

Setting

База данных для магазина интернет-коммерции (далее --- Компания). Кратко бизнес процесс: Компания предоставляет продавцам площадку для продажи товаров на едином сайте в интернете, пользователи их выбирают, оплачивают и получают доставку.

Наличие таковой довольно естественно в условиях высокотехнологичности бизнеса (существенная часть операционной деятельности ведётся в интернете). База данных поможет систематизировать и выделить все составляющие деятельности.

Описание предметной области совмещено с описанием атрибутов таблиц БД. Оно приведено ниже. Таблицы описываются в логичном и естественном для первого знакомства с областью порядке.

Физическая схема

Таблица orders

Центральная таблица, отпросится к операционной деятельности компании — посредничестве между продавцом и покупателями. Каждая строка — одна позиция заказа, или **item_id** (в рамках которой может быть K штук — см. атрибут **items_num**).

Чисто с продуктовой точки зрения (то есть с позиции пользователя **customer_id**) один заказ характеризует атрибут **multioorder_id**. Необходимость атрибута **order_id** объясняется разными способами и сроками доставки внутри одного чека, грубо говоря единый **order_id** — это одна «коробка» для части заказа. (Если все товары приехали одной «коробкой», то **multioorder** = **order**)¹.

Процесс передачи товара от продавца покупателю отслеживается по атрибутам **status** (\in 'created'², 'shipped', 'delivered', 'cancelled'), **created_dttm**, **shipped_dttm**, **delivered_dttm**. Здесь и далее все атрибута с типом **timestamp** полагаются предварительно приведёнными к Московскому времени GMT+3.

¹ К примеру, человек заказывает одним чеком айфон, макароны 2 пачки, и книгу. В таблице окажется 3 записи, в одной из них будет **items_count** = 2. Пусть айфон доставляет «барыга с Горбушки», то есть хранение и доставка на нём; а макароны и книга хранятся на складе Компании и доставляются ей же. В таком случае будет два разных **order_id** (в рамках одного **multioorder_id**). Пусть сразу после оформления пользователь вспомнил, что макарон нужно 3 пачки и дозаказал ещё одну. Тогда это будет другой мультиказ, и даже если в конце концов курьер привезёт пользователю 3 пачки вместе (очень вероятно если заказ сделан сразу же и если макароны те же), пусть даже в одной коробке — **order_id** будут разными.

² Статус **created** означает, что заказ оформлен и пользователь его оплатил.

Каждому товару на площадке Компании присваивается уникальный айди, именуемый **sku** — store keeping unit.

В современной коммерции скидки — главный инструмент привлечения. За айди применявшейся скидки отвечает атрибут `promo_id`.

Цена товара в рублях (которую видит пользователь, за которую он купил товар) отражена в атрибут `price`.

таблица	атрибут	тип	требования
orders	<code>item_id</code>	integer	non-null
orders	<code>items_num</code>	integer	non-negative
orders	<code>multiorder_id</code>	integer	non-null
orders	<code>order_id</code>	integer	non-null
orders	<code>customer_id</code>	integer	non-null
orders	<code>rus_name</code>	text	
orders	<code>status</code>	text	non-null
orders	<code>created_msk_dttm</code>	timestamp	<code>created_msk_dttm</code> <= <code>shipped_msk_dttm</code> <= <code>delivered_msk_dttm</code>
orders	<code>shipped_msk_dttm</code>	timestamp	<code>created_msk_dttm</code> <= <code>shipped_msk_dttm</code> <= <code>delivered_msk_dttm</code>
orders	<code>delivered_msk_dttm</code>	timestamp	<code>created_msk_dttm</code> <= <code>shipped_msk_dttm</code> <= <code>delivered_msk_dttm</code>
orders	<code>partner_id</code>	integer	non-null
orders	<code>warehouse_id</code>	integer	non-null
orders	<code>promo_id</code>	integer	non-null
orders	<code>price</code>	integer	positive otherwise null

Таблица **schemes**

Следующая таблица, требующая развёрнутых пояснений. Отражает схемы сотрудничества продавцов с Компанией в зависимости от `ownership`, `warehousing` и `delivery` товаров (атрибут **name**)

- 'O W D': храним и доставляем собственный товар,
- 'NO W D': храним чужой товар на нашем складе и доставляем его,
- 'NO NW D': доставка со склада партнёра силами Компании,
- 'NO NW ND': образный «барыга с Горбушки» — сам хранит и сам доставляет.

Выбор схемы по-разному влияет на цепочку исполнения заказов и на их маржинальность (как для Компании, так и для продавца), но в этой таблице мы отразим только обязательства продавцов перед компанией за пользование её услугами: атрибута **fix_fee** и **var_fee** (издержки бывают постоянные и переменные).

таблица	атрибут	тип	требования
schemes	name	text	non-null
schemes	fix_fee	float	positive xor null
schemes	var_fee	float	positive xor null

Таблица **sku**

Представляет из себя характеристики конкретного товара³ (**sku**). У каждого есть наименование (русским языком) **rus_name**. Может лежать на нескольких складах (warehouse_id_list). Все sku телефонов логично объединить в одну категорию, а категории телефонов и планшетов можно объединить в «надкатегорию». Категории будут разных уровней, по мере увеличения «уровня» будут укрупняться. Они могут называться «электроника», «продукты» и т.п., но для экономии времени у меня они будут закодированы натуральными числами. Пусть количество категорий третьего уровня **category_lvl_3** равняется 30, второго (**category_lvl_2**) — 20, первого (**category_lvl_1**) — 10.

таблица	атрибут	тип	требования
sku	sku	integer	non-null
sku	rus_name	text	non-null
sku	warehouse_id_list	integer[]	non-null
sku	category_lvl_1	integer	non-null, \in {0, ..., 9}
sku	category_lvl_2	integer	non-null, \in {0, ..., 19}
sku	category_lvl_3	integer	non-null, \in {0, ..., 29}

Таблица **promo**

Промоакции (читай — скидки) бывают разные, но мы остановимся на процентных (**percent_discount**) и фиксированных (**fixed_discount**). Зачастую давать скидки в Москве выгоднее, чем в Самаре или Хабаровске. Порой давать сутки в каких-то городах не выгодно вовсе. Та же логика распространяется на склады. Именно поэтому атрибута **region_id** и **warehouse_id** являются ключевыми. У каждой акции есть длительность, за что отвечают атрибута **start_msk_dttm**, **end_msk_dttm** и **active_flag**. Атрибут **active_flag** для комбинации id, region_id, warehouse_id, где акцию было принято решение НЕ

³ Например, телефоны с разным объёмом памяти считаются разными sku, то же с обувью разных размеров. А вот макароны разного срока годности, но одного наименования — одна sku.

проводить равняется FALSE с момента её заведения, а атрибуты начала и конца акции пустуют.

таблица	атрибут	тип	требования
promo	id	integer	non-null
promo	percent_discount	integer	positive xor null \in (0 100)
promo	fixed_discount	integer	positive xor null
promo	region_id	integer	non-null
promo	warehouse_id	integer	non-null
promo	start_msk_dttm	timestamp	
promo	end_msk_dttm	timestamp	
promo	active_flag	boolean	non-null

Таблица customers

Покупатели не всегда добровольно регистрируются на онлайн-ресурсах, а заставлять их было бы неприбыльно. Поэтому пользователь, не предоставивший никаких данных о себе, тем не менее способен оформить заказ, и в таком случае его уникальный идентификатор **customer_id** будет формироваться, например, хэшем от кук, ip и других девайс-специфичных признаков. Customer_id всегда не null.

Если пользователь всё-таки решился на регистрацию, пускай это возможно только про телефону и паролю. В явном виде их держать в первую очередь небезопасно, а во вторую почти не нужно — в подавляющем большинстве use-cases можно обойтись хэшем от них, то есть атрибутами **hash_login** и **hash_password**.

Контактная информация будет содержаться в словаре **contact_info**: не хочется их держать в виде атрибутов потому что такую информацию, как паспорт или почтовый индекс будут заполнять очень мало пользователей и столбцы окажутся разреженными.

Атрибуты first_order_dttm, last_order_dttm и multiorders_num говорят многое об активности пользователя. И хоть их можно вычислить по таблице orders, мы всё-таки оставим их тут.

таблица	атрибут	тип	требования
customers	id	integer	non-null
customers	hash_login	integer	8 chatacters
customers	hash_password	integer	8 chatacters
customers	contact_info	jsonb	
customers	first_delivered_order_msk_dttm	timestamp	first_delivered_order_msk_dttm <=

			last_delivered_order_msk_dttm
customers	last_delivered_order_msk_dttm	timestamp	first_delivered_order_msk_dttm <= last_delivered_order_msk_dttm
customers	multiorders_num	integer	non-negative

Таблица **partners**

У каждого партнёра (продавца) есть уникальный **id**, и статус активности личного кабинета **active_flag** (после завершения сотрудничества информация остаётся). Контактная информация хранится в атрибут **contact_info** в формате jsonb по причинам, аналогичным полю customers.contact_info. Также, для партнёра известна дата регистрации **joined_dttm** личного кабинета и его схема **scheme**.

таблица	атрибут	тип	требования
partners	id	integer	non-null
partners	contact_info	jsonb	
partners	joined_msk_dttm	timestamp	non-null
partners	scheme	integer	non-null
partners	active_flag	boolean	non-null

Таблица **warehouses**

У каждого склада (атрибут **id**) есть своё внутреннее название **name** и регион нахождения (федеральный) **region_id**.

таблица	атрибут	тип	требования
warehouses	id	integer	non-null
warehouses	name	text	
warehouses	region_id	integer	non-null

Таблица **regions**

Компания работает в разных регионах страны. В РФ их 8, поэтому **id** принимает значения от 0 до 7. У региона есть название **federal_name**.

таблица	атрибут	тип	требования
regions	id	integer	non-null, \in {0, ..., 7}
regions	federal_name	text	non-null

Типичные запросы

- Общая выручка Компании за месяц/год во всех федеральных округах или в некоторых
- Динамика доли доставленных заказов среди всех оформленных (слово динамика означает в дополнительной гранулярности `DATE(created_msk_dttm)` для конкретного склада
- Средняя/медианная цена покупки залогина/незалогина в федеральном округе `region_id`
- X-ый перцентиль скорости доставки по схеме `scheme`
- Динамика количества регистраций продавцов на площадке
- Динамика объёма скидки
- Самая маржинальная категория уровня K