

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**
Кафедра математической кибернетики и компьютерных наук

**РЕАЛИЗАЦИЯ ПРОСТОЙ ВІ-ПЛАТФОРМЫ НА
ОСНОВЕ СУБД POSTGRESQL**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 5 курса 551 группы
направления 09.03.04—Программная инженерия
факультета КНиИТ
Егорова Артема Анатольевича

Научный руководитель

доцент, к. ф.-м. н.

М. И. Сафрончик

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

ВВЕДЕНИЕ

В условиях стремительного роста объемов данных в компаниях особое значение приобретает эффективная система их анализа и интерпретации. В современном мире важнейшими инструментами принятия решений на всех уровнях управления стали т.н. инструменты «Бизнес-аналитики» (англ. BI - Business Intelligence). Ключевым компонентом любой современной BI системы является поддержка технологии многомерного анализа данных в реальном времени – OLAP (On-Line Analytical Processing), позволяющей выполнять сложные агрегированные запросы к большим массивам информации с высокой производительностью.

Многие BI-инструменты строятся на основе внешних OLAP-движков и специализированных хранилищ данных, имплементация и поддержка которых требует значительных ресурсов. К тому же, в связи с ужесточением санкционной политики, крупнейшие вендоры, предоставляющие комплексные BI-решения, ушли с рынка РФ, что подтолкнуло бизнес и разработчиков реализовывать и внедрять собственные BI-решения, реализованные на open-source стеке. В то же время современные реляционные СУБД «из коробки» предоставляют достаточно функциональных возможностей для реализации OLAP-подхода средствами самой СУБД. В частности, с использованием материализованных представления (materialized view). Такой подход позволяет существенно упростить архитектуру BI системы, сохранив при этом преимущества многомерного анализа. Таким образом, задача построения собственной BI-платформы с поддержкой OLAP средствами СУБД (без выделенного OLAP-движка) является актуальной и практически значимой.

Цель выпускной работы: разработка и внедрение BI-платформы, ориентированной на OLAP аналитику, с использованием материализованных представлений в PostgreSQL.

Задачи:

1. Ознакомиться с теорией моделирования OLAP и подходов к реализации многомерного анализа данных.
2. Исследовать возможности реализации OLAP-функционала средствами PostgreSQL.
3. Спроектировать архитектуру BI-платформы с учетом особенностей предметной области.
4. Реализовать прототип системы.
5. Провести анализ и подготовить вывод о функциональности разработанной системы, ее производительности и возможности дальнейшего развития.

По результатам реализации выпускной квалификационной работы была разработана собственная BI-платформа, обеспечивающая сбор, хранение и многомерный анализ данных по продажам. В качестве основы для OLAP-аналитики было использовано материализованное представление в PostgreSQL, позволяющее выполнять функции загрузки данных (ETL), агрегации, фильтрации, drill-down и roll-up.

Методологические основы реализации простой BI-платформы на основе СУБД POSTGRESQL представлены в работах Ridani M., Amnai M., С. Архипенкова, Д. Голубева, О. Максименко, Petkovic D.

Практическая значимость бакалаврской работы заключается в снижении накладных расходов на аналитику при внедрении собственной малой BI платформы средствами реляционной СУБД PostgreSQL.

Структура и объём работы. Бакалаврская работа состоит из введения, 3 разделов, заключения, списка использованных источников и 13 приложений. Общий объем работы – 48 страниц, из них 46 страниц – основное содержание, включая 23 рисунка и 0 таблиц, цифровой носитель в качестве приложения, список использованных источников информации – 21 наименование.

Краткое содержание работы

Первая глава «Теоретические основы многомерного анализа в BI-системах» посвящен определению бизнес аналитики, BI-систем, обзору технологии OLAP, хранилищ данных, а также теоретическим основам возможности многомерного анализа данных в современных СУБД на примере POSTGRESQL.

Современные предприятия в условиях цифровизации и конкуренции ежедневно накапливают большие объемы разнородных данных, которые требуют качественной аналитической обработки. Бизнес-аналитика (Business Intelligence, BI) — это комплекс процессов и инструментов, обеспечивающих анализ данных и поддержку управленческих решений. Ключевую роль в BI играет технология OLAP (On-Line Analytical Processing), обеспечивающая многомерный анализ информации на основе измерений и агрегированных показателей.

OLAP-модель основывается на многомерной структуре данных, где основными элементами являются:

1. Измерения (dimensions) — оси анализа (время, продукт, регион и т.д.), содержащие атрибуты и иерархии;
2. Иерархии (hierarchies) — упорядоченные уровни внутри измерений, позволяющие реализовать drill-down и roll-up;
3. Меры (measures) — количественные показатели (выручка, количество, среднее значение), подвергаемые агрегации;

OLAP-куб — структура, организующая меры в пересечениях измерений для быстрого доступа к агрегированным данным. Ключевые операции OLAP включают:

- 1) Drill-down — детализация данных до более низкого уровня;
- 2) Roll-up — агрегирование на более высокий уровень;
- 3) Slice-and-dice — фильтрация и формирование подмножеств;
- 4) Pivoting — переориентация осей анализа.

Выделяют три основных типа OLAP-систем:

1. MOLAP (Multidimensional OLAP) — использует специализированные многомерные хранилища, обеспечивает высокую скорость, но плохо масштабируется.
2. ROLAP (Relational OLAP) — работает поверх реляционных СУБД, гибкий и масштабируемый подход, особенно актуален при динамически изменяющихся данных.
3. HOLAP (Hybrid OLAP) — сочетает преимущества MOLAP и ROLAP: агрегаты хранятся в кубах, детали — в реляционных таблицах.

На практике наблюдается рост популярности ROLAP благодаря технологической зрелости СУБД и снижению затрат на внедрение. Для эффективного OLAP-анализа необходима проработанная структура хранения. Наиболее распространенные модели:

- 1) Схема «звезда» — денормализованная структура с центральной таблицей фактов и таблицами измерений;
- 2) Схема «снежинка» — нормализованная модель с иерархиями в измерениях, более гибкая, но сложнее в реализации.

Хранилище данных (Data Warehouse) служит основой BI и включает подсистемы ETL, администрирования и представления данных. Агрегационные слои в хранилищах позволяют ускорить аналитику за счет предрасчета типовых срезов.

Современные СУБД, в частности PostgreSQL, позволяют реализовать OLAP-функционал без внешнего аналитического движка. Использование материализованных представлений позволяет хранить предрасчитанные агрегаты, индексировать их и ускорять аналитические запросы. Это особенно актуально для компактных BI-решений с фиксированной структурой анализа.

PostgreSQL предоставляет богатый инструментарий: оконные функции, агрегаты, индексы, планировщики обновлений, что делает его подходящим для построения автономных ROLAP-систем малого и среднего масштаба.

OLAP остается ключевым элементом в BI-архитектуре, обеспечивая многомерный анализ данных. Современные тенденции демонстрируют переход

от классических MOLAP-систем к ROLAP и HOLAP решениям, основанным на реляционных базах данных. Использование PostgreSQL и материализованных представлений позволяет строить легковесные, но мощные OLAP-подсистемы без дорогостоящей инфраструктуры. Эффективная архитектура хранилища данных, включая агрегационные слои и гибкие модели, служит фундаментом надежной и масштабируемой аналитики, приближенной к бизнесу и его потребностям.

Вторая глава «Реализация собственной BI-платформы» посвящен реализации собственной простой BI платформы на базе СУБД POSTGRESQL.

Краткое содержание раздела 2: Реализация собственной BI-платформы

Разработка собственной BI-платформы ставит перед собой цель создания простой, но функционально насыщенной системы для многомерного анализа бизнес-данных без применения внешних аналитических движков. Решение должно включать механизмы загрузки данных, построение OLAP-срезов, реализацию drill-down и roll-up операций, а также визуализацию результатов анализа.

Платформа должна обеспечивать:

- 1) Импорт данных из CSV-файлов с автоматическим или ручным сопоставлением полей;
- 2) Формирование пользовательских аналитик;
- 3) Многомерный анализ: агрегации, фильтрация, динамика;
- 4) Визуализация данных (таблицы, графики, диаграммы);
- 5) Минимальную зависимость от сторонних компонентов.

Пользовательский интерфейс состоит из двух страниц: главной аналитической и страницы загрузки данных. Каждая поддерживает интерактивное взаимодействие с серверной частью через API.

Архитектура системы и компоненты

В качестве СУБД используется PostgreSQL. Модель данных реализована в виде схемы «звезда», где таблица фактов FactSales хранит числовые показатели и внешние ключи к измерениям: дата, регион, товар, канал, магазин и способ

оплаты. Агрегаты предрасчитываются в материализованном представлении SalesCube, обновляемом после загрузки данных. Это обеспечивает высокую производительность при аналитических запросах.

Сервер реализован на Python с использованием FastAPI. Два основных направления — загрузка данных и аналитика. Работа с базой осуществляется через ORM SQLAlchemy, позволяющую описывать таблицы как Python-классы, избегая прямых SQL-запросов. Сервер предоставляет API для агрегации (/analytics/cube, /analytics/drilldown, /analytics/timeseries) и для загрузки данных (/data/upload, /data/finalize_load).

Логика агрегации позволяет группировать данные по произвольным измерениям, рассчитывать сумму продаж и количество. Запросы к SalesCube оптимизированы индексами, а drill-down реализован через динамическую детализацию по выбранным аналитикам. В процессе загрузки производится сопоставление полей, проверка обязательных значений и обновление куба.

Интерфейс разработан на HTML, CSS и JS, обслуживается FastAPI как статические файлы. Визуализация данных реализована с помощью библиотеки Chart.js (круговые, линейные, столбчатые диаграммы). Все действия пользователя обрабатываются асинхронными JS-функциями. Интерфейс реализует slice-and-dice, сохранение фильтров и динамическое отображение табличных и визуальных аналитик.

Платформа проверяет корректность параметров на всех уровнях: валидация дат, проверка обязательных полей, контроль структуры CSV. Все критические действия обернуты в try/except, предусмотрены откаты транзакций и удаление временных файлов в случае ошибок.

Система разворачивается через Docker Compose и Uvicorn (ASGI-сервер). Вся инфраструктура состоит из двух контейнеров: backend (FastAPI + PostgreSQL) и frontend (статические файлы). Поддерживается автономный режим работы без сторонних сервисов.

Разработанная BI-платформа демонстрирует возможность построения автономной OLAP-системы с использованием PostgreSQL и минимального

набора инструментов. Благодаря применению материализованных представлений и продуманной архитектуре удалось обеспечить высокую производительность, гибкость в работе с данными и удобство визуализации без сложной внешней инфраструктуры. Такой подход особенно эффективен для малых и средних организаций, нуждающихся в быстрой аналитике при ограниченных ресурсах.

Третья глава «Анализ возможностей и практическое применение» посвящен детальному разбору аналитических сценариев на платформе, а также возможные зоны роста.

BI-платформа демонстрирует широкий спектр аналитических сценариев, направленных на поддержку управленческих решений. Пользователи могут исследовать продажи по регионам, категориям товаров, каналам продаж, магазинам и способам оплаты, переключаясь между суммой и количеством. Поддержка drill-down и slice-and-dice позволяет детализировать данные до уровня отдельных товаров и гибко перестраивать срезы анализа, повышая доверие к данным за счет прозрачности.

В табличном представлении реализован удобный интерактивный анализ с возможностью проваливания в детали (drill-down), а также гибкой настройки отображаемых полей. Возможность наложения фильтров по регионам и построения временных рядов (динамика по месяцам) делает платформу гибкой и наглядной для анализа продаж за различные периоды и по выбранным регионам.

Архитектурные ограничения платформы связаны с упрощенной моделью хранения данных и отсутствием модулей авторизации и ролевого доступа. При росте объема данных требуется реализация партиционирования для повышения масштабируемости. Отсутствие полноценного фреймворка во frontend усложняет дальнейшее развитие и интеграцию с внешними системами. Однако такая архитектура обеспечивает минимальные зависимости и высокую прозрачность, что делает платформу удобным стартовым решением для малых и средних команд. При дальнейшем развитии возможна интеграция с более

сложными BI-решениями и масштабирование на уровне корпоративной аналитики.

ЗАКЛЮЧЕНИЕ

В ходе выполнения выпускной квалификационной работы были продемонстрированы возможности современных технологий в части реализации BI-инструментария с OLAP-функциональностью. Разработанная в практической части BI-платформа демонстрирует, что в рамках ограниченного стека технологий можно полноценно обеспечить качественный многомерный анализ данных. Веб-платформа обеспечивает полный цикл пользовательского анализа от загрузки данных до получения визуальной аналитики, при этом все шаги цикла реализованы в рамках единого интерфейса, что снижает порог вхождения и ускоряет получение результата. Были реализованы основные функции BI-системы: фильтрация, группировка, drill-down, roll-up с помощью единого хранилища и одного RESTfull-сервиса.

С практической точки зрения платформа может быть использована для легких корпоративных решений в условиях ограниченного бюджета или отсутствия доступа к коммерческим BI-инструментам. В современном мире стоимость хранения данных стала совсем низкой, при этом основным драйвером расходов выступают вычисления, когда каждая выборка требует соединения отдельных таблиц, бизнес в конечном итоге платит за каждое такое соединение [21]. Использование материализованных представлений позволяет уменьшить накладные расходы на аналитику.

Таким образом, работа подтвердила актуальность темы исследования и практическую значимость использования BI-инструментария в анализе данных. Развитие технологий обработки данных, включая машинное обучение и предиктивную аналитику влечет дальнейшее совершенствование BI-инструментария. В свою очередь, внутренние средства современных СУБД продолжают развиваться, предлагая новые возможности для реализации эффективной OLAP-функциональности, что делает тему перспективной для дальнейшего изучения и практического применения в бизнесе.