Орищенко Виталий

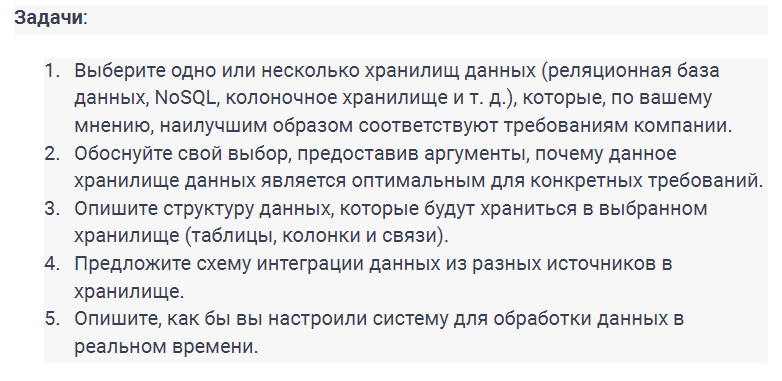
@Vitosh992

Группа 1

**Архитектор ИИ**

**Задание 14**





С учетом вводных данных можно сделать выбор в сторону NoSQL-баз данных.

**Причины:**

* Легкое горизонтальное масштабирование, подходит для большого объема данных. Учитывая, что по условиям у нас миллионы записей ежедневно, NoSQL по этому критерию подойдет.
* Учитывая техническую специфику NoSQL, а, точнее, их работу с неструктурированными данными, это позволяет более гибко проводить различные преобразования. Это как раз и удовлетворяет критерии аналитики в условиях.
* NoSQL могут масштабироваться и имеют разные инструменты для оптимизации работы БД с учетом изменения данных.
* NoSQL прекрасно справляются с выполнением задач в режиме реального времени, что также обеспечивает удовлетворение условий.

**Пример структуры данных:**

В рамках NoSQL можно выбрать СУБД MongoDB. С учетом наших условий и выбранной архитектуры БД можно создать следующие коллекции данных:

«sales» для хранения данных о продажах,

«customers» для информации о клиентах,

«inventory» для информации об инвентаре.

Каждая коллекция будет иметь свои поля и индексы. MongoDB не позволяет создавать связи между коллекциями, как на реляционных БД. Соответственно, основная нагрузка на взаимодействие с БД будет реализована непосредственно на ПО.

**Интеграция данных из разных источников:**

Можно использовать один из инструментов ETL, например, Apache NiFi, который позволит автоматизировать процесс преобразования данных в желаемый формат.

**Настройка системы для обработки данных в реальном времени:**

Для обработки данных в реальном времени будем использовать Apache Kafka (как самый распространенный инструмент). Он позволит нам обрабатывать потоковые данные и отслеживать заказы и складскую информацию и уведомлять о событиях в реальном времени.