

РТУ МИРЭА

Создание системы для OLAP – кубов

Актуальность

- Необходимость компаниям быстро внедрять аналитику и принимать решения на основе данных, в том числе анализ сценарное моделирования финансового результата компании с помощью тесно интегрированных моделей финансового и инвестиционного планирования.
- Аналитическая обработка данных в реальном времени.
- Системы OLAP предоставляют интерактивный и удобный интерфейс для оперативной отчетности и анализа. Аналитики могут быстро анализировать данные, применять фильтры и создавать собственные отчеты, не полагаясь на навыки ИТ или программирования.
- Организациям эффективно обрабатывать огромные объемы данных.
- Хранение данных в предварительно агрегированной форме, что делает их эффективными для анализа больших наборов данных.

Гипотезы решения предполагаемой задачи

Создание своей системы для формирования OLAP-кубов с формой хранения в виде индексов каждого из полученных значений.

Технологический подход:

- Использование для хранения данных куба в `postgres sql` всей информации об измерениях куба в одной сложной координате (отображение n -мерного пространства на прямую) позволяет получать любые срезы (даже состоящие из объединения ортогональных плоскостей) за минимально время – время одного запроса получения (`select`).
- Использование библиотеки `pandas` (на текущий момент одна из самых проработанных библиотек для анализа данных) и работа не с плоскостями при построении агрегаций, а с гиперплоскостями позволяет получать искомые значения куба за время, сопоставимое с временем загрузки этих данных в базу.

Скорость расчетов и получения:

- Возможность расчет больших объемов данных за короткий срок.
- Возможность получения данных в агрегированном виде за скорость запроса в базу данных.

Функциональная доработка и техническая поддержка:

- Возможность добавления функционала для специализированных организаций.
- Обеспечение постоянной технической поддержки и обновлений.
- Внедрение системы обратной связи для оперативного реагирования на отзывы и запросы пользователей.

Концепция

Разработка платформы: Для разработки платформы необходимо подготовить визуальную часть, в которой будут отображены агрегированные данные полученные из базы данных, для этого будет использоваться фреймворк языка JavaScript, React, для расчетов был выбран язык программирования Python использование библиотеки pandas (на текущий момент одна из самых проработанных библиотек для анализа данных) и работа не с плоскостями при построении агрегаций, а с гиперплоскостями позволяет получать искомые значения куба за время, сопоставимое с временем загрузки этих данных в базу. В качестве базы данных был выбран PostgreSQL.

Способ хранения агрегированных данных:

Для хранения каждого из значений сформированного куба необходимо составлять индекс. Данный индекс будет хранить в себе информацию о том к какому измерению, к каким данным и к какому уровню иерархии в измерении он относится.

Преимущества

- Кубы OLAP представляют собой предварительно агрегированные структуры данных. Они хранят обобщенные данные, что повышает производительность запросов по сравнению с запросом необработанных данных транзакций. Пользователи могут разрезать, детализировать и сводить данные за считанные секунды, даже при работе с большими наборами данных.
- Благодаря правильной индексации и оптимизированному выполнению запросов кубы OLAP могут обеспечить высокую производительность даже при работе с огромными наборами данных, что делает их пригодными для аналитики на уровне предприятия.
- Кубы OLAP обеспечат согласованность отчетов и анализа. Поскольку кубы построены на основе определенной модели данных, все пользователи, обращающиеся к кубу, получают согласованные результаты. Это гарантирует, что лица, принимающие решения во всей организации, имеют общее понимание данных, что приводит к улучшению сотрудничества и принятия решений.

Перспектива коммерциализации

- Актуальность проекта в более быстром формировании OLAP-куба и способе его представления
- Переход на свою платформу и ее поддержки.

План монетизации

- Разработка и последующая продажа различным организациям

Перспективы роста

- Возможность добавлять специальные функции с сборку OLAP-куба
- Возможность доработки под нужды компании
- Возможность использовать платные подписки при использовании специальных доработок

Смета расходов для MVP

1. Подготовка (1 месяц)

- 1.1 Поиск frontend-разработчика со знаниями JavaScript (200000 руб.)
- 1.2 Поиск backend-разработчика со знаниями Python (200000 руб.)
- 1.3 Поиск системного аналитика (100000 руб.)
- 1.4 Специалист по обучению (30000 руб.)
- 1.5 Техническая поддержка (30000 руб.)

2. Разработка (1 год)

- 2.1 Настройка сервера (1 месяц)
- 2.2 Создание wsgi-сервера (1 месяц)
- 2.3 Создание визуальной составляющей платформы (3-4 месяца)
- 2.4 Создание архитектуры базы данных (1 месяц)
- 2.5 Создание функциональной части (8 месяцев)

3. Тестирование (3 месяца)

- 3.1 Тестирование продукта (3 месяца)

Итого: 560 000 руб.

Полная смета расходов

Этап 1: Анализ и проектирование (2)

- 1.1. Анализ требований и подготовка проекта
Системный аналитик: 200 000 руб. (2 месяца)
- 1.2. Настройка сервера
Бэкенд-разработчик: 400 000 руб. (2 месяца)

Этап 2: Разработка (9 месяцев)

- 2.1. Разработка платформы
Фронтенд-разработчик: 600 000 руб. (3 месяца)
Бэкенд-разработчик: 600 000 руб. (3 месяца)

Этап 3: Интеграция и тестирование (4 месяца)

- 3.1. Интеграция с существующими системами
Бэкенд-разработчик: 400 000 руб. (2 месяца)
- 3.2. Тестирование и оптимизация
Бэкенд-разработчик : 200 000 руб. (2 месяца)

Этап 3: Запуск и поддержка (3 месяца)

- 3.1. Обучение и поддержка пользователей
Специалист по обучению: 30 000 руб. (1 месяц)
Техническая поддержка: 90 000 руб. (3 месяца)

Итого: 3 720 000 руб.

Этот график не включает накладные расходы, такие как аренда офиса, оборудование, и дополнительные расходы, которые могут возникнуть в ходе проекта

СПАСИБО ЗА ВНИМАНИЕ!

Миронов Д.С.

РТУ МИРЭА, Институт информационных технологий