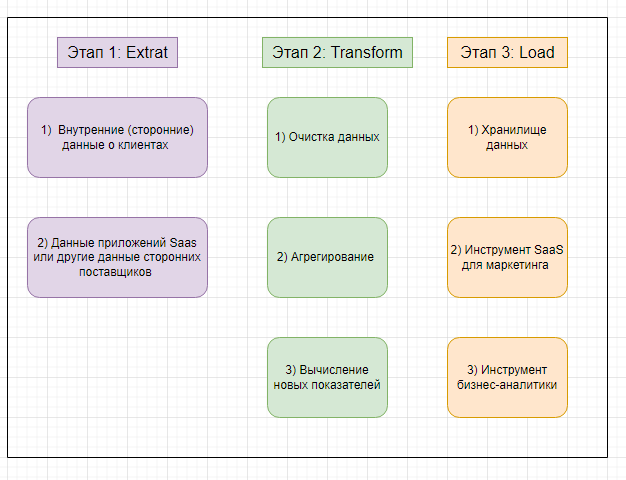
ETL: автоматизация подготовки данных

Урок 1. Модели данных и нормализация таблиц. Схема «звезда»

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

На первом этапе данные извлекаются из различных источников, будь то база данных, файлы или внешние системы:

1. Внутренние (сторонние) данные о клиентах, к ним относятся файлы с информацией о клиентах (CSV, Microsoft Excel и т.д.), наборы данных о покупках клиентов и с информацией о доставке (Amazon S3).
2. Данные приложений SaaS или другие данные сторонних поставщиков (электронные письма клиентов для CRM–платформ, рекламные данные, информация о продуктах поставщика услуг электронной коммерции).

Этап 2: Преобразование данных

Полученные данные проходят через процесс трансформации, где они подвергаются различным операциям:

1. Очистка данных. Удаляются дублированные или поврежденные записи, отфильтровывается информация, которая не нужна для проведения маркетинговой компании, а также ошибки и несоответствия из полученных наборов данных.
2. Агрегирование данных: данные, которые собираются на Этапе 1, бывают часто слишком детализированы. Не нужно, чтобы каждое событие, связанное с кликом клиента, показывалось отдельно. Вместо этого нам необходим агрегированное общее количество кликов, чтобы определить, какая реклама показала наилучшие результаты.

Этап 3 Загрузка данных

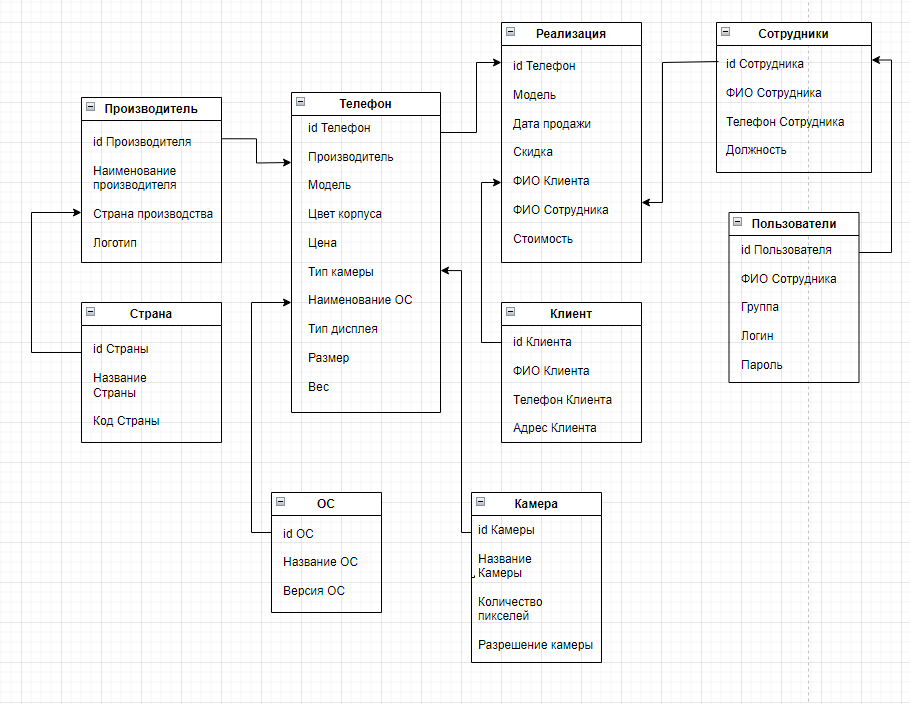
Преобразованные данные загружаются в целевую систему или хранилище данных, готовые для дальнейшего анализа и использования:

1. Хранилище данных для всех полученных данных о клиентах, которое обеспечивает согласованность данных. Хранилище данных позволяет членам маркетинговой команды получать доступ к одним и тем же данным, без ущерба для согласованности данных или дублирования работы с ETL.
2. Инструмент Saas для маркетинга: использование полученных данных в данном инструменте поможет подготовить их к проведению маркетинговой компании. Например, создается набор данных, который включает всех, кто приобрёл последние продукты. Этот набор данных можно использовать для проведения гипертаргетированной кампании.
3. Инструмент бизнес-аналитики. Загрузив данные в такой инструмент как PowerBI (или в Tableau) сразу готовим их к анализу. Создавая информационные панели с маркетинговыми показателями и ключевыми показателями эффективности, чтобы выявить тенденции.

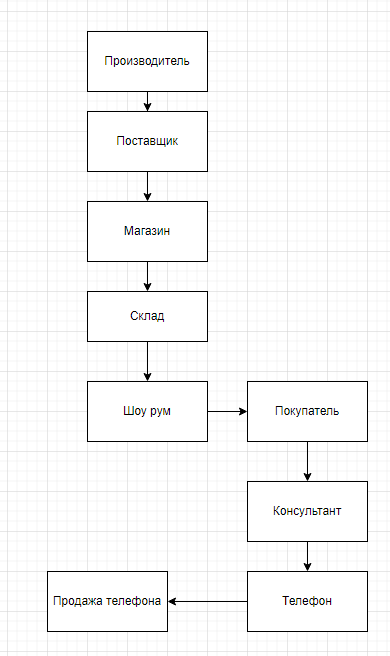
Задание 2.

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

Реляционная модель данных:

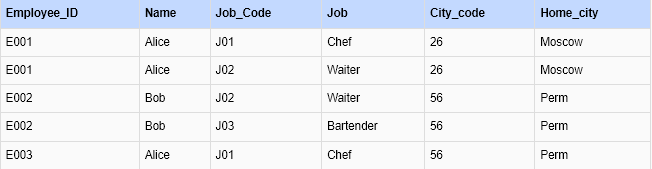


Иерархическая модель данных:

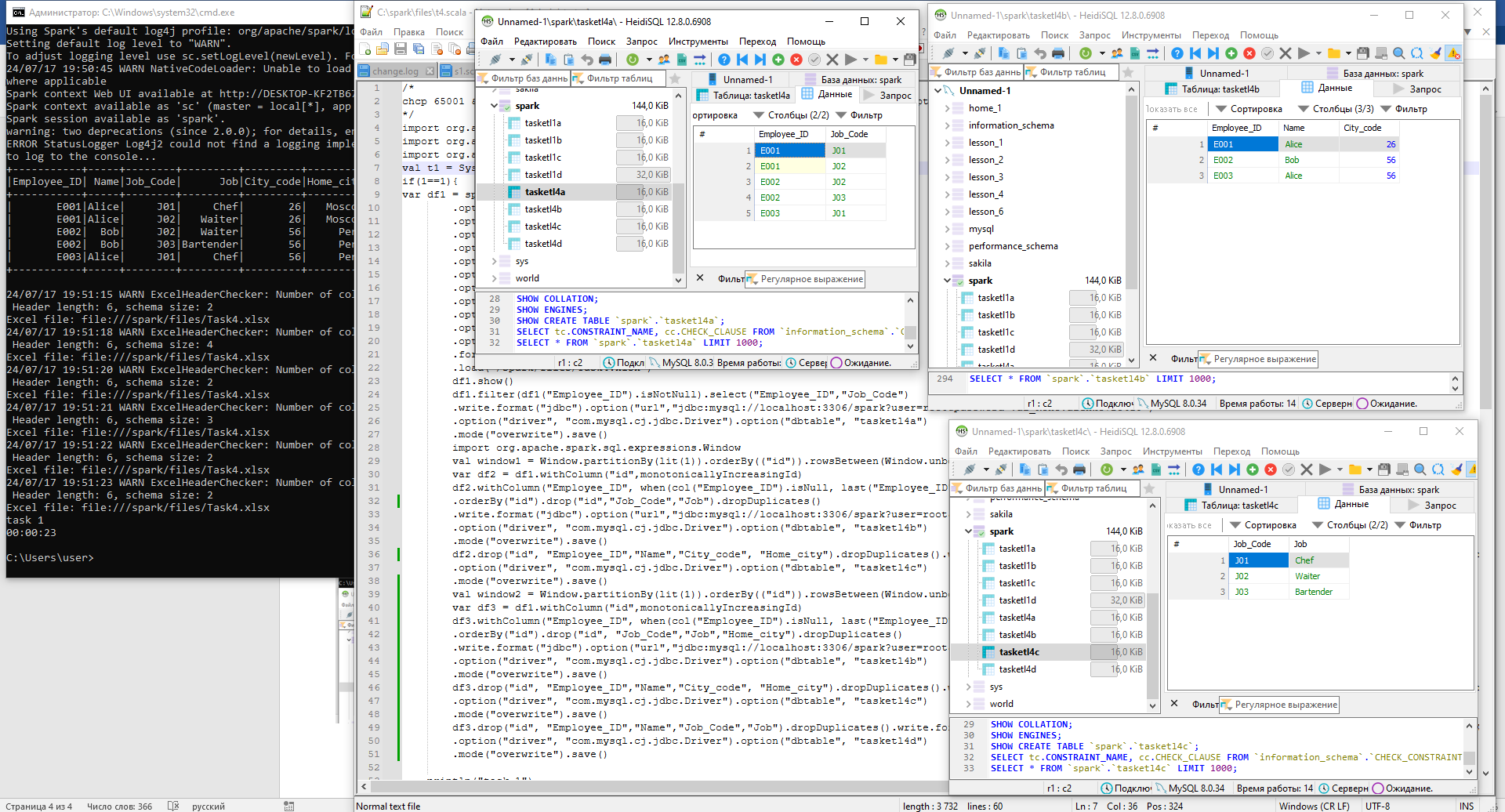


Задание 3.

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



2 Нормальная форма:



3 Нормальная форма:

