

Создание и нормализация базы данных

Крючкова Юлия Николаевна

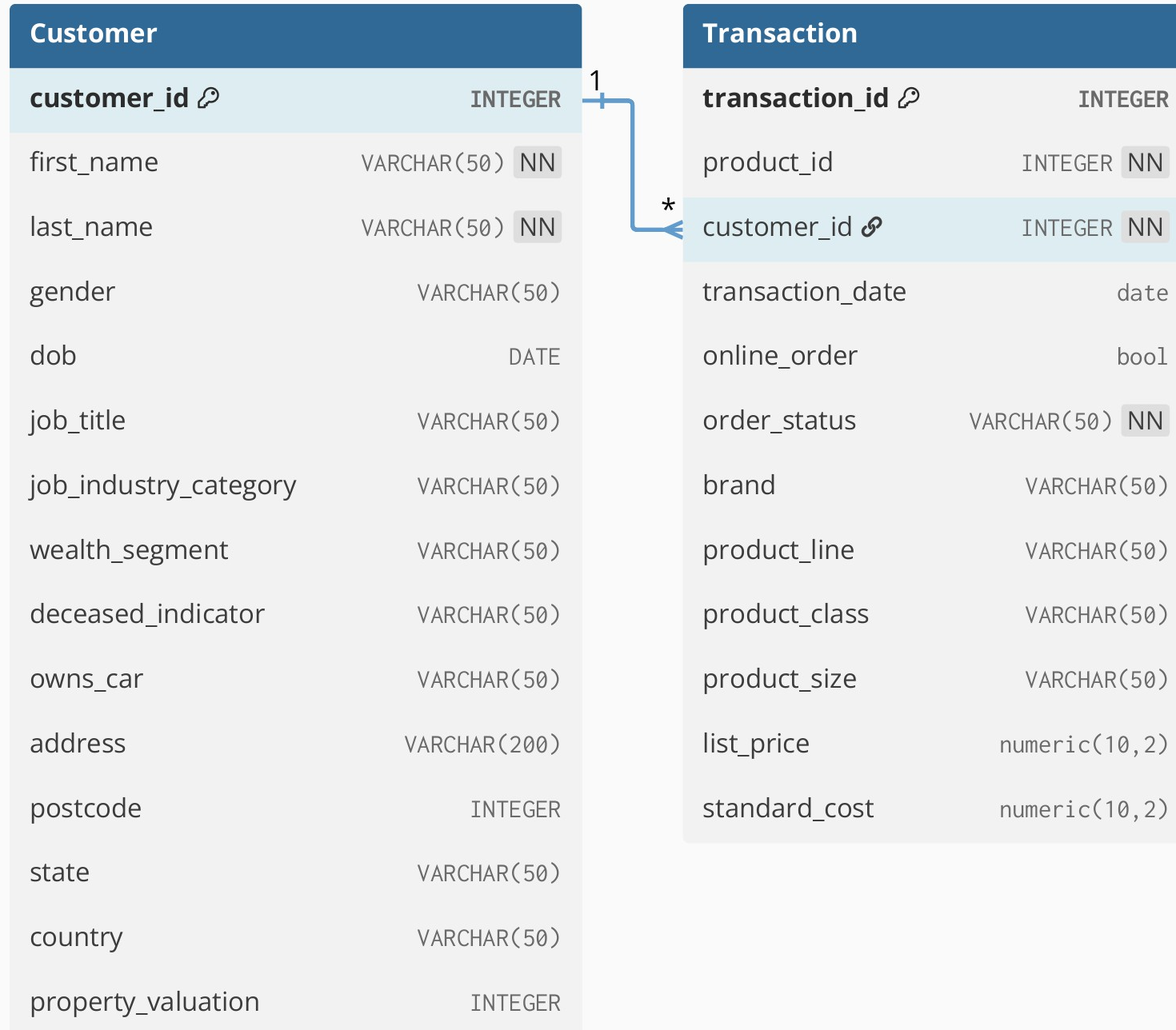
Sales	
transaction_id 	INTEGER
product_id	INTEGER NN
customer_id	INTEGER NN
transaction_date	date
online_order	bool
order_status	VARCHAR(50) NN
brand	VARCHAR(50)
product_line	VARCHAR(50)
product_class	VARCHAR(50)
product_size	VARCHAR(50)
list_price	numeric(10,2)
standard_cost	numeric(10,2)
first_name	VARCHAR(50)
last_name	VARCHAR(50)
gender	VARCHAR(50)
dob	date
job_title	VARCHAR(50)
job_industry_category	VARCHAR(50)
wealth_segment	VARCHAR(50)
deceased_indicator	VARCHAR(50)
owns_car	VARCHAR(50)
address	VARCHAR(200)
postcode	INTEGER
state	VARCHAR(50)
country	VARCHAR(50)
property_valuation	INTEGER

1НФ:

- Все значения атомарны
- Каждая строка уникальна по первичному ключу
- Нет повторяющихся групп

Несоответствие 2НФ:

- Есть **дублирование** информации о клиентах.
 - пример: если клиент сделал 5 заказов, то его имя, адрес, профессия будут повторятся 5 раз. Это приводит к избыточности данных и потенциальным аномалиям при обновлении (например, если изменится адрес клиента, нужно обновлять все строки с его заказами).

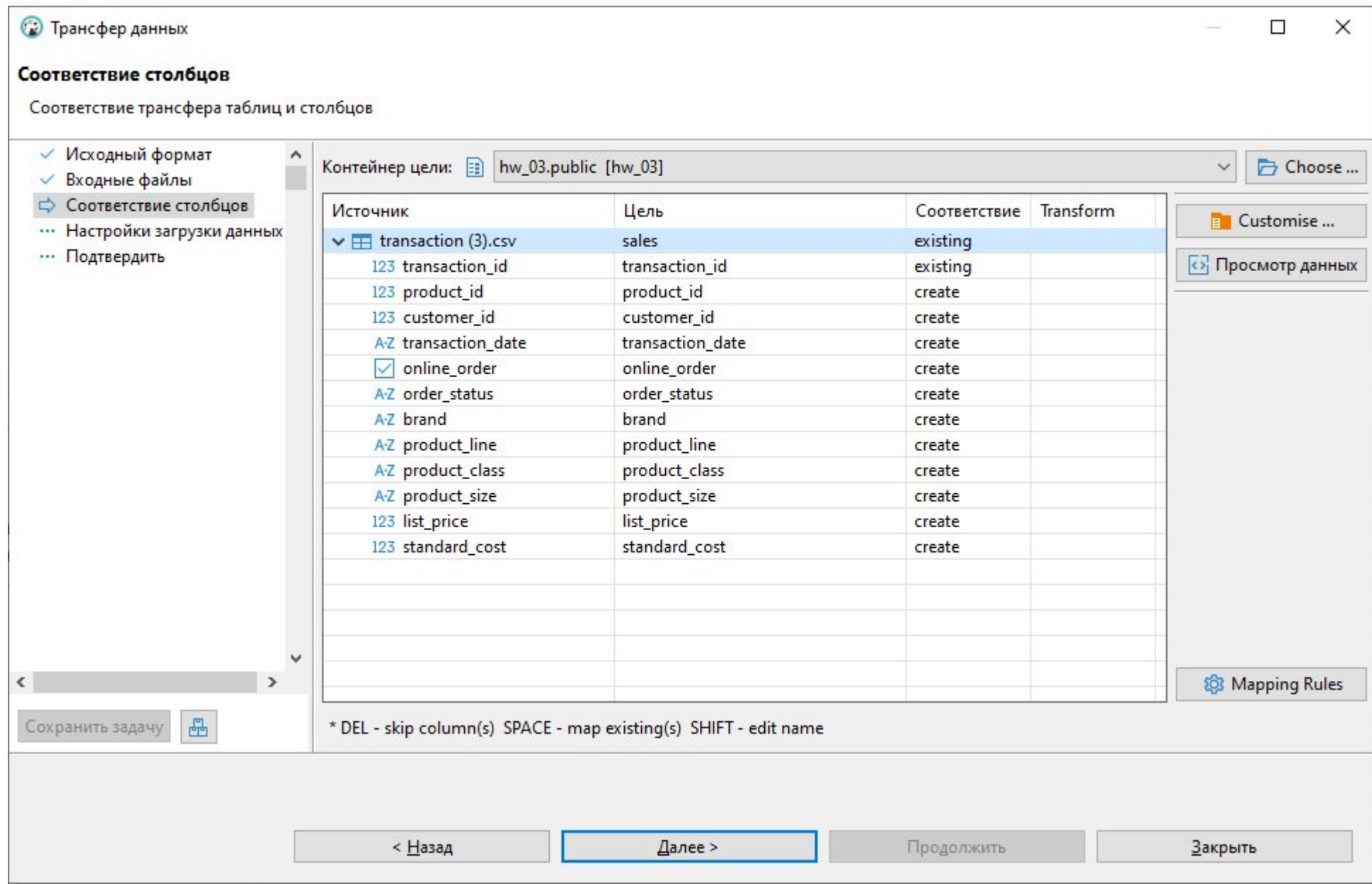


2НФ:

- Таблицы в 1НФ
- Первичный ключ не составной - нет частичных зависимостей (ни одна из колонок не зависит от части первичного ключа)
- Избыточность данных частично устранена, данные о клиентах вынесены в отдельную таблицу

Несоответствие 3НФ:

- Есть транзитивные зависимости
 - пример: в таблице Transaction *brand*, *product_line*, *product_class*, *product_size*, *list_price*, *standart_cost* зависят от *product_id*, а не от *transaction_id(pk)*
 - пример: в таблице Customer *postcode* и *property_valuation* зависят от *address*, а не от *customer_id(pk)*
 - пример: в таблице Customer *state*, *country* зависят от *postcode*, а не от *customer_id(pk)*



Импортируем данные в таблицу **transaction** с предыдущего слайда через Dbeaver

	123 transaction_id	123 product_id	123 customer_id	AZ transaction_date	<input checked="" type="checkbox"/> online_order	AZ order_status	AZ brand	AZ product
1	1	2	2 950	02-25-17	[]	Approved	Solex	Standard
2	2	3	3 120	05-21-17	[v]	Approved	Trek Bicycles	Standard
3	3	37	402	10-16-17	[]	Approved	OHM Cycles	Standard
4	4	88	3 135	08-31-17	[]	Approved	Norco Bicycles	Standard
5	5	78	787	10-01-17	[v]	Approved	Giant Bicycles	Standard
6	6	25	2 339	03-08-17	[v]	Approved	Giant Bicycles	Road
7	7	22	1 542	04-21-17	[v]	Approved	WeareA2B	Standard
8	8	15	2 459	07-15-17	[]	Approved	WeareA2B	Standard
9	9	67	1 305	08-10-17	[]	Approved	Solex	Standard
10	10	12	3 262	08-30-17	[v]	Approved	WeareA2B	Standard
11	11	5	1 986	01-17-17	[]	Approved	Trek Bicycles	Mountain
12	12	61	2 783	01-05-17	[v]	Approved	OHM Cycles	Standard
13	13	35	1 243	02-26-17	[v]	Approved	Trek Bicycles	Standard
14	14	16	2 717	09-10-17	[]	Approved	Norco Bicycles	Standard
15	15	12	247	06-11-17	[]	Approved	Giant Bicycles	Standard
16	16	3	2 961	10-10-17	[]	Approved	Trek Bicycles	Standard
17	17	79	2 426	04-03-17	[]	Approved	Norco Bicycles	Standard
18	18	33	1 842	06-02-17	[]	Approved	Giant Bicycles	Standard
19	19	54	2 268	04-06-17	[v]	Approved	WeareA2B	Standard
20	20	25	3 002	01-28-17	[v]	Approved	Giant Bicycles	Road
21	21	27	1 582	10-09-17	[]	Approved	Trek Bicycles	Standard
22	22	37	595	06-29-17	[v]	Approved	OHM Cycles	Standard
23	23	37	2 001	04-08-17	[v]	Approved	OHM Cycles	Standard
24	24	82	515	10-18-17	[]	Approved	Giant Bicycles	Road
25	25	89	2 822	06-11-17	[]	Approved	WeareA2B	Touring
26	26	64	2 596	01-10-17	[]	Approved	Trek Bicycles	Standard
27	27	64	2 666	04-11-17	[v]	Approved	Trek Bicycles	Standard
28	28	19	76	12-23-17	[]	Approved	Trek Bicycles	Mountain

Свойства

Данные

Диаграмма

Название:

transaction

Табличное пространство:

pg_default

Комментарий:

Has Row-Level Security

Секционирование по:

ID объекта:

16506

Колонки

Ограничения

Внешние ключи

Название	Колонка	Владелец	Тип	Выражение	Комментарий
transaction_pkey	—	transaction	PRIMARY KEY		
123 transaction_id	123transaction_id	—	—	—	—

С помощью sql запросов типы колонок и ограничения были приведены в соответствие с диаграммой со слайда 3. Покупка не может существовать без статуса, покупателя и покупки, поэтому колонки *product_id*, *customer_id* и *order_status* отмечены как Not Null

Свойства

Данные

Диаграмма

Название:

transaction

Табличное пространство:

pg_default

Владелец:

Дополнить

Комментарий:

Has Row-Level Security

Секционирование

Секционирование по:

ID объекта:

16542

Колонки

Ограничения

Внешние ключи

Индексы

Зависимости

Ссылки

Секции таблиц

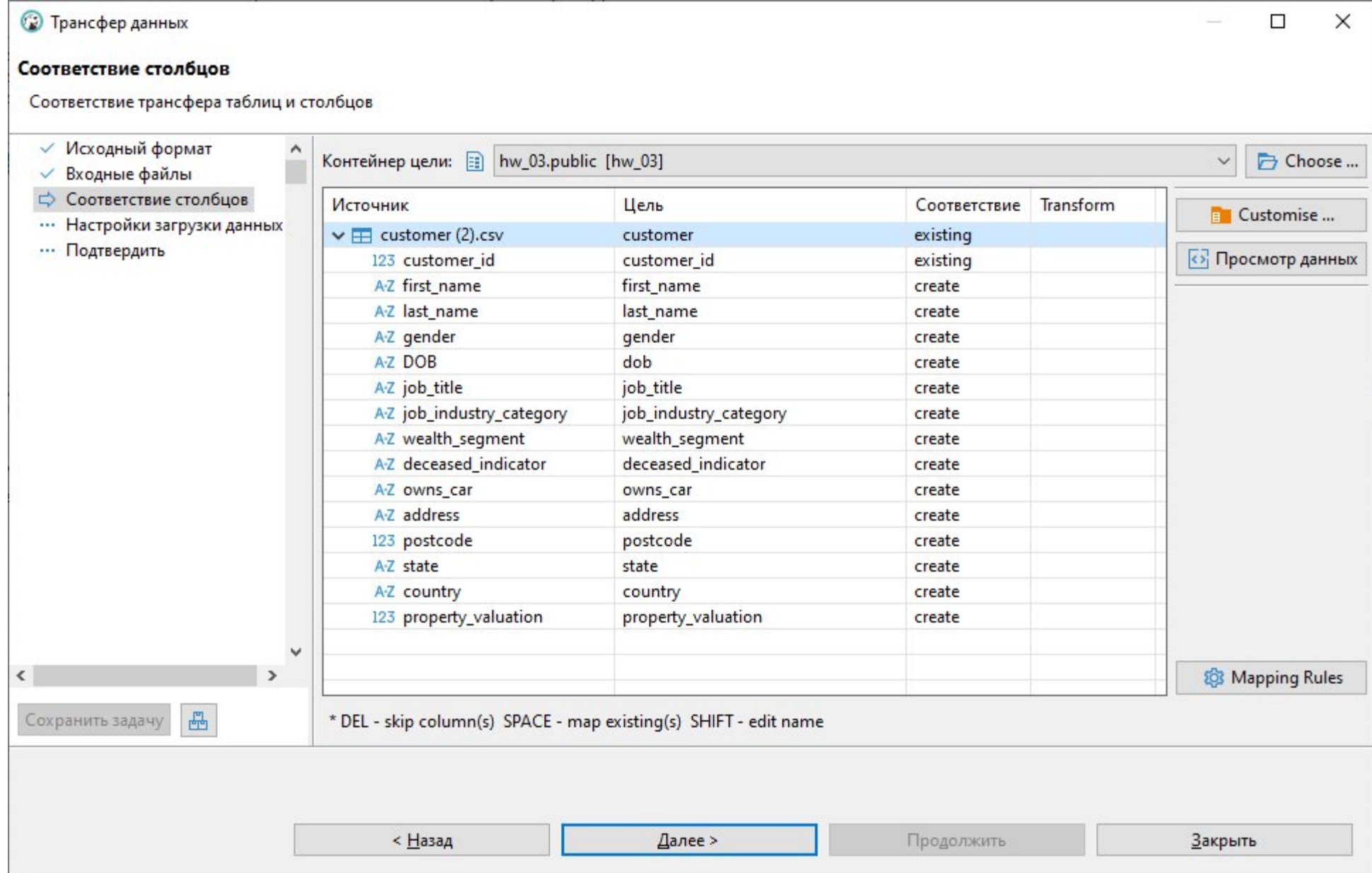
Триггеры

Правила

Policies

Statistics

Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
123 transaction_id	1	int4			[v]		
123 product_id	2	int4			[v]		
123 customer_id	3	int4			[v]		
transaction_date	4	date			[]		
online_order	5	bool			[]		
A-Z order_status	6	varchar(50)		default	[v]		
A-Z brand	7	varchar(50)		default	[]		
A-Z product_line	8	varchar(50)		default	[]		
A-Z product_class	9	varchar(50)		default	[]		
A-Z product_size	10	varchar(50)		default	[]		
123 list_price	11	numeric(10, 2)			[]		
123 standard_cost	12	numeric(10, 2)			[]		



Импорт данных со второго листа в таблицу **customer** через Dbeaver

Свойства Данные Диаграмма

Название: customer Табличное пространство: pg_d

Комментарий:

Секционирование по:

ID объекта: 1651

Колонки	Название	Колонка	Владелец	Тип	Выражение	Комментарий
Ограничения	customer_pkey	—	customer	PRIMARY KEY	—	—
Внешние ключи	123 customer_id	123 customer_id	—	—	—	—

С помощью sql запросов типы колонок и ограничения были приведены в соответствие с диаграммой со слайда 3. Поля *first_name* и *last_name* отмечены как Not Null, т.к. являются минимальным идентификатором человека.

Свойства Данные Диаграмма

Название: customer Табличное пространство: pg_default Владелец:

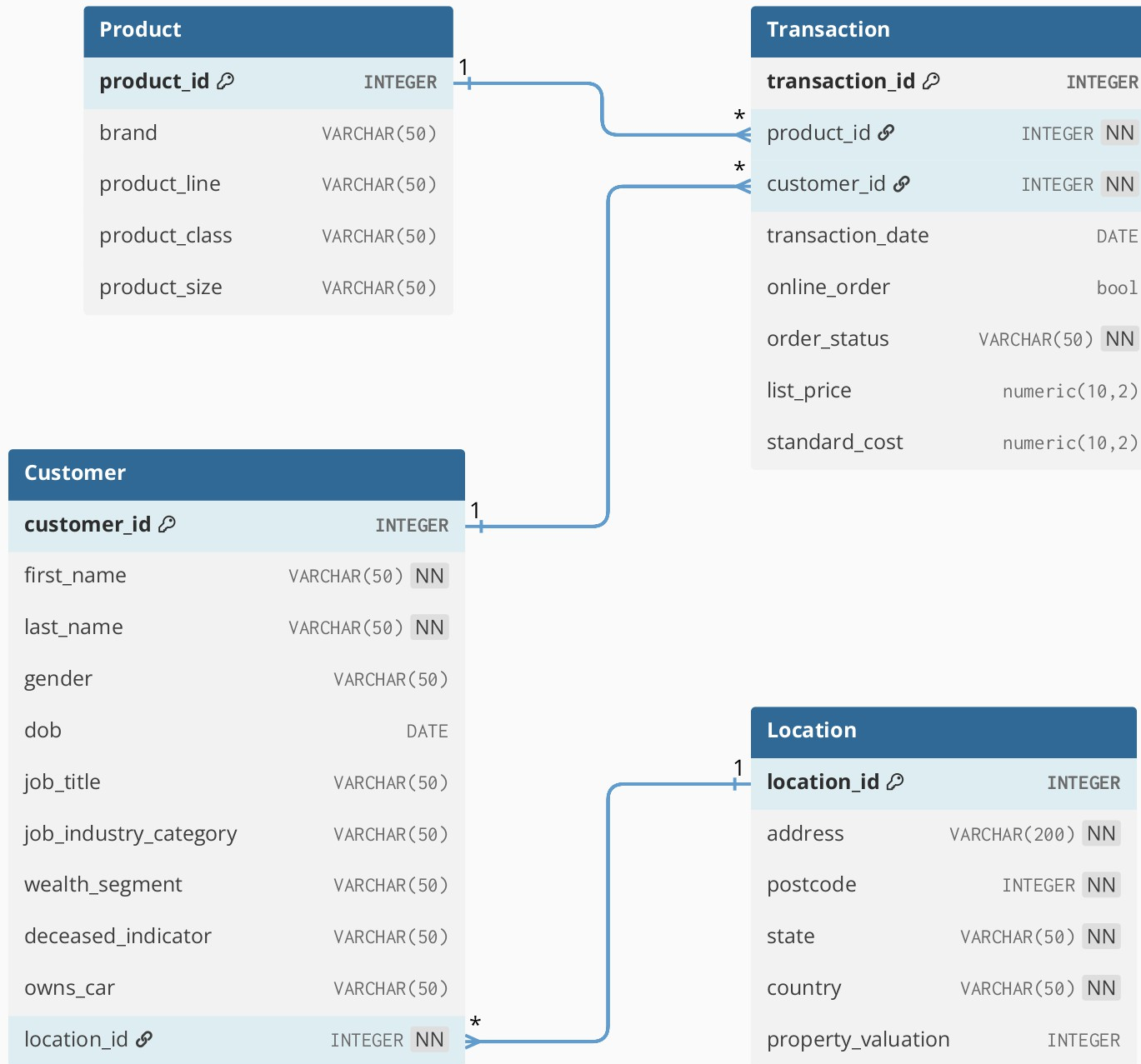
Комментарий:

Секционирование по:

ID объекта: 16516

Has Row-Level Security Секционирование

Колонки	Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
Ограничения	123 customer_id	1	int4			[v]		
Внешние ключи	A-Z first_name	2	varchar(50)		default	[v]		
Индексы	A-Z last_name	3	varchar(50)		default	[v]		
Зависимости	A-Z gender	4	varchar(50)		default	[]		
Ссылки	dob	5	date			[]		
Секции таблиц	A-Z job_title	6	varchar(50)		default	[]		
Триггеры	A-Z job_industry_category	7	varchar(50)		default	[]		
Правила	A-Z wealth_segment	8	varchar(50)		default	[]		
Policies	A-Z deceased_indicator	9	varchar(50)		default	[]		
Statistics	A-Z owns_car	10	varchar(50)		default	[]		
Права доступа	A-Z address	11	varchar(200)		default	[]		
DDL	123 postcode	12	int4			[]		
	A-Z state	13	varchar(50)		default	[]		
	A-Z country	14	varchar(50)		default	[]		
	123 property_valuation	15	int4			[]		



3НФ:

- Все таблицы в 2НФ
- Устранены транзитивные зависимости для таблиц Customer и Transaction
- В текущих таблицах нет транзитивных зависимостей

В результате нормализации каждая таблица хранит данные, относящиеся только к одному объекту предметной области. Это снижает избыточность, упрощает обновление данных и повышает целостность базы.

Свойства Данные Диаграмма

Название: transaction Табличное пространство: pg_default Владелец: Дополнит

Комментарий:

☐ Has Row-Level Security ☐ Секционирование

Секционирование по:

ID объекта: 16542

Колонки	Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
	123 transaction_id	1	int4			[v]		
	123 product_id	2	int4			[v]		
	123 customer_id	3	int4			[v]		
	🕒 transaction_date	4	date			[]		
	☑ online_order	5	bool			[]		
	A-Z order_status	6	varchar(50)		default	[v]		
	123 list_price	11	numeric(10, 2)			[]		
	123 standard_cost	12	numeric(10, 2)			[]		

Название: product Табличное пространство: pg_default Владелец: Дополнит

Комментарий:

☐ Has Row-Level Security ☐ Секционирование

Секционирование по:

ID объекта: 16611

Колонки	Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
	123 product_id	1	int4			[v]		
	A-Z brand	2	varchar(50)		default	[]		
	A-Z product_line	3	varchar(50)		default	[]		
	A-Z product_class	4	varchar(50)		default	[]		
	A-Z product_size	5	varchar(50)		default	[]		

Из таблицы **transaction** вынесена таблица **product**.

Все поля **product** nullable, потому что в момент транзакции мы можем не знать детали продукта

Свойства Данные Диаграмма

Название: customer Табличное пространство: pg_default Владелец: Дополните

Комментарий:

☐ Has Row-Level Security ☐ Секционирование

Секционирование по:

ID объекта: 16516

	Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
Колонки	123 customer_id	1	int4			[v]		
Ограничения	A-Z first_name	2	varchar(50)		default	[v]		
Внешние ключи	A-Z last_name	3	varchar(50)		default	[v]		
Индексы	A-Z gender	4	varchar(50)		default	[]		
Зависимости	🕒 dob	5	date			[]		
Ссылки	A-Z job_title	6	varchar(50)		default	[]		
Секции таблиц	A-Z job_industry_cat	7	varchar(50)		default	[]		
Триггеры	A-Z wealth_segment	8	varchar(50)		default	[]		
Правила	A-Z deceased_indicat	9	varchar(50)		default	[]		
Policies	A-Z owns_car	10	varchar(50)		default	[]		
	123 location_id	17	int4			[]		

Свойства Данные Диаграмма

Название: location Табличное пространство: pg_default Владелец: Дополните

Комментарий:

☐ Has Row-Level Security ☐ Секционирование

Секционирование по:

ID объекта: 16647

	Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
Колонки	123 location_id	1	serial4			[v]	nextval('locatio...	
Ограничения	A-Z address	2	varchar(200)		default	[v]		
Внешние ключи	123 postcode	3	int4			[v]		
Индексы	A-Z state	4	varchar(50)		default	[v]		
Зависимости	A-Z country	5	varchar(50)		default	[v]		
Ссылки	123 property_valuatic	6	int4			[]		

Из таблицы **customer** вынесена таблица **location**. Все поля таблицы **location** not nullable, кроме **property_valuation** потому что при отсутствии любого из полей невозможно идентифицировать адрес.