

Лекция 2. Практическое задание

Курс: разработчик ХД



Copyright © 2018–2025 by DataTech. All Rights Reserved.





Практическое задание на курс (1 из 2)

Практическое задание на курс: создание собственного миниатюрного хранилища данных для сети продовольственных магазинов шаговой доступности.

Требование заказчика №1: необходимо реализовать витрину для анализа маржинальности продаж.

- **Измерения:** чем больше, тем лучше, но не идентификаторы, а текстовые наименования.
- **Метрики:**
 - Количество проданного товара в штуках.
 - Сумма продаж в рублях.
 - Маржинальность в процентах $((\text{стоимость продажи} - \text{ себестоимость}) / \text{стоимость продажи} * 100)$ с округлением до 2 знаков после запятой.

Требование заказчика №2: необходимо реализовать пользовательскую витрину с **матрицей** продаж в рублях по месяцам за 2022 и 2023 годы для оценки эффективности работы точек продаж.

- **Измерение:** торговая точка (название магазина).
- **Метрика:** суммарные продажи товаров в разбивке по месяцам в отдельных колонках.
- Размеры торговых площадей, стоимость аренды, плотность населения и т.п. в расчет не берем.



Практическое задание на курс (2 из 2)

- Постановка задачи на курс:
 - Проработать **архитектуру** данных.
 - Создать **логическую** модель будущего ХД.
 - Создать **физическую** модель ХД на технологиях СУБД Greenplum.
 - Реализовать процесс **доставки** данных из источников и их **загрузки** в ХД как средствами СУБД, так и внешним инструментом Apache Airflow.
 - Построить пользовательские **бизнес-витрины** данных.
 - Добраться приемлемой **производительности** работы хранилища.
- **Подсказка:** зайдите в ближайший магазин, пройдитесь по нему. Проанализируйте его деятельность в комплексе и ответьте самому себе на следующие **вопросы** (и другие, если сочтете их важными):
 - Какой ассортимент товаров продает магазин, и почему товары расставлены на полках именно так?
 - Откуда товары попадают на полки? Какие внешние факторы могут влиять на ассортимент?
 - Кто является покупателями, и что подтверждает факт купли-продажи?
 - Как и почему регулярно варьируются цены на товары?
 - Какова география распространения точек продаж этой Сети в нашей стране?

Лекция 2: Практическое задание

1. Определите **архитектуру** данных будущего хранилища:
 - Состав **слоев**.
 - Тип логической **модели** для каждого слоя (типы моделей могут отличаться).
2. Разработайте логическую модель **детального** (интеграционного) слоя данных ХД.
 - Начните работу с создания концептуальной **бизнес-модели**: сущности и связи между ними. Основную информацию для этого вам поможет предоставить поход в магазин и анализ его деятельности.
 - Дополните модель **атрибутами** сущностей, первичными и внешними ключами.
 - **Опционально** укажите другие ограничения колоночного уровня, если сочтете необходимым.
3. Задание 3. Разработайте логическую модель пользовательских **витрин** согласно требованиям.
 - Результаты практического задания лекции 2 принимаются в файлах форматов .jpg, .png (картинки), и предпочтительно .docx, .pptx, (фигуры и значки MS Office) или .drawio.
 - Если платформа обучения не принимает файл вашего формата, заархивируйте его.



Вносите в ЛМД только сущности и атрибуты, которые, по вашему мнению, напрямую влияют на бизнес торговой сети.