

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**РАЗРАБОТКА СИСТЕМЫ ПОСТРОЕНИЯ АНАЛИТИЧЕСКИХ
ОТЧЕТОВ С ИСПОЛЬЗОВАНИЕМ ЯЗЫКА MDX НА БАЗЕ
OLAP-СЕРВЕРА MONDRIAN**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студентки 5 курса 551 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Казанцевой Снежаны Дмитриевны

Научный руководитель
доцент, к. ф.-м. н.

М. И. Сафрончик

Заведующий кафедрой
к. ф.-м. н., доцент

С. В. Миронов

Саратов 2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Основы OLAP-технологий и языка MDX.....	5
1.1 Принципы многомерного анализа данных и OLAP-технологий	5
1.2 Язык MDX: синтаксис, операторы и функции	5
1.3 Архитектура OLAP-сервера Mondrian	6
1.4 Сравнение OLTP и OLAP-подходов к хранению данных	6
2 Проектирование системы аналитической отчетности	8
2.1 Анализ предметной области книжного магазина	8
2.2 Проектирование звездообразной схемы хранилища данных	8
2.3 Разработка OLAP-куба в Mondrian	9
2.4 Архитектура клиент-серверного приложения.....	9
3 Реализация и тестирование системы	10
3.1 Настройка OLAP-сервера	10
3.2 Разработка веб-клиента на React	10
3.3 Примеры аналитических отчетов с MDX-запросами	11
ЗАКЛЮЧЕНИЕ	12

ВВЕДЕНИЕ

В эпоху цифровых технологий данные превращаются в один из главных ресурсов для компаний. Но их настоящая ценность раскрывается только когда есть хорошие инструменты для анализа - те, что превращают необработанные данные в полезную информацию для принятия решений.

Традиционные системы управления базами данных, основанные на реляционной модели, иногда не справляются с более сложными аналитическими запросами. Именно поэтому появились OLAP-системы — инструменты, которые используют многомерные структуры данных в виде кубов. MDX (англ. Multidimensional Expressions) — это основной язык для работы с такими кубами, который помогает создавать запросы к многомерным данным. Одной из популярных и бесплатных является OLAP-сервер Mondrian — он реализует многомерные модели поверх классических реляционных баз и умеет работать с MDX, что делает его отличным инструментом для построения аналитических отчетов. Актуальность темы исследования обусловлена растущей конкуренцией на рынке книжной торговли, где успешное функционирование магазинов зависит от способности эффективно анализировать продажи, динамику спроса и общую эффективность работы.

Целью выпускной квалификационной работы является разработка системы, которая позволяет формировать аналитические отчёты на базе Mondrian и адаптирована под книжный магазин.

В рамках работы решаются следующие задачи:

1. требуется спроектировать хранилище данных в PostgreSQL;
2. необходимо реализовать OLAP-куб с помощью Mondrian;
3. разработать веб-клиент на React;
4. обеспечить интуитивно понятный интерфейс для работы с MDX-запросами;
5. протестировать систему на реальных аналитических сценариях.

Практическая значимость работы состоит в том, что разработанная система аналитической отчетности может быть внедрена в книжных магазинах среднего масштаба без существенных финансовых затрат, обеспечивая комплексный анализ продаж, маржинальности товарных категорий и эффективности торговых точек на основе открытых технологий (PostgreSQL, Mondrian, React).

Структура и объем работы. Выпускная квалификационная работа состо-

ит из введения, трёх разделов, заключения, списка использованных источников и трёх приложений. Общий объём работы — 73 страницы, из них 46 страниц составляет основное содержание, включая 12 рисунков, а также Flash-носитель в качестве приложения, список использованных источников – 20 наименований. Работа сопровождается программным проектом, реализованным на JavaScript (React) и Python (FastAPI), приведённым в приложениях на электронном носителе.

1 Основы OLAP-технологий и языка MDX

1.1 Принципы многомерного анализа данных и OLAP-технологий

Многомерный анализ данных (MDX) организует информацию в виде гиперкуба с несколькими измерениями, такими как время, место или товар, что помогает выявлять взаимосвязи и тенденции, не заметные при традиционном анализе.

OLAP (анг. Online Analytical Processing), разработанный Эдгаром Коддом в 1993 году, обеспечивает быстрый и интерактивный доступ к агрегированной информации. Основные элементы OLAP включают многомерные модели данных, где измерения представляют оси куба, и факты — числовые метрики. MDX — это язык запросов для работы с многомерными данными.

Системы OLAP бывают ROLAP, MOLAP и HOLAP, каждая из которых имеет свои преимущества. В проекте выбран ROLAP-подход с OLAP-сервером Mondrian для гибкости и интеграции с веб-приложениями.

1.2 Язык MDX: синтаксис, операторы и функции

MDX (анг. Multidimensional Expressions) — это стандартный язык для работы с многомерными данными в OLAP-системах, позволяющий выполнять запросы к OLAP-кубам. Он был разработан Microsoft в 1997 году для взаимодействия с SQL Server Analysis Services (SSAS) и предоставляет мощные инструменты для анализа бизнес-данных.

MDX имеет декларативную парадигму, что упрощает написание запросов, позволяя пользователям сосредоточиться на том, что они хотят проанализировать. Язык работает с метаданными, такими как иерархии и уровни измерений, что облегчает создание сложных аналитических моделей. Основные компоненты MDX включают перечисления, кортежи, иерархии, члены и наборы.

MDX поддерживает множество встроенных функций для агрегации, фильтрации и сортировки данных. Запросы имеют четкую структуру, где оси определяются для строк и столбцов, а фильтрация позволяет сосредоточиться на конкретных аспектах данных. MDX также предлагает арифметические, сравнительные и логические операторы для выполнения различных операций.

В системе Mondrian MDX используется для оптимизации запросов, например, с помощью конструкции EXISTS, что улучшает производительность. MDX и SQL различаются тем, что MDX работает с кубами и поддерживает иерар-

хия данных нативно, в то время как SQL требует явной агрегации и эмуляции иерархий.

Таким образом, MDX является мощным инструментом для анализа данных в OLAP-системах, позволяя создавать гибкие и масштабируемые решения для бизнес-анализа.

1.3 Архитектура OLAP-сервера Mondrian

Mondrian — это популярный OLAP-сервер с открытым исходным кодом, предоставляющий мощные инструменты для многомерного анализа данных. Его архитектура основана на модульности и расширяемости, что позволяет интегрировать его с различными источниками данных. Основные компоненты включают кубы, измерения, факты и MDX-движок, который обрабатывает запросы и выполняет вычисления. Mondrian также использует механизмы кэширования для ускорения выполнения запросов и предоставляет API для интеграции с другими приложениями.

Архитектура состоит из четырех слоёв: интерфейсного, аналитического, метаданных и хранилища данных. При отправке MDX-запроса он проходит через эти слои, что обеспечивает высокую производительность и гибкость. Однако Mondrian имеет ограничения, такие как отсутствие встроенного ETL-процесса, сложности с обновлениями в реальном времени и зависимость от реляционных баз данных. Понимание архитектуры и ограничений Mondrian поможет эффективно использовать его возможности в аналитических решениях. В дальнейшем будет рассмотрен процесс разработки системы отчётности на базе Mondrian и MDX.

1.4 Сравнение OLTP и OLAP-подходов к хранению данных

Системы управления базами данных (СУБД) делятся на OLTP (анг. Online Transaction Processing) и OLAP (анг. Online Analytical Processing) в зависимости от назначения и архитектуры. OLTP-системы оптимизированы для обработки транзакций в реальном времени, выполняя множество коротких операций, таких как вставка и обновление данных. Они используются в банковских системах и электронных коммерческих платформах.

OLAP-системы, напротив, предназначены для анализа больших объёмов данных и поддержки принятия решений, выполняя сложные запросы и многомерный анализ. Данные в OLTP хранятся в нормализованных таблицах, тогда

как в OLAP — в денормализованных схемах, что ускоряет выполнение запросов.

Типичные операции в OLTP включают вставку и обновление, тогда как в OLAP преобладают выборки. OLTP-системы требуют высокой частоты процессоров и SSD для низкой задержки, в то время как OLAP-системы выигрывают от многоядерных процессоров и большого объема оперативной памяти.

С недавних пор наблюдается конвергенция OLTP и OLAP, что привело к появлению HTAP-систем, таких как ClickHouse. Облачные решения, такие как AWS Aurora и Redshift, становятся популярными, а выделение аналитических нагрузок в отдельные хранилища данных продолжает набирать популярность.

В данном проекте OLTP-слой реализован в PostgreSQL через таблицы книг и продаж, а OLAP-слой представлен витриной `sales_fact` с агрегированными данными для Mondrian. Синхронизация между слоями осуществляется через периодический ETL-процесс.

2 Проектирование системы аналитической отчетности

2.1 Анализ предметной области книжного магазина

Книжный магазин — сложная предметная область с ключевыми процессами: управление ассортиментом, закупка книг, учёт характеристик, динамическое ценообразование, продажи, управление скидками, программы лояльности и складская логистика. Для эффективного анализа разработана сущностная модель с основными сущностями: книги (isbn, жанр, цена, год издания), продажи (время, способ оплаты, количество, скидка) и клиенты (уровень лояльности, история покупок, география).

Ключевые показатели эффективности включают выручку, маржинальность, средний чек, топ-авторов и оборачиваемость запасов. Существующие системы ограничены отсутствием историчности, сложностью кросс-анализа и ручным формированием отчетов. Проект предусматривает OLAP-систему с многомерным анализом, автоматическим расчётом прибыли и интерактивными отчётами с быстрым откликом.

2.2 Проектирование звездообразной схемы хранилища данных

Проектирование звездообразной схемы хранилища данных является ключевым этапом для создания эффективной аналитической системы в книжном магазине. Звездообразная схема организует данные в центральной таблице фактов (sales_fact), окруженной таблицами измерений (dim_date, dim_book, dim_store и др.), что обеспечивает высокую производительность запросов и гибкость в анализе. Основные характеристики включают простоту структуры, высокую производительность и легкость в поддержке. Однако есть недостатки, такие как избыточность данных и ограниченная гибкость.

В отличие от звездообразной схемы, снежинообразная схема более сложная и использует нормализацию для уменьшения избыточности. Для реализации хранилища данных выбрана звездообразная схема, что обеспечивает высокую скорость отклика на запросы и простоту поддержки.

Процесс ETL включает извлечение данных из OLTP-системы, их трансформацию с учётом денормализации и агрегации, а затем загрузку в OLAP-хранилище. Спроектированная схема поддерживает ключевые показатели эффективности и обеспечивает быстрое выполнение запросов, что позволяет эффективно принимать управленческие решения.

2.3 Разработка OLAP-куба в Mondrian

Разработка OLAP-куба начинается с создания XML-схемы, которая определяет структуру данных. Сначала указывается фактовая таблица с помощью элемента `<Table>`, которая связывается с измерениями через ключи. Каждое измерение описывается иерархиями с уровнями детализации, например, временное измерение включает уровни год и квартал, а товарное измерение группирует данные по жанрам. Меры привязываются к числовым колонкам фактов с указанием функций агрегации, таких как сумма. Вычисляемые меры позволяют создавать новые показатели, например, средний чек на основе существующих метрик. Для контроля доступа используются элементы `<Role>` и `<MemberGrant>`. После настройки схема проверяется с помощью утилиты валидации. Производительность повышается активацией кэширования и настройкой пулов соединений сервера Tomcat.

2.4 Архитектура клиент-серверного приложения

Система реализована по принципу разделения ответственности между компонентами, что обеспечивает гибкость и масштабируемость. Клиентский слой на React включает компоненты, такие как QueryEditor.js для редактирования MDX-запросов и ResultGrid.js для отображения результатов. Прокси-слой на FastAPI преобразует REST-запросы в SOAP-запросы для Mondrian, обеспечивая безопасность через JWT-аутентификацию и ограничение частоты запросов.

OLAP-сервер на Tomcat с Mondrian обрабатывает запросы с использованием ядра ROLAP и кэша сессий. PostgreSQL служит хранилищем данных с оптимизациями для OLAP, включая индексы и параллелизм. Система поддерживает асинхронную коммуникацию и обеспечивает высокую производительность, безопасность и масштабируемость, что позволяет эффективно обрабатывать запросы и предоставлять аналитические данные.

3 Реализация и тестирование системы

3.1 Настройка OLAP-сервера

В рамках реализации системы аналитической отчётности была проведена настройка OLAP-сервера, включая установку и конфигурацию Apache Tomcat и Mondrian. Apache Tomcat — это открытый сервер приложений, поддерживающий спецификации Java Servlet и JSP, что позволяет разрабатывать веб-приложения на Java. Основные характеристики Tomcat включают поддержку Java EE, легковесность, модульную архитектуру, высокую производительность и безопасность.

Процесс настройки OLAP-сервера включает несколько этапов. Сначала была установлена версия Tomcat 9.0.105 с помощью пакетного менеджера. Затем произведена конфигурация параметров Java Virtual Machine (JVM) для выделения необходимого объема памяти. После успешной установки проверялась доступность сервера.

Следующим этапом стала интеграция Mondrian в Tomcat. Для этого был собран WAR-артефакт с помощью Maven, после чего файл mondrian.war был размещен в директории webapps для автоматического развертывания. Конфигурация пула соединений к базе данных PostgreSQL была выполнена в файле context.xml, что обеспечило эффективное управление соединениями.

Для поддержки кросс-доменных запросов был создан кастомный сервлет XmlServlet.java, обрабатывающий CORS-заголовки. Регистрация сервлета была выполнена в файле web.xml. Для повышения производительности активирована функция кэширования схем кубов и настроена параллельная обработка запросов.

Для диагностики работы сервера использовался JMX, а также организовано логирование медленных запросов. Валидация работоспособности OLAP-сервера проводилась с помощью smoke-теста, отправляя запросы на сервер. Таким образом, настройка OLAP-сервера была успешно завершена, создавая надежную основу для системы аналитической отчетности.

3.2 Разработка веб-клиента на React

В рамках реализации веб-клиента для системы аналитической отчётности была разработана архитектура на основе React, обеспечивающая гибкость и масштабируемость приложения. Ключевые функциональные модули включают.

1. MDX Query Editor. реализован на базе Monaco Editor с поддержкой автодополнения через Language Server Protocol, что упрощает написание MDX-запросов;
2. Metadata Explorer. модуль для динамической загрузки структуры куба через эндпоинт /discover, визуализирующий иерархии в виде дерева, что помогает пользователям находить необходимые элементы;
3. Result Grid. использует библиотеку @mui/x-data-grid для отображения результатов с возможностью сортировки, фильтрации и экспорта в форматы CSV и Excel;
4. Графики. Многоуровневый конвертер для преобразования табличных данных в графики с использованием библиотеки Recharts, позволяющий создавать интерактивные графики.

Для управления состоянием приложения используется Redux Toolkit, что улучшает пользовательский опыт. Асинхронные thunk-экшены обрабатывают сетевые запросы и ошибки, предоставляя информативные сообщения.

Оптимизации, такие как виртуализация рендеринга, мемоизация компонентов и кэширование метаданных в IndexedDB, повышают производительность. Тестирование ключевых сценариев показало время отклика менее 100 мс. Реализованы Unit-тесты с использованием Jest для проверки корректности функций. В результате был создан полнофункциональный интерфейс для ad-hoc анализа данных, соответствующий современным требованиям к производительности и удобству использования.

3.3 Примеры аналитических отчетов с MDX-запросами

Система аналитической отчетности позволяет делать ad-hoc анализ с помощью MDX-запросов, что даёт пользователям возможность гибко исследовать данные и получать ценные выводы. Были разработаны оптимизированные MDX-запросы для ключевых сценариев, учитывающие структуру OLAP-куба: анализ по времени показал устойчивый рост продаж с пиком в 2022 году, товарная аналитика выявила самые прибыльные жанры и категории; сегментация клиентов показала, как VIP-пользователи влияют на общий доход, а географический анализ указал на повышенный спрос в Москве. В отчётах используются динамические параметры, иерархические операции и условное форматирование, что помогает лучше визуализировать данные и обеспечивает возможность глубокого ad-hoc анализа для принятия обоснованных решений.

ЗАКЛЮЧЕНИЕ

В результате работы была создана полноценная аналитическая OLAP-система для книжного магазина, охватывающая весь процесс от проектирования витрины в PostgreSQL до разработки веб-клиента на React. Система поддерживает нерегламентированные отчеты с использованием MDX, что позволяет аналитикам исследовать данные без знания SQL.

Технологическая интеграция включает сборку OLAP-куба в Mondrian с определением измерений (книги, время, магазины) и мер (продажи, прибыль). XMLA-сервис на Tomcat и прокси-сервер на FastAPI обеспечивают выполнение MDX-запросов. Примеры запросов, такие как продажи по жанрам и динамика по месяцам, подтверждают гибкость анализа.

Практическая значимость работы заключается в автоматизации формирования отчетов для менеджеров, что позволяет сравнивать эффективность магазинов и выявлять популярных авторов. Веб-интерфейс снижает порог входа для пользователей благодаря предустановленным шаблонам MDX.

В заключение, разработанная система продемонстрировала эффективность связки PostgreSQL, Mondrian и React для OLAP-аналитики, успешно решая задачи бизнеса книжного магазина. Модульная архитектура открывает возможности для адаптации системы в других отраслях, таких как ритейл и логистика.