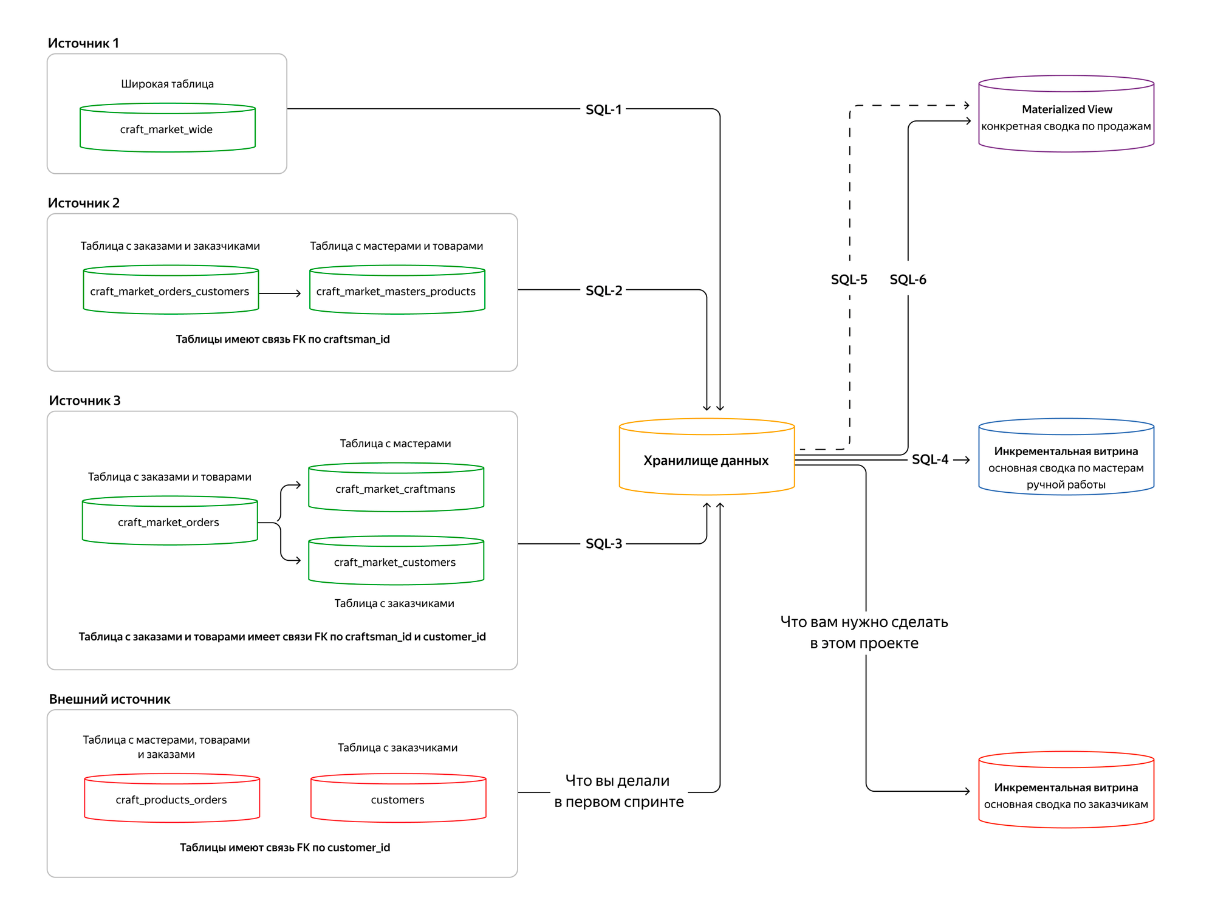
**Проект**

Важная особенность корпоративных хранилищ — возможность интегрировать новые источники. В этом проекте вы как раз и добавите новый источник в DWH.

Маркетплейс товаров ручной работы набирает популярность и расширяет свою аудиторию. Чтобы привлечь больше покупателей, приобрели ещё один сайт, который предоставляет похожие услуги. Формальная часть сделки завершена, нужно интегрировать данные купленного сайта в хранилище.

Добавление нового источника — тот же процесс, что и загрузка предыдущих источников: нужно интегрировать данные и построить новую витрину. Вы уже строили инкрементальную витрину с отчётом по мастерам, теперь постройте похожую витрину с отчётом по заказчикам. Она пригодится, чтобы внедрить программу лояльности, изучить интересы пользователей, сделать более релевантной систему рекомендаций товаров и скорректировать маркетинговый план по акциям.

Общая схема проекта выглядит следующим образом:



Бо́льшую часть работы вы сделали, осталось только загрузить данные нового источника в хранилище и построить витрину.

**Инструкция по выполнению**

**Шаг 1. Подключение к базе**

Вам понадобится та же БД, с которой вы работали в течение всего спринта. Напомним схему подключения, таблицы и модель данных:

* В схеме dwh — таблицы измерений для мастеров ручной работы (d\_craftsman), заказчиков (d\_customer) и товаров (d\_product). Также там есть таблица фактов о продажах (f\_order), инкрементальная витрина по мастерам за отчётные периоды (craftsman\_report\_datamart) и MATERIALIZED VIEW с отчётом продаж по всему маркетплейсу (orders\_report\_materialized\_view).
* В схеме source1 — широкая таблица первого источника craft\_market\_wide.
* В схеме source2 — две ненормализованные таблицы: таблица мастеров и товаров (craft\_market\_masters\_products) и таблица заказчиков и заказов (craft\_market\_orders\_customers).
* В схеме source3 — три таблицы: мастеров (craft\_market\_craftsmans), заказчиков (craft\_market\_customers), заказов с товарами (craft\_market\_orders).
* Схема external\_source — новая. Там как раз находятся таблицы нового источника, которые нужно подключить к хранилищу.

**Шаг 2. Изучение данных нового источника**

Изучите данные в источнике. Там указаны две таблицы: в одной данные по мастерам, товарам и заказам (craft\_products\_orders), а в другой — только по заказчикам (customers). Правильно разложите данные по измерениям в хранилище. Сделайте это так же, как вы делали в первом спринте.

Данные новой схемы external\_source должны оказаться в dwh в таблицах измерений и фактов.

**Шаг 3. Напишите скрипт переноса данных из источника в хранилище**

Загрузка данных из нового источника — операция не на один раз: какое-то время сайт всё ещё будет существовать, пока не будет выполнен ребрендинг. А это значит, что данные из источника нужно постоянно забирать и собирать в хранилище.

Данные должны попасть в схему dwh по тому же принципу, что и из схем source1, source2, source3. Дополните скрипт загрузки данных из старых источников новым источником.

Если нужно, изучите скрипт загрузки из старых источников.

Скрипт загрузки из старых источников

**Шаг 4. Изучите потребности бизнеса в новой витрине**

Изучите постановку задачи от бизнес-аналитика и исследуйте хранилище на соответствие данных. Посмотрите, какие данные понадобятся для создания DDL новой витрины.

Бизнесу требуются следующие данные:

* идентификатор записи;
* идентификатор заказчика;
* Ф. И. О. заказчика;
* адрес заказчика;
* дата рождения заказчика;
* электронная почта заказчика;
* сумма, которую потратил заказчик;
* сумма, которую заработала платформа от покупок заказчика за месяц (10% от суммы, которую потратил заказчик);
* количество заказов у заказчика за месяц;
* средняя стоимость одного заказа у заказчика за месяц;
* медианное время в днях от момента создания заказа до его завершения за месяц;
* самая популярная категория товаров у этого заказчика за месяц;
* идентификатор самого популярного мастера ручной работы у заказчика. Если заказчик сделал одинаковое количество заказов у нескольких мастеров, возьмите любого;
* количество созданных заказов за месяц;
* количество заказов в процессе изготовки за месяц;
* количество заказов в доставке за месяц;
* количество завершённых заказов за месяц;
* количество незавершённых заказов за месяц;
* отчётный период, год и месяц.

**Шаг 5. Напишите DDL новой витрины**

Напишите DDL для витрины, определив типы, названия полей и комментарии. Учитывайте, что витрина должна быть инкрементальной: понадобится дополнительная таблица с датой загрузки данных.

Витрину назовите dwh.customer\_report\_datamart. Возьмите за основу витрину по мастерам ручной работы. Новая витрина и витрина по мастерам будут во многом схожи.

**Шаг 6. Напишите скрипт для инкрементального обновления витрины**

Напишите скрипт, который будет выполнять расчёт для новой витрины. Это инкрементальная витрина, значит, и расчёт должен быть инкрементальным. За основу можно взять скрипт обновления инкрементальной витрины по мастерам.

Скрипт обновления витрины по мастерам

Чтобы получить идентификатор самого популярного мастера у заказчика, можно попробовать написать ещё один INNER JOIN по аналогии с определением самой популярной категории товаров. Самый популярный мастер определяется по количеству заказов от заказчика. Если заказчик сделал одинаковое количество заказов у нескольких мастеров, возьмите идентификатор любого. Для этого воспользуйтесь [командами](https://www.postgresql.org/docs/current/functions-window.html) DISTINCT ON (COLUMN) или FIRST\_VALUE.