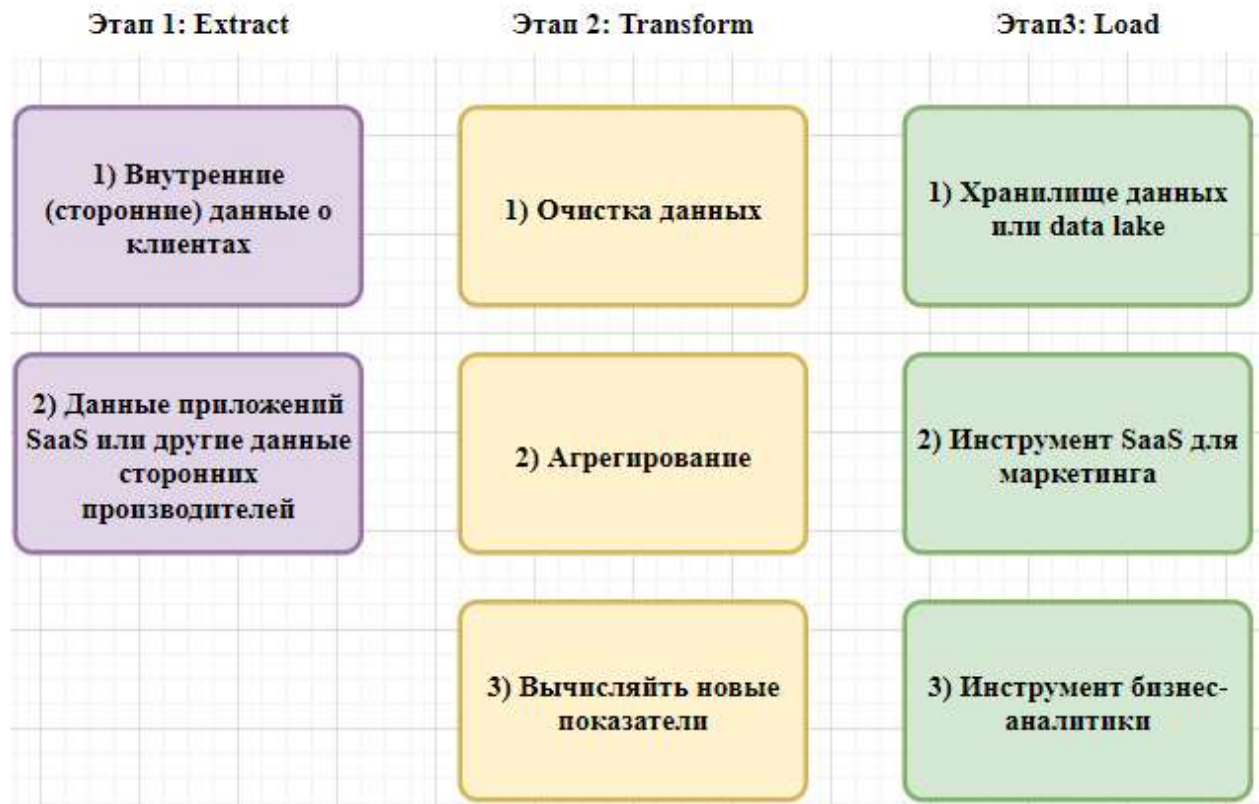


Задание 1.

Нарисуйте архитектуру ETL процесса для сбора и анализа данных компанией которая хочет провести маркетинговую кампанию, используя app.diagrams.net. Сделайте описание почему вы считаете что архитектура должна выглядеть именно так.



Этап 1: Извлечение данных

Каждая маркетинговая команда использует данные из разных источников:

- 1) Внутренние (сторонние) данные о клиентах: к ним относятся файлы с информацией о клиентах (CSV, Microsoft Excel, Google Sheets и т.д.), наборы данных с покупками клиентов и информацией о доставке (например, в локальной реляционной базе данных SQL или облачном хранилище, таком как Amazon S3, для транзакций электронной коммерции).
- 2) Данные приложений SaaS или другие данные сторонних производителей: В эту категорию входят электронные письма клиентов для CRM-платформ, таких как Hubspot и Salesforce, рекламные данные с Facebook ads и Google ads, взаимодействия с веб-сайтами из Google Analytics и информация о продуктах поставщика услуг электронной коммерции.

Этап 2: Преобразование данных

Преобразование данных включает в себя следующие этапы:

- 1) Очистка данных: Вы удаляете дублированные или поврежденные записи, отфильтровываете информацию, которая не нужна для маркетинговой деятельности, а также ошибки и несоответствия из ваших наборов данных.
- 2) Агрегирование данных: данные, которые собираются на этапе 1, часто слишком детализированы. Например не нужно, чтобы каждое событие, связанное с кликом клиента, демонстрировалось отдельно. Вместо этого нам нужно агрегированное общее количество кликов, чтобы определить, какая реклама показала наилучшие результаты.
- 3) Вычислять новые показатели: данные, которые извлекаются из различных источников, не всегда оптимизированы для анализа. Чтобы решить эту проблему, необходимо рассчитать свои маркетинговые показатели, такие как ARPU или пожизненная ценность клиента.

Этап 3: Загрузка данных

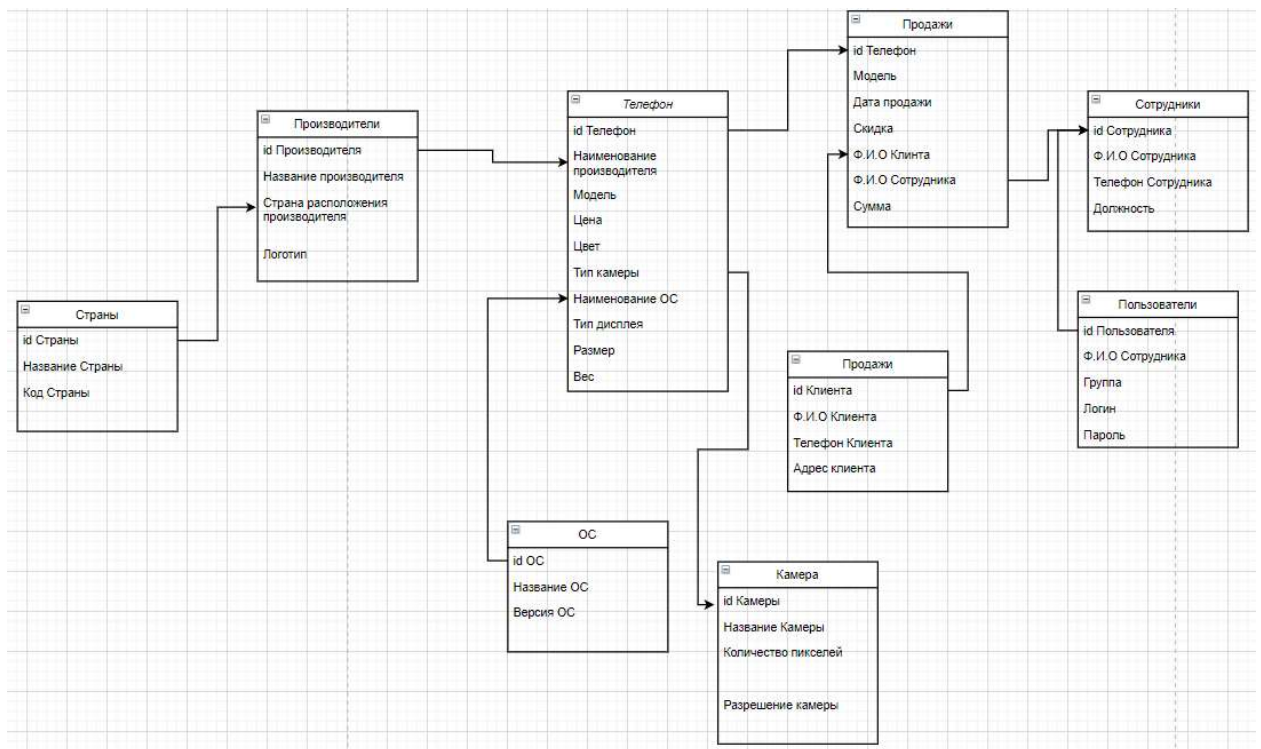
Загрузка данных, проще говоря, означает сохранение данных. Ваши варианты включают:

- 1) Хранилище данных или data lake: Наилучший способ переместить очищенные данные в хранилище данных. Наличие данных в Snowflake или BigQuery поможет создать “единый источник достоверности”. Единое хранилище для всех ваших данных о клиентах, которое обеспечивает согласованность ваших данных. Хранилище данных также позволяет членам команды получать доступ к одним и тем же данным без ущерба для согласованности данных или дублирования работы с ETL.
- 2) Инструмент SaaS для маркетинга: используйте свои данные в инструменте SaaS, чтобы подготовить их к маркетинговым кампаниям. Например, когда создается набор данных, включающий всех, кто приобрел ваши последние продукты, вы можете направить этот набор данных в Hubspot (список адресов электронной почты) или Facebook Ads (аудитории) для проведения гипертаргетированных кампаний
- 3) Инструмент бизнес-аналитики: загружая данные в такие инструменты, как PowerBI, Tableau или Google Data Studio, данные сразу же готовятся к анализу. Создавайте информационные панели с маркетинговыми показателями и ключевыми показателями эффективности, чтобы выявлять тенденции.

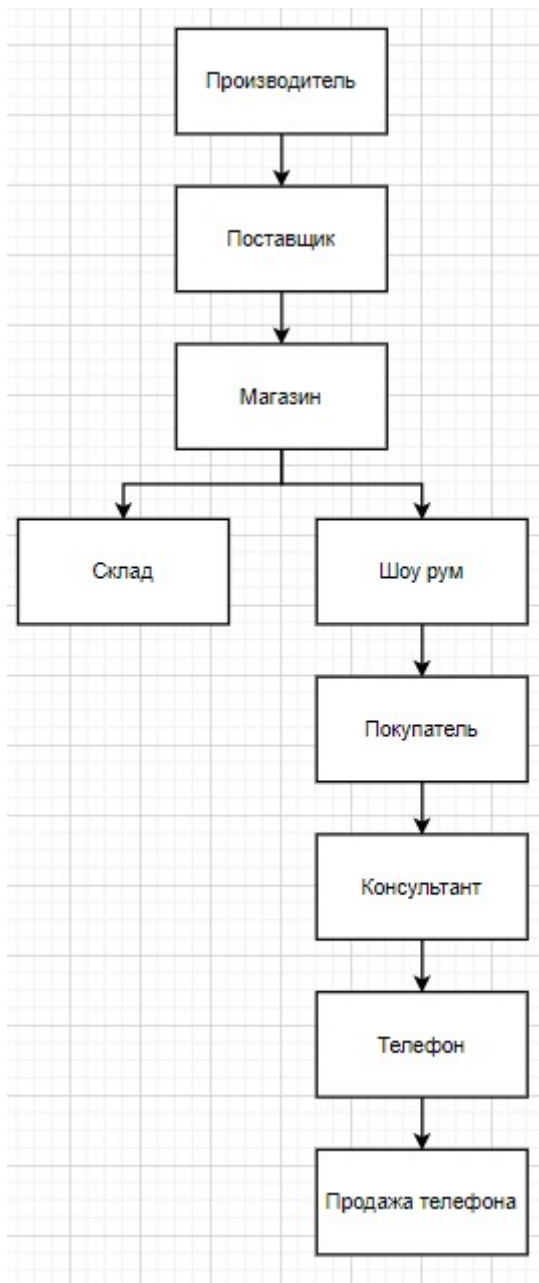
Задание 2.

Постройте реляционную и иерархическую модели данных для магазина который продает телефоны.

Реляционная



Иерархическая



Задание 3.

Определите в какой нормальной форме данная таблица, приведите её ко 2 и 3 нормальной формам последовательно.

Employee_ID	Name	Job_Code	Job	City_code	Home_city
E001	Alice	J01	Chef	26	Moscow
E001	Alice	J02	Waiter	26	Moscow
E002	Bob	J02	Waiter	56	Perm
E002	Bob	J03	Bartender	56	Perm
E003	Alice	J01	Chef	56	Perm

Таблица находится в форме NF1.

NF2:

Database management tool interface showing a table with the following data:

Employee_ID	Name	City_code	Home_city
E001	Alice	26	Moscow
E002	Bob	56	Perm
E003	Alice	56	Perm

NF3:

Database management tool interface showing a table with the following data:

Job_Code	Job
301	Chef
302	Waiter
303	Bartender