

Описание кейса

Компания получает данные из нескольких источников, формирует бизнес-метрики и использует AI агентов для автоматизации процессов. Необходимо спроектировать и реализовать базовый data и AI пайплайн

Часть 1.

Есть REST API по адресу
<https://api.example.com/ads>

API возвращает объекты со следующими полями

date
campaign_id
spend
clicks

Необходимо реализовать Python скрипт, который выполняет следующие действия:

Получает данные за последние 7 дней
Корректно обрабатывает ошибки API
Логирует статусы выполнения
Сохраняет данные в CSV файл или в базу данных

Часть 2.

Есть таблица orders со следующими полями:

order_id
user_id
order_date
revenue

Необходимо написать SQL запросы, которые рассчитывают следующие показатели:

Выручка по дням
Количество заказов по дням
ARPU
Количество повторных заказов

Часть 3.

Необходимо описать DAG в Apache Airflow, который состоит из следующих этапов:

Получение данных из рекламного API
Получение данных о заказах из базы данных
Трансформация и расчет ежедневных метрик
Загрузка агрегированных данных в хранилище

Необходимо настроить:

Расписание запуска
Ретраи
Логирование и обработку ошибок

Часть 4.

Необходимо описать структуру таблицы `daily_business_metrics`

Таблица должна содержать следующие поля:

`date`

`revenue`

`orders_count`

`ad_spend`

`roi`

Необходимо описать формулы расчета каждого показателя

Часть 5.

Необходимо описать концепцию AI агента, который использует данные из витрины `daily_business_metrics` и выполняет одну из следующих бизнес задач:

Контроль аномалий в выручке

Контроль эффективности рекламы

Уведомление о падении ROI

Формирование ежедневных бизнес отчетов

Нужно описать:

Какие данные использует агент

Как часто он работает

Какие решения он принимает

Куда отправляет результат работы

Нужно разработать критерии оценки работоспособности агента.