**Практическая работа: Анализ данных интернет-магазина с использованием NoSQL-подхода**

**Тема:** Проектирование гибкой модели данных и анализ продаж с использованием документно-ориентированной СУБД

**Цель практики:**

* Научиться проектировать модель данных для NoSQL-хранилища на основе бизнес-требований.
* Освоить базовые операции преобразования данных (CSV to JSON) и их загрузки.
* Получить навыки написания агрегационных запросов для аналитики средствами NoSQL-СУБД.
* Сравнить эффективность аналитической обработки в SQL и NoSQL-подходах.

**Условие (контекст):**  
Вы — data-аналитик в онлайн-магазине электроники. Компания решила перевести часть данных из реляционной БД (MySQL) в документную NoSQL-БД (MongoDB) для ускорения формирования аналитических отчетов. Вам поручено спроектировать модель хранения данных о заказах и товарах, подготовить данные и выполнить аналитические запросы.

**Исходные данные:**  
Вам предоставлены два CSV-файла:

1. orders.csv — содержит данные о заказах: order\_id, customer\_id, order\_date, total\_amount.
2. order\_items.csv — содержит данные о товарах в заказах: item\_id, order\_id, product\_id, product\_name, quantity, price.

**Задачи:**

**1. Проектирование модели данных (Class Diagram)**  
Спроектируйте UML-диаграмму классов, отражающую документную модель данных для хранения заказов в MongoDB. Учтите, что в документной БД данные о заказе и его товарах часто хранятся в одном документе (денормализация). Опишите, какие embedding и referencing вы будете использовать.

**Ожидаемый результат:**

* Диаграмма с классами Order и Product.
* Пояснение, почему выбран такой способ хранения (например, для ускорения чтения данных о заказе).

**2. Преобразование данных и загрузка (Data Preparation)**  
Напишите скрипт (на Python или укажите логику), который:

* Считывает данные из orders.csv и order\_items.csv.
* Преобразует их в коллекцию JSON-документов, где каждый документ представляет собой заказ со вложенным массивом товаров.
* Сохраняет результат в файл orders.json.

**Пример фрагмента JSON-документа:**

json

{

"order\_id": 1001,

"customer\_id": 201,

"order\_date": "2023-10-25",

"total\_amount": 1500.00,

"items": [

{

"product\_id": 301,

"product\_name": "Смартфон",

"quantity": 1,

"price": 1500.00

}

]

}

**3. Аналитические запросы (Aggregation Pipeline)**  
Напишите запросы агрегации в MongoDB для решения следующих задач:

1. Найти топ-5 самых продаваемых товаров за последний месяц.
2. Посчитать общую выручку за указанный период.
3. Определить средний чек заказа.

**Пример запроса:**

javascript

*// Общая выручка за октябрь 2023*

db.orders.aggregate([

{

$match: {

order\_date: {

$gte: ISODate("2023-10-01"),

$lte: ISODate("2023-10-31")

}

}

},

{

$group: {

\_id: null,

total\_revenue: { $sum: "$total\_amount" }

}

}

])

**4. Предложить оптимизацию**  
Предложите, как можно оптимизировать модель данных или запросы, если:

* Отчеты стали строиться медленнее из-за роста данных.
* Появилась необходимость учитывать возвраты товаров (используйте знания из предыдущих практик).

**Ожидаемые результаты оптимизации:**

* Возможное введение индексов на поля order\_date и product\_id.
* Предложение по использованию колоночной БД (например, ClickHouse) для исторических данных.

**5. Сравнение с SQL**  
В виде таблицы сравните, как бы решались те же аналитические задачи в SQL и в MongoDB.

| **Задача** | **SQL-запрос** | **MongoDB Aggregation** |
| --- | --- | --- |
| Топ-5 товаров | SELECT product\_name, SUM(quantity) FROM order\_items ... GROUP BY product\_id | [$group, $sort, $limit] |
| Общая выручка | SELECT SUM(total\_amount) FROM orders WHERE ... | [$match, $group] |
| Средний чек | SELECT AVG(total\_amount) FROM orders | [$group] |

**Критерии оценки:**

* Правильность проектирования модели данных (учет денормализации, связей).
* Корректность преобразования CSV в JSON (полнота данных, структура).
* Точность агрегационных запросов (использование $match, $group, $sort).
* Логичность предложений по оптимизации и сравнению подходов.

Это задание позволяет применить знания о NoSQL на практике: от моделирования и конвертации данных до аналитических запросов и сравнения с SQL. Оно соответствует формату ваших образцов и фокусируется на практических аспектах работы аналитика с данными.