1. ***Выберите одно или несколько хранилищ данных (реляционная база данных, NoSQL, колоночное хранилище и т. д.), которые, по вашему мнению, наилучшим образом соответствуют требованиям компании***

Гибридное хранилище данных.

1. PostgreSQL для хранения структуры торговой площадки, личных кабинетов, покупателей, товаров и их количества, заказов, маркетинговых акций и скидок, и т.д. (почти все CMS для e-commerce реализованы MySQL).

2. Колоночную БД, ClickHouse, для сбора логов статистики работы сайта, действий пользователя. Аналитики, (например, брошенных корзин, наиболее просматриваемых товаров, прогнозирования закупок).

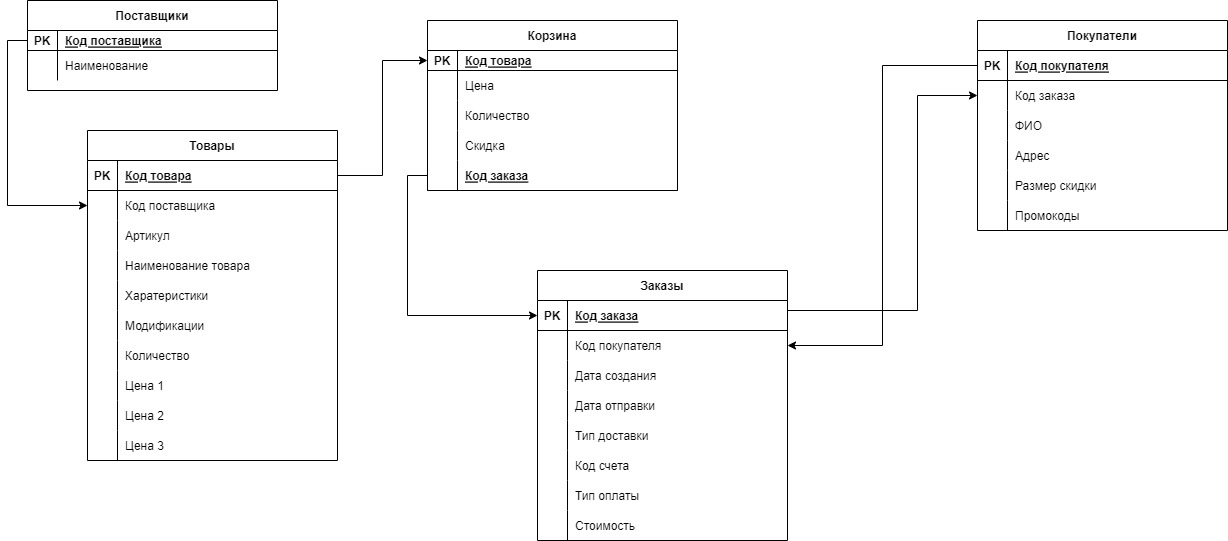
1. ***.Обоснуйте свой выбор, предоставив аргументы, почему данное хранилище данных*** *является оптимальным для конкретных требований.*

В части применения функционала обоснование приведено в п.1. В части удобства использования, обе БД имеют синтаксис SQL и могут работать с большими данными.

Обе БД масштабируются шардированием и работают в кластерах.

1. ***Опишите структуру данных, которые будут храниться в выбранном хранилище (таблицы, колонки и связи)***

Пример небольшой части формирования заказа покупателя



1. ***Предложите схему интеграции данных из разных источников в хранилище***

Информация из внешних источников данных может быть интегрирована с помощью различных сервисов, подключаемых к API источников данных (например, СПАРК для проверки поставщиков, ОФД или облачная касса). При очень большом количестве данных (например, розничных продаж) можно использовать Apache Kafka, для отправки информации о платежах в ОФД и подтверждении платежа покупателю.

1. ***Опишите, как бы вы настроили систему для обработки данных в реальном времени***

Отслеживание количества покупок, оплат, суммы в режиме реального времени можно реализовать подключением Apache Kafka к ClickHouse (движок Kafka в ClickHouse). Настроить разнообразные витрины и подключить средства визуализации (Apache Superset, Visiology и т.д.)