятием ГПК предназначено для решения задач диспетчерского, производственно-технологического, организационно-экономического управления технологическими процессами шахты, а также повышения безопасности работ проводимых в шахте. При этом информационное обеспечение управления предприятием ГПК представляет собой многоуровневую систему сбора и обработки данных с определенным кругом задач, эффективная интеграция которых в автоматизированной системе достигается в результате правильной организации системы управления данными. При этом распределенный характер и нестабильность условий функционирования шахты обуславливает целесообразность использования для управления данными распределенные базы данных.

Применение технологий РБД для эффективной организации и ведения специализированного информационного обеспечения требует решения задач по конфигурированию репликации. При этом под репликацией понимают процесс копирования информации из одного хранилища данных в другое с последующей синхронизацией их содержимого, направленный на: поддержание узлов данных в актуальном состоянии; обеспечение резервирования данных; объединение информации из нескольких массивов данных; поддержание в работоспособном состоянии узлов с непостоянным соединением с ядром системы. Каждый раз при изменении копии она начинает отличаться от всех прочих. Соответственно, для сохранения непротиворечивости эти изменения должны быть перенесены и на остальные копии, причем как можно быстрее.

Однако в настоящее время в РБД предприятия ГПК "ШахтИнвестКузбасс" репликация проводится по расписанию, что порождает высокую интенсивность удаленных запросов. При этом нерациональная нагрузка на вычислительные и сетевые ресурсы нега-

тивно сказывается на временных задержках обслуживания запросов на различных этапах их обработки, что, в свою очередь, влечет за собой снижение реактивности РБД.

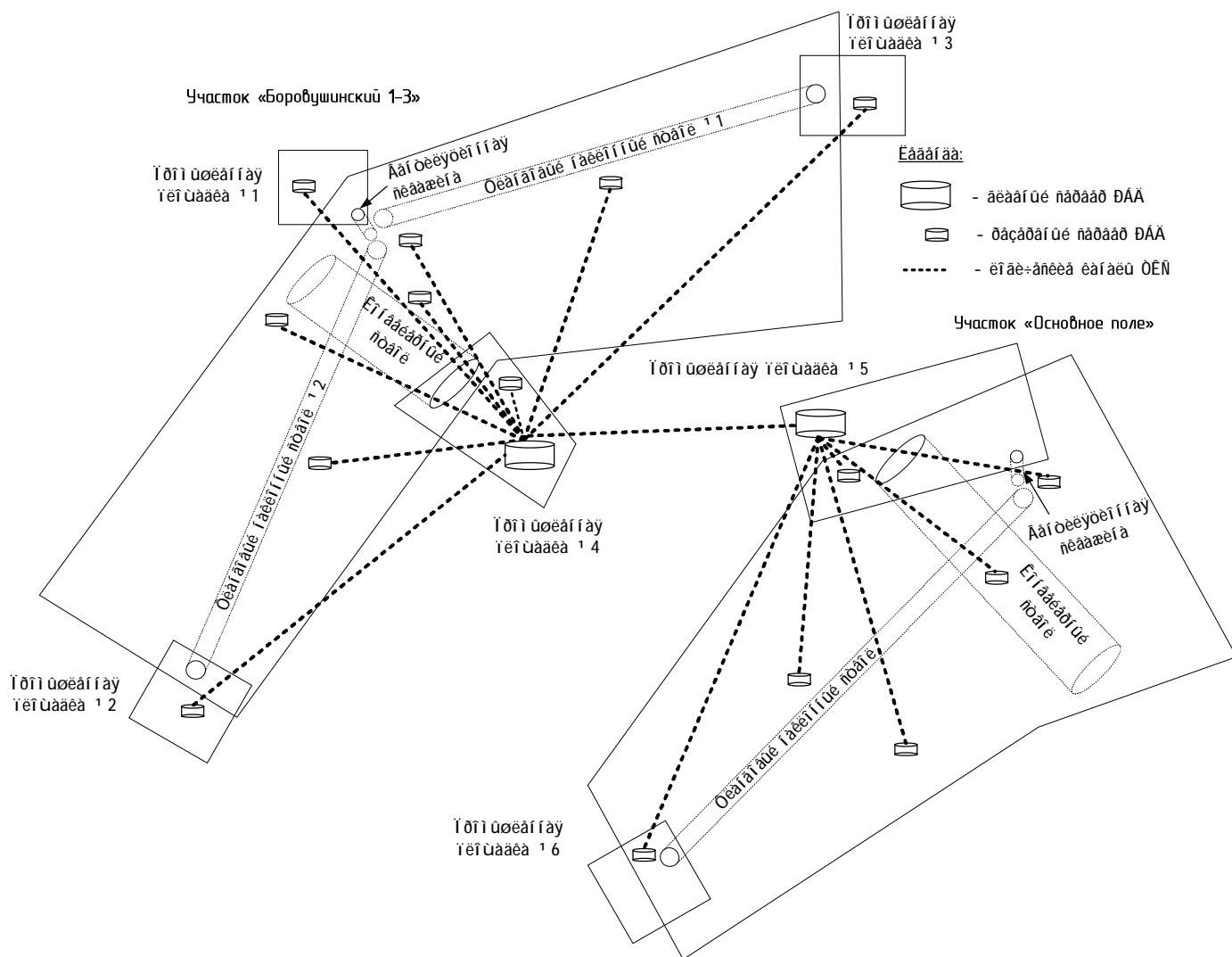


Рисунок 1 – План размещения главного и резервных серверов РБД предприятия ГПК

РБД предприятия ГПК "ШахтИнвестКузбасс", состоящая из одного главного сервера и множества резервных серверов, обрабатывает данные от служб: аэрогазового контроля; вентиляционного проветривания; теплоэнергетики шахты; пожарного водоснабжения; водоотлива; электроснабжения; конвейерного транспорта; технологического оборудования очистных и подготовительных участков шахты; геосейсмического мониторинга; мониторинга параметров шахтного пространства; виброакустического мониторинга горного массива; радиологического мониторинга горного массива; мониторинга и контроля гидрогеологической обстановки; радиозондирования горного массива; системы наблюдения и оповещения персонала.

Общая площадь ГПК "ШахтИнвестКузбасс" составляет более 30 квадратных километров, рисунок 1. Управление технологическими процессами производится на 10 угольных пластах: Выклинившийся, Надартельный 2, Артельный, Абрамовский, Лыжинский, Кумпановский, Верхний, Двойной-Промежуточный (1-ая пачка), Двойной-Промежуточный (2-ая пачка).

В диссертационном исследовании доработана модель двухуровневой информационной системы с репликацией данных, представленная в работе Л. И. Мейкшан, на предмет снятия ограничения по управляющим параметрам и детализации времени ожидания и передачи данных по телекоммуникационной подсистеме (ТКС).

ВТОРАЯ ГЛАВА посвящена разработке математической модели отклика РБД на запросы при репликации, базирующейся на модели двухуровневой информационной системы с репликацией данных, отличающейся учетом совокупности параметров: интенсивности запросов на обновление ($I'u$) и интенсивности поисковых запросов ($I'q$), обрабатываемых на резервных серверах, на уровне физической интерпретации. Новизна и возможность применения представленной модели подтверждается патентом на полезную модель № 126161 "Система децентрализованного управления структурой распределенной базы данных" и изобретением (положительное решение от 25.10.2013 по заявке № 2012116021 о выдаче патента на изобретение "Способ репликации информации в распределенных базах данных с конкурентным распределением потоков").

В работе показано, что закон распределения статистических данных поступления запросов на резервные узлы РБД предприятия ГПК "ШахтИнвестКузбасс" согласуется с

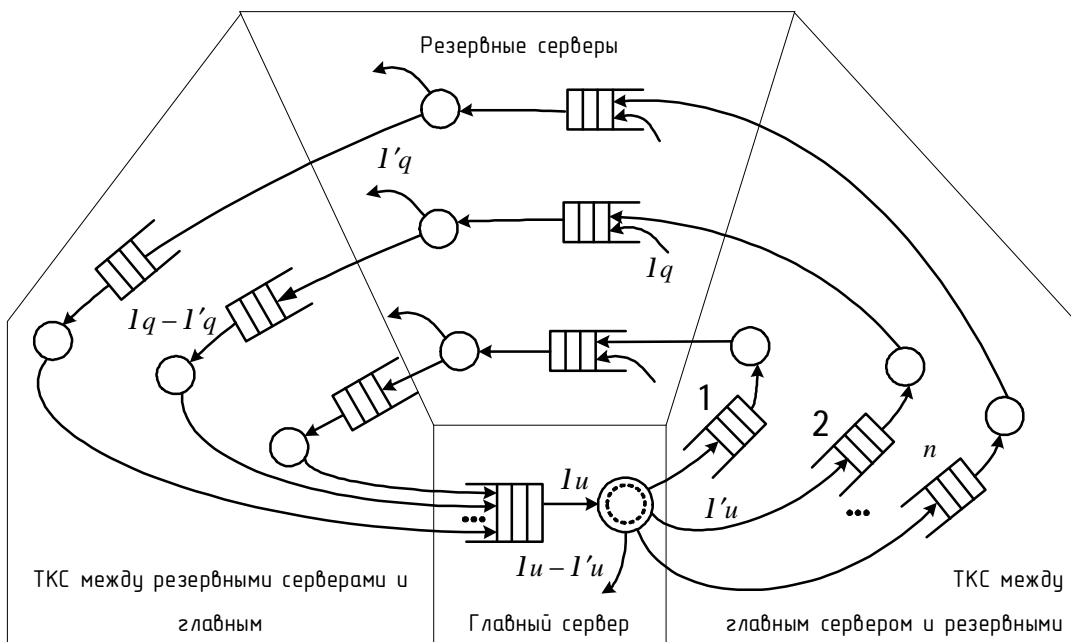


Рисунок 2 – Структура РБД предприятия ГПК при репликации

распределением Пуассона, что подтверждается результатами проверки на основе критерия согласия χ^2 Пирсона. При этом закон распределения случайной величины времени обслуживания заявок в резервных узлах – произвольный.

Поступление запросов в соответствии с распределением Пуассона, а

их обслуживание по произвольному закону позволяет моделировать отдельные элементы рассматриваемой РБД с помощью одноканальных СМО типа 1/M/G/FCFS. Таким образом, в соответствии с теорией массового обслуживания модель отклика РБД на запросы при репликации можно представить в виде совокупности одноканальных СМО типа 1/M/G/FCFS, описывающих обработку запросов на главном и резервном серверах, а также передачу данных по ТКС от главного сервера до резервных и обратно, рисунок 2.

Такие модели относятся к классу непрерывно-стохастических (Q-схемы). Имеет место многофазовая Q-схема с одноканальным оператором сопряжения:

$$Q = \langle W, U, H, Z, R, AL \rangle, \quad (1)$$

где W - подмножество входящих потоков, U - подмножество потоков обслуживания, H - подмножество собственных параметров, Z - подмножество состояний системы, R - оператор сопряжения, AL - оператор алгоритма обслуживания заявок.

Доработка модели двухуровневой информационной системы с репликацией данных Л. И. Мейкшан в соответствии с особенностями РБД предприятия ГПК позволила получить математическую модель отклика на запросы при репликации, в рамках которой среднее время отклика РБД на запросы при репликации определяется как:

$$\begin{aligned} \bar{T}(I'q, I'u) = & I'q/Iq \cdot (\bar{Tr}(I'q, I'u) + M[trq]) + \\ & + (1 - I'q/Iq) \cdot (\bar{Tgr}(I'q, I'u) + \bar{Trg}(I'q, I'u) + \bar{Tg}(I'q, I'u) + \\ & + M[tgq] + M[tgrqotkl] + M[trgq]) \end{aligned} \quad (2)$$

где $M[trq]$ - математическое ожидание времени обработки поискового запроса на резервном сервере; $M[trgq]$ - математическое ожидание времени передачи запроса с резервного сервера на главный; I_u - общая интенсивность запросов на обновление; I_q - общая интенсивность поисковых запросов; $I'u$ - интенсивность запросов на обновление, обрабатываемых на резервном сервере; $I'q$ - интенсивность поисковых запросов, обрабатываемых на резервном сервере; $\bar{Tr}(I'q, I'u)$ - среднее время ожидания обслуживания запроса на резервном сервере; $\bar{Tg}(I'q, I'u)$ - среднее время ожидания обслуживания запроса на главном сервере; $\bar{Trg}(I'q, I'u)$ - среднее время ожидания обслуживания запроса при передаче с резервного сервера на главный; $\bar{Tgr}(I'q, I'u)$ - среднее время ожидания обслуживания запроса при передаче с главного сервера на резервный.

Для расчета $\bar{Tr}(I'q, I'u)$, $\bar{Tg}(I'q, I'u)$, $\bar{Trg}(I'q, I'u)$, $\bar{Tgr}(I'q, I'u)$ используется формула

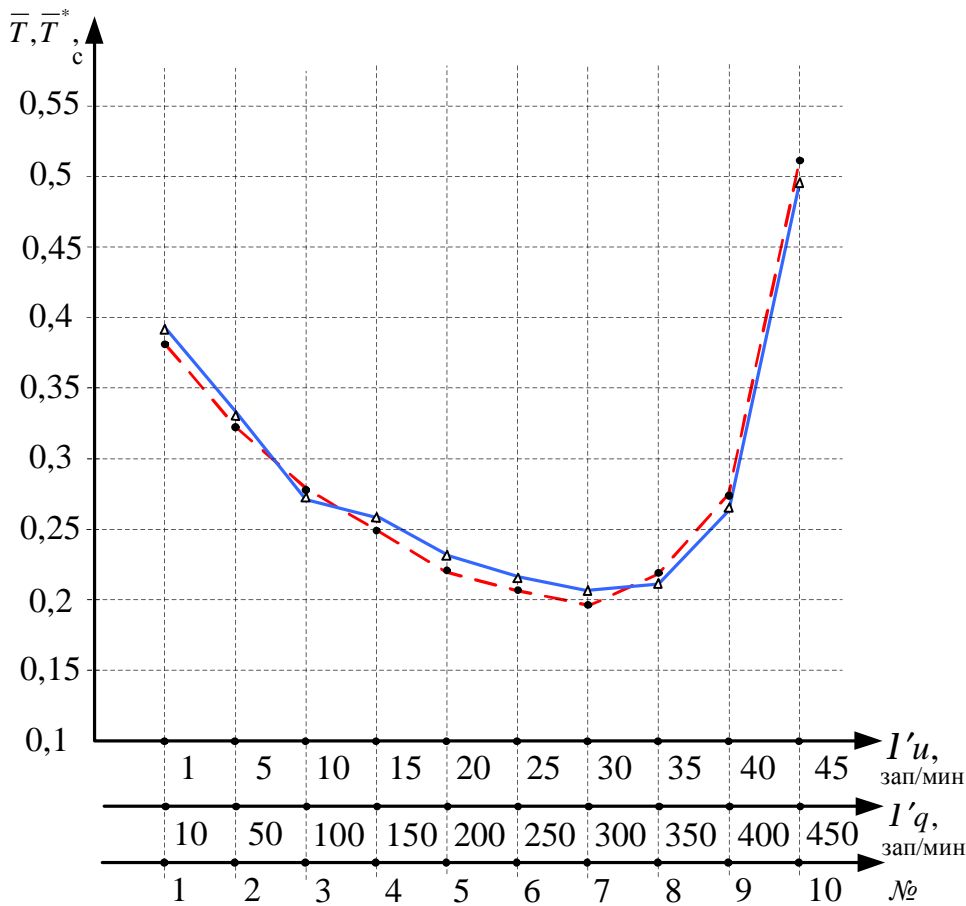


Рисунок 3 – Графики в условиях производства (сплошная линия) и модельного (пунктирная линия) среднего времени отклика РБД на запросы при увлечении значений управляющих параметров

Оценка чувствительности модели отклика РБД на запросы при репликации проводилась по каждому управляющему параметру в отдельности на основании приращений наблюдаемой переменной. Пары значений приращений оцениваемых параметров: для $I'q$ - (40%, 83%), для $I'u$ - (40%, 35%) позволили сформулировать вывод о достаточной чувствительности модели к изменению этих параметров.

Поллачека-Хинчина, справедливая для СМО типа 1/M/G/FCFS.

На рисунке 3 графически представлено расхождение данных, полученных в условиях производства, и модельного среднего времени отклика РБД на запросы при изменении значений управляющих параметров модели. Модельные данные получены на основе программы "ctrReplic", зарегистрированной в Федеральной службе по интеллектуальной собственности (свидетельство о государственной регистрации программы для ЭВМ № 2013618670 от 13 сентября 2013 года).

Оценка чувствительности модели отклика РБД на запросы при репликации проводилась по каждому управляющему параметру в отдельности на основании приращений наблюдаемой переменной.

В ТРЕТЬЕЙ ГЛАВЕ представлены алгоритм вычисления оптимальной загруженности резервного узла распределенной базы данных при репликации и алгоритм выбора фрагментов данных для немедленной репликации, проведена оценка их свойств.

Целевой функционал задачи уменьшения среднего времени отклика распределенной базы данных на запросы при заданных ограничениях на временные задержки обработки запросов в различных её элементах за счет обоснованного выбора значений интенсивностей обработки запросов на резервных серверах имеет вид:

$$\bar{T}(I'q(\{X'\}, type), I'u(\{X'\}, type)) \rightarrow \min_{I'q, I'u} \left\{ \begin{array}{l} (\bar{T}) \\ "A", \bar{T} \leq T\delta \end{array} \right. \quad (3)$$

где "А" - совокупность ограничений по свойствам: результативность и ресурсоемкость; $I'u$ - интенсивность запросов на обновление, обрабатываемых на резервном сервере; $I'q$ - интенсивность поисковых запросов, обрабатываемых на резервном сервере.

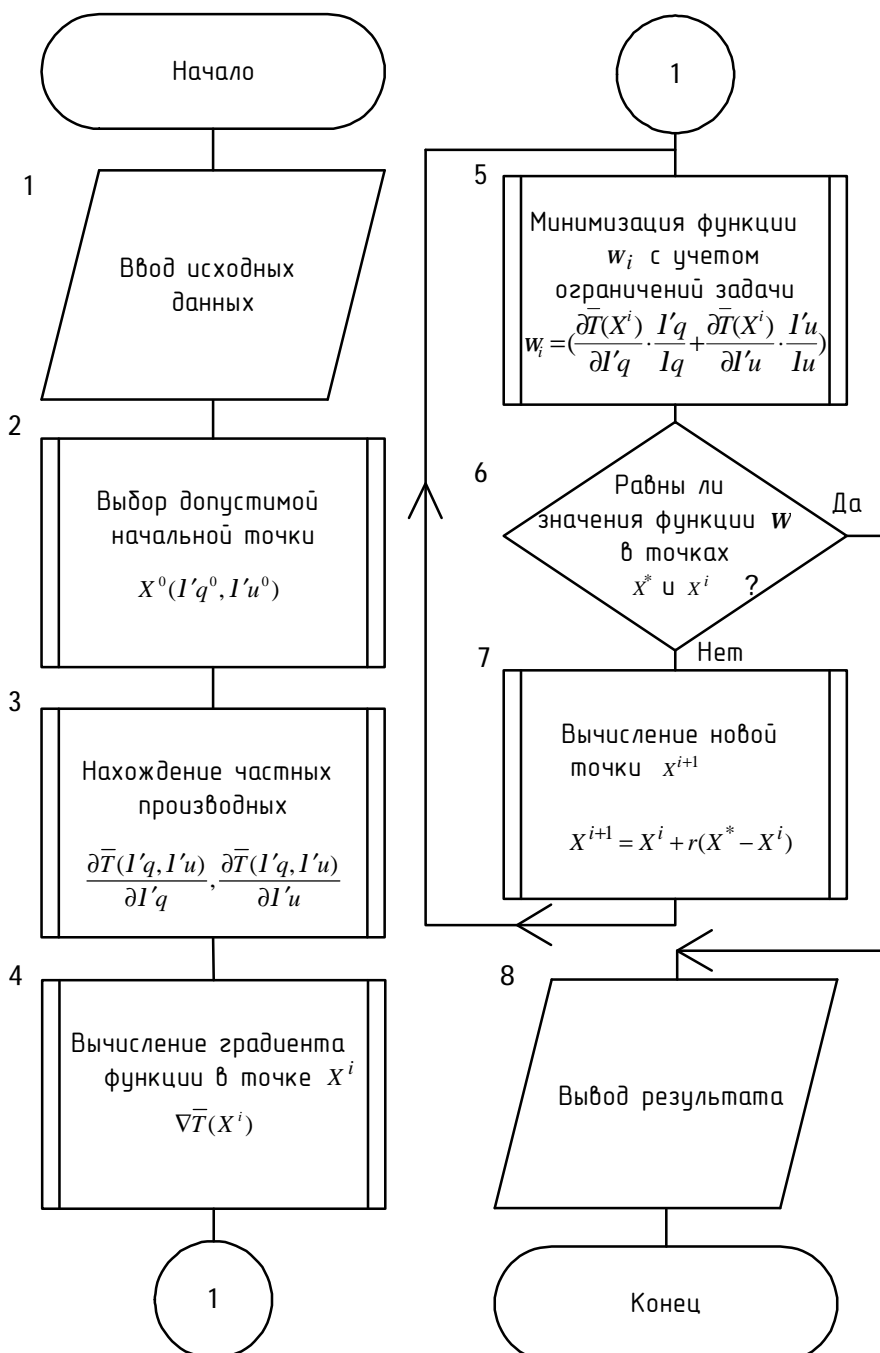


Рисунок 4 – Схема алгоритма вычисления оптимальной загруженности резервного узла РБД при репликации

Алгоритм вычисления оптимальной загруженности резервного узла при репликации в РБД, описываемой математической моделью отклика на запросы, основанный на модифицированном методе линейных комбинаций, отличающийся формированием ограничений, обеспечивающих режим функционирования РБД предприятия ГПК без блокировки, позволяет определять значения параметров репликации, при которых достигается снижение среднего времени отклика РБД на запросы, рисунок 4.

Решение задачи аналитически сопряжено с нахождением корней системы уравнений шестого порядка. По этой причине алгоритм построен с использованием численных методов. В основе алгоритма вычисления оптимальной загруженности резервного узла РБД при репликации лежит градиентный метод наискорейшего спуска, модифицированный для применения при наличии линейных ограничений.

Отдельные этапы алгоритма реализуются на основе симплекс-метода и метода золотого сечения.

Алгоритм позволяет при заданных значениях характеристик обслуживания запросов на различных этапах их обработки в РБД вычислять с заданной точностью значения интенсивности запросов на обновление $I'u$, обрабатываемых на резервном сервере, и интенсивности поисковых запросов $I'q$, обрабатываемых на резервном сервере, при которых достигается минимум среднего времени отклика на запросы при репликации.

Оценка корректности алгоритма проведена согласно методике Бейбера. Корректность алгоритма гарантирована при заданной точности проверки условия равенства результатов его работы в смежных итерациях $e_{\text{итер}} \leq 10^{-6}$ секунд и выполнении требований по загруженности РБД $0 \leq R_N < 1$. Алгоритм обладает достаточной точностью и устойчивостью, относится к классу алгоритмов с полиномиальной сложностью.

Минимум среднего объема пересылаемой по сети информации достигается в том случае, если совокупный размер фрагментов РБД, задействованных в немедленной репликации, минимальный, при условии обслуживания этими фрагментами максимального числа поисковых запросов, обрабатываемых на резервных серверах.

Алгоритм выбора фрагментов данных для немедленной репликации, основанный на оптимизированном методе частично-целочисленного линейного программирования с аддитивным алгоритмом для задач с двоичными переменными, отличающийся процедурой принятия решения по критерию минимума объема пересылаемых реплик, представлен на рисунке 5.

Полученный алгоритм позволяет находить решения за полиномиальное время с учетом снижения точности вычислений менее 5 %.

Критической частью алгоритма является цикл 6-9 с выходом по условию. При этом условием корректного завершения цикла 6-9 является конечное число точек зондирования, определяемое степенью фрагментации данных и производительностью ЭВМ, с использованием которой будут проводиться вычисления.

Показано, что алгоритм выбора фрагментов данных для немедленной репликации корректен по Бейберу при соблюдении предусловия: $n \leq 1000$, где n - количество обрабатываемых фрагментов данных. Определено, что вычислительная сложность представленного алгоритма соизмерима со сложностью $O(n^4)$, алгоритм является устойчивым.

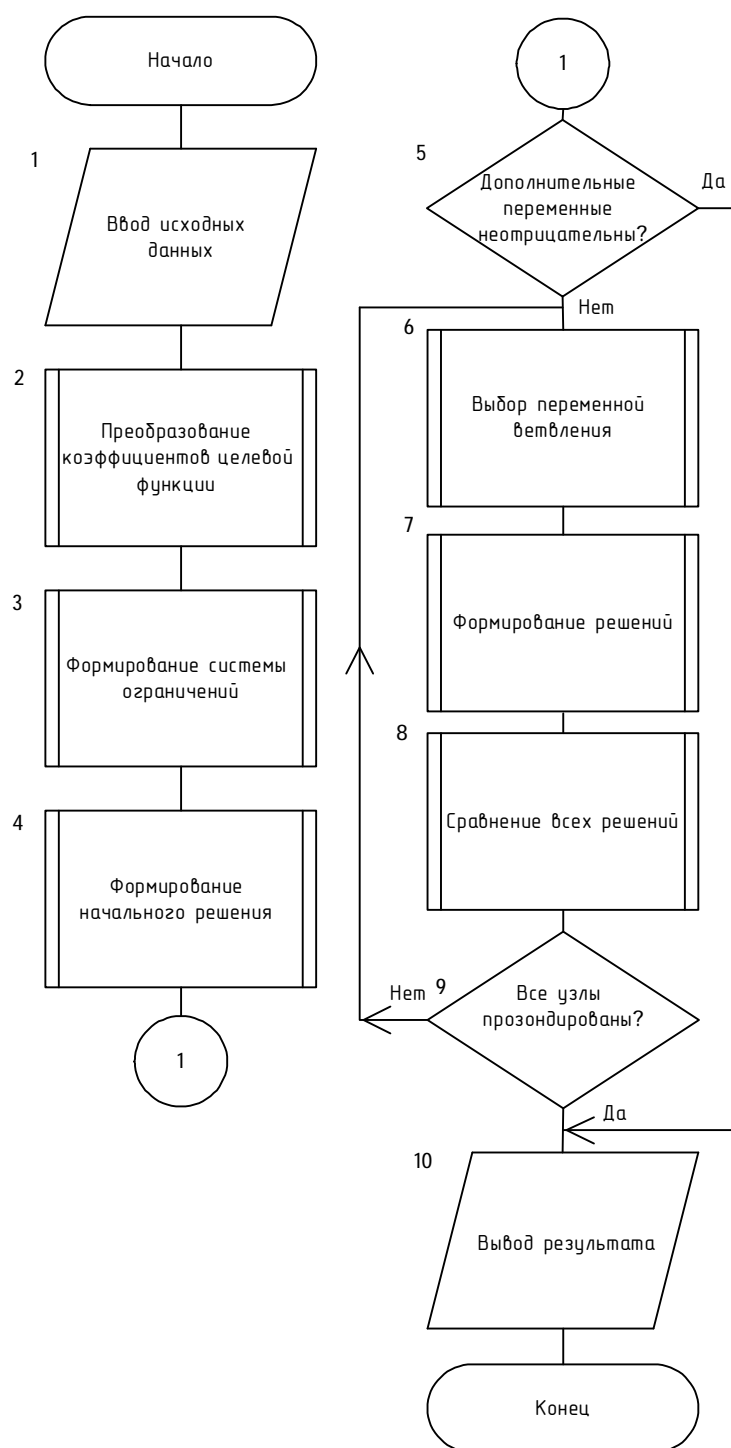


Рисунок 5 – Схема алгоритма выбора фрагментов данных для немедленной репликации

Целевой функционал задачи имеет вид:

$$Vq(x_1, x_2, \dots, x_n) + Vu(x_1, x_2, \dots, x_n) \rightarrow \min(x_1, x_2, \dots, x_n), \quad (4)$$

при наличии ограничений:

$$\begin{cases} x_1 \cdot Iu_1 + x_2 \cdot Iu_2 + \dots + x_n \cdot Iu_n \leq Iu^{opt} + d \\ Iu^{opt} - d \leq x_1 \cdot Iu_1 + x_2 \cdot Iu_2 + \dots + x_n \cdot Iu_n \\ x_1 \cdot Iq_1 + x_2 \cdot Iq_2 + \dots + x_n \cdot Iq_n \leq Iq^{opt} + d \\ Iq^{opt} - d \leq x_1 \cdot Iq_1 + x_2 \cdot Iq_2 + \dots + x_n \cdot Iq_n \\ x_1, x_2, \dots, x_n \in \{0,1\} \end{cases} \quad (5)$$

При этом:

$$Vq(x_1, x_2, \dots, x_n) = \bar{x}_1 \cdot Iq_1 \cdot n_1 + \bar{x}_2 \cdot Iq_2 \cdot n_2 + \dots + \bar{x}_n \cdot Iq_n \cdot n_n, \quad (6)$$

где Vq - средний объем информации, пересылаемой по сети за единицу времени в рамках удаленного обслуживания поисковых запросов к фрагментам РБД с репликацией по требованию ($x_i = 0$), а:

$$Vu(x_1, x_2, \dots, x_n) = x_1 \cdot Iu_1 \cdot n_1 + x_2 \cdot Iu_2 \cdot n_2 + \dots + x_n \cdot Iu_n \cdot n_n, \quad (7)$$

где Vu - средний объем информации, пересылаемой по сети за единицу времени в рамках обновления фрагментов РБД с немедленной репликацией ($x_i = 1$).

В ЧЕТВЕРТОЙ ГЛАВЕ представлена реализация способа управления репликацией в РБД, защищенного патентом на изобретение, в виде алгоритма автоматизации процесса конфигурирования репликации в РБД предприятия ГПК с учетом минимума среднего времени её отклика на запросы (рисунок 6), проведена оценка свойств алгоритма. Получены графики зависимостей интенсивностей обработки поисковых и запросов на обновление на резервных узлах, значения которых соответствуют минимуму среднего времени отклика РБД на запросы.

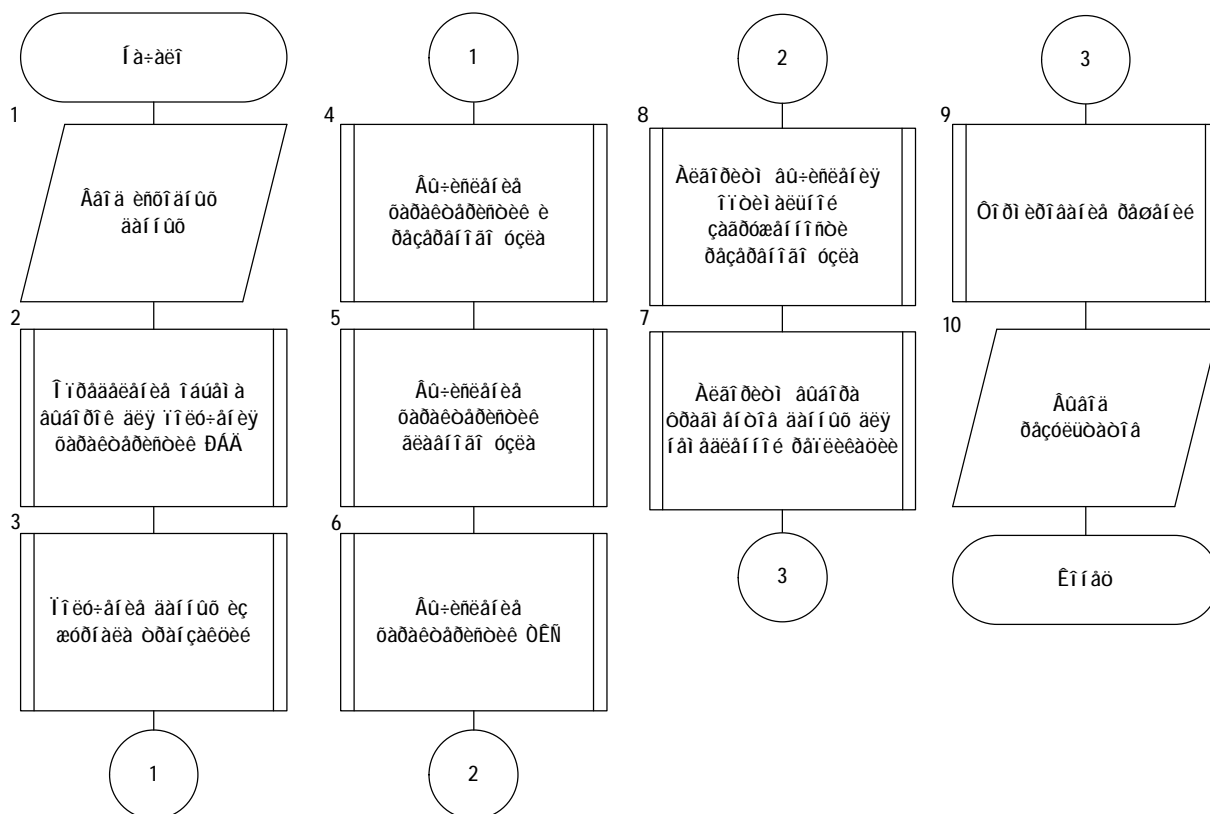


Рисунок 6 – Схема алгоритма автоматизации процесса конфигурирования репликации в РБД предприятия ГПК

Алгоритм отличается обеспечением условий функционирования РБД предприятия ГПК с пониженным средним временем отклика на запросы, позволяет вычислять значения параметров репликации в РБД предприятия ГПК и формировать решения для администратора по её конфигурированию.

Разработанный алгоритм автоматизирует процесс конфигурирования репликации в РБД предприятия ГПК за счет вычисления характеристик узлов РБД и ТКС, нахождения решений в соответствии с алгоритмом вычисления оптимальной загруженности резервного узла РБД при репликации и алгоритмом выбора фрагментов данных для немедленной репликации и формирования решений, позволяющих администратору РБД добиваться снижения среднего времени её отклика на запросы.

Вычисление характеристик узлов РБД и ТКС выполняется в рамках теории математической статистики путем корректного сведения средних значений к их математическим ожиданиям. Для нахождения минимального объема выборки, при котором среднее значение характеристики сводится к ее математическому ожиданию, с заданной точностью и надежностью используется выражение:

$$n = \frac{t^2 \cdot s^2}{d^2}, \quad (8)$$

где t - коэффициент доверия, s - среднее квадратическое отклонение вариационного признака, d - предельная ошибка выборки.

Показано, что вычисление значений количественных признаков РБД предприятия на примере ГПК "ШахтИнвестКузбасс" с учетом обеспечения требуемой точности и надежности вычислений возможно за допустимое время. При этом в качестве значения надежности выбран наиболее трудоемкий вариант, применяемый на практике, $g = 0,999$.

На основе алгоритма получены значения интенсивностей поисковых и запросов на обновление при изменении одного из параметров во всем рабочем диапазоне значений, при которых среднее время отклика РБД на запросы минимально, рисунок 7, 8.

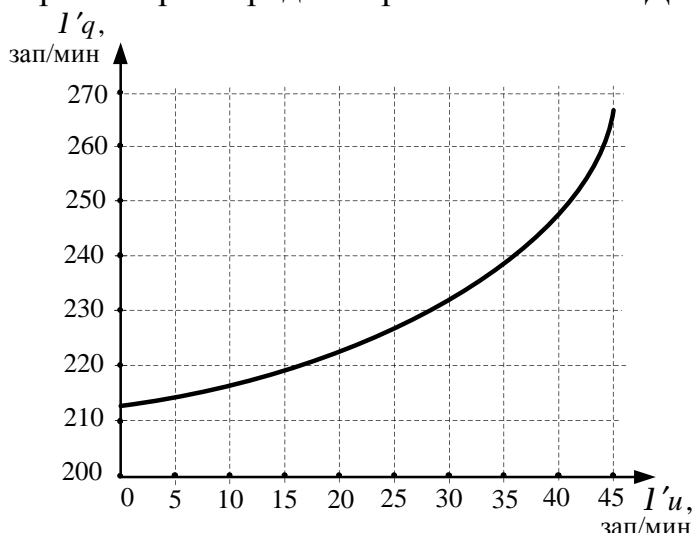


Рисунок 7 – График зависимости интенсивности поисковых запросов от изменения интенсивности запросов на обновление во всем рабочем диапазоне значений, при которых среднее время отклика РБД на запросы минимально

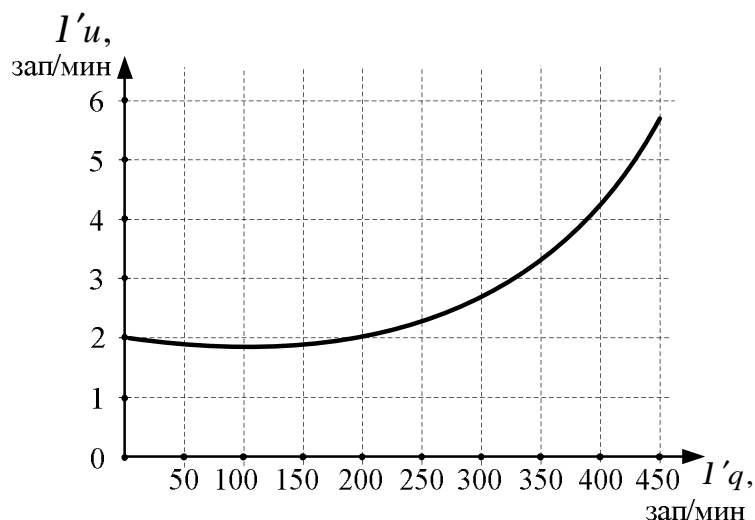


Рисунок 8 – График зависимости интенсивности запросов на обновление от изменения интенсивности поисковых запросов во всем рабочем диапазоне значений, при которых среднее время отклика РБД на запросы минимально

В диссертационной работе представлены шаблоны SQL-запросов, позволяющие администратору применять полученные решения для настройки репликации в РБД предприятия на примере ГПК "ШахтИнвестКузбасс".

Результаты оценки свойств алгоритма автоматизации процесса конфигурирования репликации в РБД позволили сформулировать вывод, что алгоритм является корректным, устойчивым, обладает полиномиальной сложностью и достаточной точностью.

На основе представленного алгоритма получен вариант набора фрагментов данных для немедленной репликации в условиях допустимого отклонения $d = 5\%$ от оптимальных значений: $I'q^{opt} = 214,31$ запросов в минуту и $I'u^{opt} = 2,07$ запроса в минуту.

При данном варианте набора фрагментов данных для немедленной репликации на узлах РБД предприятия ГПК "ШахтИнвестКузбасс" среднее время отклика на один запрос составляет $\bar{T} = 0,223$ секунды, тогда как существующее среднее время отклика $\bar{T}_{реальное} = 0,239$ секунды. Тем самым достигнут выигрыш по среднему времени отклика на запросы в 6,69% по сравнению со штатным режимом функционирования системы. При реагировании на аварийные ситуации функционирования предприятия ГПК, отличающимся значительным всплеском интенсивности поисковых запросов, данный выигрыш обеспечит запас времени для принятия обоснованных решений по оперативному управлению.

В ЗАКЛЮЧЕНИИ перечислены основные результаты диссертационной работы и сформулированы предложения по применению полученных результатов.

В ПРИЛОЖЕНИИ приведены характеристики фрагментов РБД предприятия ГПК "ШахтИнвестКузбасс".

ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ ДИССЕРТАЦИОННОЙ РАБОТЫ

В диссертационной работе решена актуальная научно-техническая задача разработки модели и алгоритмов управления параметрами репликации в РБД предприятия ГПК, позволяющих снизить время отклика РБД предприятия ГПК на запросы.

В рамках проведенных исследований получены следующие основные результаты:

1. Разработана математическая модель отклика РБД на запросы при репликации, базирующаяся на модели двухуровневой информационной системы с репликацией данных, отличающаяся учетом совокупности параметров: интенсивности запросов на обновление ($I'u$) и интенсивности поисковых запросов ($I'q$), обрабатываемых на резервных серверах, на уровне физической интерпретации. Проверка модели на основе сравнения модельного и времени обработки запросов, полученного в условиях производства, выявила достаточную адекватность и точность заявленной модели.

2. На основе математической модели разработаны алгоритмы вычисления оптимальной загруженности резервного узла распределенной базы данных при репликации и выбора фрагментов данных для немедленной репликации.

Алгоритм вычисления оптимальной загруженности резервного узла при репликации в РБД, описываемой математической моделью отклика на запросы, основанный на модифицированном методе линейных комбинаций, отличающийся формированием ограничений, обеспечивающих режим функционирования РБД предприятия ГПК без блокировки, позволяет определять значения параметров репликации, при которых достигается снижение среднего времени отклика РБД на запросы. Алгоритм является корректным, обладает достаточной точностью, вычислительной устойчивостью и допустимой сложностью.

Алгоритм выбора фрагментов данных для немедленной репликации, основанный на оптимизированном методе частично-целочисленного линейного программирования с аддитивным алгоритмом для задач с двоичными переменными, отличающийся процедурой принятия решения по критерию минимума объема пересылаемых реплик, позволяет находить решения за полиномиальное время с учетом снижения точности вычислений не

более чем на 5 %, является корректным, вычислительно устойчивым и обладает допустимой сложностью.

3. Способ управления репликацией в РБД, защищенный патентом на изобретение, реализован в виде алгоритма автоматизации процесса конфигурирования репликации в РБД предприятия ГПК, отличающегося обеспечением условий функционирования РБД предприятия ГПК с пониженным средним временем отклика на запросы, позволяющего вычислять значения параметров репликации в РБД предприятия ГПК и формировать решения для администратора по её конфигурированию. Оценка свойств алгоритма установила его корректность, устойчивость, полиномиальную сложность и достаточную точность.

Установлено, что для применения представленного алгоритма необходимо ограничиться числом фрагментов не превышающим 1000, так как последующая детализация выводит время реализации алгоритма за допустимые пределы в соответствии с современными требованиями к производительности рабочего мета администратора РБД предприятия ГПК "ШахтИнвестКузбасс".

5. Применение представленных алгоритмов на примере РБД предприятия ГПК "ШахтИнвестКузбасс" позволило получить выигрыш по среднему времени отклика на запросы в 6,69 % по сравнению со штатным функционированием системы.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

В ведущих рецензируемых научных изданиях, определенных Высшей аттестационной комиссией:

1. **Дунаев, В. А.** Разработка модели информационного обмена в распределенных базах данных в режиме репликации / В. А. Дунаев, О. В. Тараканов // Системы управления и информационные технологии. – 2012. – № 4.1(50). – С. 192–196 (доля участия автора – 50%, соответствует пункту 1 научной новизны)

2. **Дунаев, В. А.** Оценка времени реакции распределенной базы данных на запросы при гибридном механизме репликации / В. А. Дунаев // Информационные системы и технологии. – 2013. – № 6 (80). – С. 103–113. (соответствует пунктам 1, 2, 3 научной новизны)

3. **Дунаев, В. А.** Выбор фрагментов данных для размещения по узлам распределенной базы данных с учетом минимума среднего времени её отклика на запросы / В. А. Дунаев / Системы управления и информационные технологии. – 2013. – № 4(54). – С. 57–60. (соответствует пункту 4 научной новизны)

4. **Дунаев, В. А.** Модифицированная модель обработки запросов в распределенных базах данных / В. А. Дунаев, О. В. Тараканов // Программные продукты и системы. – 2014. – № 1(105). – С. 70–76 (доля участия автора – 50%, соответствует пункту 1 научной новизны)

5. **Дунаев, В. А.** Особенности управления параметрами репликации распределенной базы данных предприятия горнопромышленного комплекса / В. А. Дунаев, О. В. Тараканов // Информационные системы и технологии. – 2014. – № 2. – С. 45–52. (доля участия автора – 50%, соответствует пунктам 2, 3, 4 научной новизны)

На научно-технических конференциях Всероссийского и межведомственного уровня:

6. **Дунаев, В. А.** Методика выбора рациональных значений параметров информационного обмена при репликации распределенной базы данных / В. А. Дунаев // Современные проблемы информатизации в моделировании и социальных технологиях : сборник статей 17-ой Международной открытой научной конференции. – Воронеж : Издательство "Научная книга", 2012. – Вып. 17. – С. 188–190.

7. **Дунаев, В. А.** Состояние и перспективы развития технологий управления распределенными базами данных / В. А. Дунаев // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. – Ростов-на-Дону : ПЦ "Университет" СКФ МТУ-СИ, 2012. – С. 26–28.

8. **Дунаев, В. А.** Выбор режимов репликации в базах данных распределенного типа / В. А. Дунаев // Научная сессия ТУСУР–2013 : Материалы Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых. – Томск : В-Спектр, 2013. – Ч. 2. – С. 241–243.

9. **Дунаев, В. А.** Информационный обмен в распределенных базах данных при репликации / В. А. Дунаев, О. В. Тараканов // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. – Ростов-на-Дону : ПЦ "Университет" СКФ МТУ-СИ, 2013. – С. 92–94. (доля участия автора – 40%)

10. **Дунаев, В. А.** Анализ режимов репликации в базах данных распределенного типа / В. А. Дунаев // Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов : сборник статей всероссийской научно-практической конференции. – Барнаул : Барнаул, 2013. – С. 139–142.

11. **Дунаев, В. А.** Методика выбора фрагментов данных для размещения по узлам распределенной базы данных с учетом минимума среднего времени её отклика на запросы / В. А. Дунаев // Современные проблемы информатизации : сборник статей 19-ой Международной открытой научной конференции. – Yelm, WA, USA : Science Book Publishing House, 2014. – С. 166–169.

В других изданиях:

12. **Дунаев, В. А.** Математическое моделирование информационного обмена в распределенных базах данных в режиме репликации / В. А. Дунаев, О. В. Тараканов // Информационные технологии моделирования и управления. – 2012. – № 6(78). – С. 458–465 (доля участия автора – 60 %)

Свидетельства о государственной регистрации программ для ЭВМ:

13. **Расчет** функциональной живучести информационных систем с распределенной базой данных при репликации : свидетельство о государственной регистрации программ для ЭВМ № 2013611771 / В. А. Дунаев, О. Ю. Миронов, Н. В. Покусин, Д. О. Кривошея. заявка № 2012661052 от 13.12.2012. (доля участия автора – 25%, соответствует пунктам 2, 3 научной новизны)

14. **ctrlReplіc** : свидетельство о государственной регистрации программ для ЭВМ № 2013618670 / В. А. Дунаев, А. А. Цельшков, Ю. И. Федоров, А. Ю. Кузнецов, И. О. Ковыршин, О. А. Сенотрусов. заявка № 2013616315 от 19.07.2013. (доля участия автора – 30 %, соответствует пунктам 2, 3 научной новизны)

Изобретения и полезные модели:

15. **Система** децентрализованного управления структурой распределенной базы данных : пат. на полезную модель № 126161 Рос. Федерация : МПК⁸ G 06 F 12/00 / [В. А. Дунаев, Е. В. Лебеденко и др.] ; патентообладатель Гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – №2012116023 ; заявл. 19.04.2012, опубл. 20.03.2013. (доля участия автора – 25 %, соответствует пунктам 2, 3 научной новизны)

16. Положительное решение от 25.10.2013 по заявке № 2012116021 о выдаче патента на изобретение "Способ репликации информации в распределенных базах данных с конкурентным распределением потоков". / В. А. Дунаев, О. В. Тараканов, С. В. Козлов. (доля участия автора – 40 %, соответствует пункту 4 научной новизны)

ЛР ИД № 00670 от 05.01.2000 г.

Подписано к печати « 22 » апреля 2014 г.

Усл. печ. л.1 Тираж 100 экз.

Заказ № 179.

Полиграфический отдел «Госуниверситет-УНПК»

302035, г. Орел, ул. Московская, 65