Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.6

Бернацкий Владислав Викторович

Управление хранением данных при масштабировании и распределенной Обработке

АВТОРЕФЕРАТ
на соискание академической степени
магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Бранцевич П.Ю.
к.т.н., доцент

Минск 2017

КРАТКОЕ ВВЕДЕНИЕ

Несмотря на огромное количество доступных компьютеров в мире большинство этих систем имеют проблемы с обеспечением коммуникаций посредством разделяемых ресурсов. Однако последние инновации в распределённых базах данных (РБД) делают возможным получить доступ к данным через единую систему которая называется распределённая база данных или децентрализования система баз данных. Распределённая база данных (РБД) представляет собой коллекцию, состоящую из нескольких логически взаимосвязанных баз данных распределённых в компьютерной сети

Распределённые базы данных используют децентрализованные схемы управления данными где данные разбросаны по набору отдельных автономных узлов, которые могут взаимодействовать друг с другом

Распределённая система управления данными (РСУД) - это программное обеспечение которые управляет РБД и состоит из набора узлов, каждый из которых содержит локальную базу данных и предоставляет механизм доступа, который делает распределение прозрачным пользователю, чьи данные распределены и реплицированы на нескольких машинах, в отличие от централизованных систем баз данных , где хранится лишь одна копия данных. Все эти данные доступны всем другим узлам системы через общую сеть.

Технологии распределённых баз данных являются одними из наиболее популярных в последнее десятилетие. Их популярность превзошла существующие централизованные базы данных. В течение этого периода количество исследований в данной области возросло вместе с ростом количества коммерческих продуктов «первого поколения».

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритмов и программного обеспечения для решения задач управления хранением данных при масштабировании и распределённой обработке, которое обеспечивало бы максимальный уровень доступности, устойчивости к разделению, согласованности данных.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Произвести обзор предметной области распределённого хранения данных,
2. Произвести обзор предметной области распределённого обеспечения масштабируемости
3. Проанализировать существующие алгоритмы репликации данных
4. Проанализировать существующие алгоритмы достижения согласованности данных
5. Разработать алгоритмы, методы и модель системы для обеспечения распределённого хранения данных с возможностью масштабирования
6. Разработать программное обеспечения для распределённого хранения данных с использованием разработанных алгоритмов.
7. Провести экспериментальные исследования, оценку и анализ результатов разработанной системы.

Объектом исследования являются распределённые системы хранения данных.

Предметом исследования является масштабирование распределённых систем хранения данных.

Основной *гипотезой*, положенной в основу диссертационной работы, является использование архитектуры CRDT в распределённых системах хранения данных. Данная архитектура обеспечивает *строгую согласованность в конечном итоге (strong eventually consistency)* при одновременном обеспечении высокой доступности, а также устойчивости к разделению. Но ограничения накладываемые данной архитектурой сильно сужает круг возможного применения. Данная архитектура также не описывает механизмов репликации данных. Расширение CRDT архитектуры, а также обеспечение механизмов репликации позволяют использовать её в широком кругу задач.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий» по теме «Разработка моделей, методов, алгоритмов, повышающих показатели проектирования, внедрения и эксплуатации программных средств для перспек-

тивных платформ обработки информации, решения интеллектуальных задач, работы с большими массивами данных и внедрение в современные обучающие комплексы» (ГБ № 16-2004, № ГР 20163588, научный руководитель НИР – Н.В. Лапицкая).

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя П.Ю. Бранцевича, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на 52-ой научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2016), международной научной конференции «Информационные технологии и системы» (Минск, Беларусь, 2016)

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатных работ, из них 2 статьи в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикаций автора.

В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны и проанализированы направления их решения, сформулированы требования к системам распределённого хранения данных.

Вторая глава посвящена разработке алгоритмов и моделей для обеспечения масштабируемости распределённого хранения данных. Разработаны механизмы расширения архитектуры CRDT для обеспечения дополнительных свойств данного подхода, которые позволяют расширить сферу задач решаемых данной архитектурой, а также обеспечивающие дополнительные свойства. Предложены алгоритмы репликации данных.

В третьей главе разработана архитектура приложения реализующая предложенные подходы, методы, алгоритмы и модели. Описаны структуры данных необходимые для обеспечения функционирования данной системы, а также основные процессы необходимы для её работы.

В четвёртой главе предложена программная реализация разработанной системы распределённого хранения данных, описана общая структура прило-

жения и избранные технологии. Проведены экспериментальные исследования и произведён анализ результатов работы данной системы.

Общий объем работы составляет 62 страниц, из которых основного текста – 55 страниц, 10 рисунков на 5 страницах, список использованных источников из 31 наименования на 2 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** описаны основные понятия данной предметной области. Рассмотрены распределённые системы хранения, их виды, преимущества перед нераспределёнными системами хранения данных, недостатки, функции и методы поддержки распределённых данных. Дано определение понятию масштабируемость и их разновидностям. Также затронуто описание таких понятий как фрагментация и репликация данных.

Также проведён анализ существующих алгоритмов и решений для организации распределённого хранения данных с обеспечением масштабируемости. У распределённых систем хранения данных есть 3 основных признака: согласованность данных (consistency), доступность данных (availability), устойчивость к разделению (partition tolerance). Существует теорема CAP, которая гласит, что возможно полное выполнение лишь двух из этих 3 признаков. Поэтому ввиду теоремы CAP существующие решения отказываются от соблюдения одного из признаков.

Данная глава заканчивается формированием требований к распределённым системам хранения данных. Эти требования включают в себя как требования предъявляемые к централизованным системам, так и имеют дополнительные требования для обеспечения распределённости и масштабируемости. Стоит заметить, что ввиду специфики данной области подобно теореме CAP не требуется полное соблюдение всех требований.

Во **второй главе** находится разработка алгоритмов и модели для системы распределённой обработки и масштабирования данных. Предложено использование архитектуры CRDT, которая обеспечивает устойчивость к разделению, высокую доступность, а согласованность данных ослаблена и представлена в виде согласованности в конечном счете. Предложены алгоритмы дополняющие данную архитектуру для решения задач не описанных в данной архитектуре, а также структуры данных расширяющие возможное использование данного подхода.

В CRDT предполагается, что система обеспечивает строгая согласованность в конечном счете и её состояния монотонно прогрессируют, не приводя к конфликтам. Монотонность в этом смысле означает отсутствие откатов: операции нельзя отменить, вернув систему в раннее состояние. Состояния такой сис-

темы связаны отношением частичного порядка, в математике такая система с определённой на ней операцией объединения называется полурешёткой.

CRDT системы предъявляют следующие требования к операциям разрешенным над хранимыми типам данных:

1. Они должны быть идемпотенты, т.е. повторное применение одной и той же операции должно приводить данные в одно и то же состояние.
2. Они должны быть коммутативны
3. Они должны быть ассоциативны

Выполнение данных требований приводит к тому, что последовательность состояний данных представляют собой полурешётку и гарантируют сходимость к одному результату.

Существует две разновидности CRDT систем: Operation-Based и State-Based.

Разработан алгоритм для обеспечения масштабируемости системы основанной на CRDT. Репликация отдельного фрагмента данных должна происходить не по всем вершинам, а лишь над частью. Распределение этих фрагментов должно быть равномерным. При использовании системы клиент может обращаться к любому узлу системы. Узел в свою очередь может проверить может ли он сам обработать запрос или же его нужно перенаправить в другую вершину. Это достигается за счет хранения метаданных на всех вершинах.

Алгоритмы репликации распределения реплик бывают статическими и динамическими, или адаптивными.

В статическом подходе количество реплик зафиксировано и неизменно.

В качестве динамического алгоритма распределения реплик предложен новый алгоритм, который является модификацией алгоритма CDDR для адаптации в использовании с CRDT подходом. В данном подходе ведется статистика запросов на чтение, запись, распространение изменений. Если в какую-то вершину не хранящую реплику определённых данных приходит много запросов на чтение или запись, то по достижению определённого порога эти данные реплицируются на этот узел. Также предусмотрен механизм исключения узлов из схемы репликации. Для этого сравнивается количество операций чтения и количество операций распространения изменения. Если последний тип запросов значительно превосходит первый, то вершина исключается из схемы репликации.

В **третьей главе** разрабатывается архитектура системы, структуры данных необходимые для поддержания работы системы, а также определяются основные процессы её работы. Основную работу системы выполняют два процесса.

Обработка пользовательских запросов.

Данная часть алгоритма работы системы начинается с ожидания клиентских запросов. После его получения система проверяет по своим конфигурационным данным может ли обработать этот запрос. При положительном ответе на данный вопрос производит необходимые изменения данных и отвечает клиенту. В случае невозможности самостоятельной обработки запроса используя

конфигурационные данные обращается к вершине способной обработать запрос и возвращает её ответ клиенту.

Обеспечение обмена данными между узлами сети

Эта часть рабочего процесса отвечает за обеспечение функционирования всей системы как единого целого. Здесь узел базы данных общается с другими участниками системы, проверяет присутствуют ли «соседние» вершины в сети, отсылает новые изменения им.

Это достигается путем регулярной отсылки сигнала heartbeat (пульс). Данный запрос отсылается не реже определённого времени. Он отсылает все не отосланные изменения. Если изменений нет, то отсылается пакет содержащий 0 изменений с заданным интервалом. Если ответа нет, то удалённая вершина помечается как недоступная и работа системы продолжается. Данные о недоступности какой-либо вершины может быть использована модулями для репликации и фрагментации для изменения алгоритма работы, а также клиентом в более сложных сценариях использования.

В **четвертой главе** описывается реализация приложения, а также проводятся экспериментальные исследования.

В ходе исследования было обнаружено, что работа CRDT системы в штатном режиме крайне эффективна, при не частых обновлениях реплики содержат актуальные данные. Однако с ростом нагрузки на сеть актуальность данных падает, но не сказывается на производительности. А для случаев с необходимым обеспечением временных блокировок данных производительность может значительно падать в системах с интенсивным обновлением информации

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Проанализирована предметная область распределённого хранения и масштабирования данных
2. Исследованы существующие решения обеспечивающие данную функциональность
3. Сформированы требования к системам подобного рода

Разработаны алгоритмы и модель обеспечения распределённого хранения данных с обеспечением масштабируемости системы

Разработана программная реализация описанной системы

Проведено тестирование и анализ результатов.

В ходе анализа результатов установлено, что данная система обеспечивает высокую производительность системы при невысоким требованиям к актуальности данных. С повышением требований к согласованности данных наблюдается снижение производительности системы. А предложенные модификации архитектуры CRDT значительно расширяют диапазон решамых задач.

Рекомендации по практическому использованию результатов

Разработанные алгоритмы и модель система может быть использована в реальных проектах в качестве масштабируемого распределённого хранилища данных. Однако при использовании стоит учитывать спектр решаемых задач и проанализировать подходит ли этим задачам согласованность strong eventually consistency.

Способ обеспечения сохранения данных на диск оставлен на усмотрение лиц занимающихся реализацией данных подходов.

Также в ходе исследования были обнаружены недостатки данной системы, что предоставляет поле для дальнейшего исследования способов улучшения данной модели системы.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Бернацкий, В.В. Модель системы распределённого хранения данных / В.В. Бернацкий // Компьютерные системы и сети: материалы 52-ой научной конференции аспирантов, магистрантов и студентов. – Минск: БГУИР, 2016. – с. 49-50.

2. Бернацкий, В.В. Управление и анализ информации в распределённых CRDT базах данных / В.В. Бернацкий // Информационные технологии и системы: материалы международной научной конференции – Минск: БГУИР, 2016. – с. 264.