



# Финальное задание по SQL

Описание БД, концептуальная схема

---

Белогорцев Михаил

Магистратура ВШЭ

Инженерия данных

## Описание компании

Есть компания XYZ, которая занимается производством, розничной и оптовой продажей хобби-электроники и робототехнических деталей.

У нее есть разные каналы продаж и используются разные информационные системы для их учета. Например, есть личные кабинеты на маркетплейсах (Ozon, WB), есть внутренняя ERP (для складского и производственного учета), есть CRM система (для ведения информации о клиентах), есть 1С потому что без нее нельзя.

Постоянно добавляются новые каналы продаж, сущности, новые маркетинговые инструменты и у компании нет возможности и желания постоянно дорабатывать внутренние ИС для того чтобы работать в одной единой системе, которая бы отражала всю информацию.

При этом бизнесу нужно получать ответы на конкретные вопросы, чтобы принимать решения и видеть результат тех или иных действий.

Данные содержащиеся в разных ИС тем не менее структурированы, сущности имеют уникальные идентификаторы и могут быть объединены в виде реляционной БД.

## Цель:

1. Создать единую БД для аналитических целей, куда будут загружать данные из разных ИС, где в итоге будет содержаться информация обо всех произведенных товарах, их себестоимости, продажах и клиентах компании.

## Преимущества создания единой БД:

1. БД обеспечит аналитикам компании возможность отвечать на вопросы бизнеса за счет создания отчетов, витрин данных, которые нужно создать только один раз, а использоваться будут регулярно.
2. Даст возможность настройки автоматических загрузок и обновлений данных.
3. Сэкономит время аналитиков за счет отсутствия ручных выгрузок csv, excel файлов из разных систем.

## Требования к данным и транзакциям

### Требования к данным:

#### Полнота данных:

- В БД должны загружаться все данные, необходимые для анализа, включая данные из ERP, CRM, маркетплейсов, 1С и других используемых систем.
- Данные должны включать информацию о товарах (артикулы, наименования, категории, себестоимость), данные о продажах (канал продаж, дата, сумма, количество), информацию о клиентах (идентификаторы, категории, история покупок), информацию о складских остатках (локации, количество, движение товара).

#### Актуальность данных:

- Данные в БД должны регулярно обновляться, чтобы обеспечивать возможность анализа на основе актуальной информации.

#### Качество данных:

- Все данные должны быть очищены и стандартизированы перед загрузкой в БД.
- Устранение дубликатов, проверка целостности данных, преобразование форматов (например, даты и валюты).
- Использование единой системы идентификаторов для товаров, клиентов, складов и каналов продаж.

### Требования к транзакциям:

#### ACID-свойства:

- База данных должна обеспечивать выполнение транзакций с соблюдением всех ACID-свойств.

#### **Производительность:**

- Транзакции по обновлению данных должны быть оптимизированы, чтобы минимизировать время выполнения.
- Обеспечение быстрой вставки данных для больших объемов (bulk insert).

#### **Безопасность:**

- Транзакции должны выполняться с учетом прав доступа.

#### **Историчность данных:**

- Обеспечение сохранения исторических данных о продажах, ценах и движении товаров.
- Использование механизмов версионности для критически важных данных.

## **Типичные запросы:**

#### **Анализ продаж:**

- Как изменились продажи за последние 6 месяцев по различным каналам (Ozon, WB, собственный сайт, опт) в процентах по отношению к первому месяцу.
- Какие товары обеспечивают наибольшую маржинальность (разницу между себестоимостью и продажной ценой) за последний год с разбивкой по каналам продаж.
- Сколько возвратов было по каждому товару и каналу продаж за последний месяц

#### **Маркетинг:**

- Сколько новых клиентов появилось за последний месяц в каналах: опт, розница.

