

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 повторяться 5 раз. Это приводит к избыточности данных и потенциальным аномалиям при обновлении (например, если изменится адрес клиента, нужно обновлять все строки с его заказами).

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

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)*

Трансфер данных

Соответствие столбцов

Соответствие трансфера таблиц и столбцов

✓ Исходный формат
✓ Входные файлы
➡ Соответствие столбцов
... Настройки загрузки данных
... Подтвердить

Контейнер цели: hw_03.public [hw_03] Choose ...

Customise ...
Просмотр данных

Источник	Цель	Соответствие	Transform
transaction (3).csv	sales	existing	
123 transaction_id	transaction_id	existing	
123 product_id	product_id	create	
123 customer_id	customer_id	create	
AZ transaction_date	transaction_date	create	
<input checked="" type="checkbox"/> online_order	online_order	create	
AZ order_status	order_status	create	
AZ brand	brand	create	
AZ product_line	product_line	create	
AZ product_class	product_class	create	
AZ product_size	product_size	create	
123 list_price	list_price	create	
123 standard_cost	standard_cost	create	

Mapping Rules

Сохранить задачу

* DEL - skip column(s) SPACE - map existing(s) SHIFT - edit name

< Назад Далее > Продолжить Закрыть

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



Ба... X Пр... □

<test_db> Script <test_db> Script-2 *<hw_03> Script-1 transaction X

Свойства Данные Диаграмма

Показать SQL Введите SQL выражение чтобы отфильтровать результаты

Фильтровать соединение DBeaver Sample Database (S)

> hw_03 localhost:5432

Базы данных

hw_03

Схемы

public

Таблицы

> transaction

Внешние таб

Представлен

Мат. предста

Индексы

Функции

Последовате

Типы данных

Агрегатные с

Событийные триг

Расширения

Хранилище

Системные объект

Роли

postgres

test_db

> Администрирование

> Системные объекты

> mentoring localhost:5433

> postgres localhost:5432

> test_db localhost:5432

	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:			
		Секционирование по:	
		ID объекта:	16506

Колонки

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

Ограничения

Внешние ключи

Свойства **Данные** **Диаграмма**

Название:	transaction	Табличное пространство:	pg_default	Владелец:
Комментарий:				Дополнит:
		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			[]		
<input checked="" type="checkbox"/> 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)			[]		

Ограничения

Внешние ключи

Индексы

Зависимости

Ссылки

Секции таблиц

Триггеры

Правила

Policies

Statistics

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

Трансфер данных

Соответствие столбцов

Соответствие трансфера таблиц и столбцов

✓ Исходный формат
✓ Входные файлы
➡ Соответствие столбцов
... Настройки загрузки данных
... Подтвердить

Контейнер цели: hw_03.public [hw_03] Choose ...

Customise ...
Просмотр данных

Источник	Цель	Соответствие	Transform
customer (2).csv	customer	existing	
123 customer_id	customer_id	existing	
A-Z first_name	first_name	create	
A-Z last_name	last_name	create	
A-Z gender	gender	create	
A-Z DOB	dob	create	
A-Z job_title	job_title	create	
A-Z job_industry_category	job_industry_category	create	
A-Z wealth_segment	wealth_segment	create	
A-Z deceased_indicator	deceased_indicator	create	
A-Z owns_car	owns_car	create	
A-Z address	address	create	
123 postcode	postcode	create	
A-Z state	state	create	
A-Z country	country	create	
123 property_valuation	property_valuation	create	

Mapping Rules

Сохранить задачу

* DEL - skip column(s) SPACE - map existing(s) SHIFT - edit name

< Назад Далее > Продолжить Закрыть

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



Ба... X Пр... □

<test_db> Script <test_db> Script-2 *<hw_03> Script-1 transaction customer X

Свойства Данные Диаграмма

Показать SQL Введите SQL выражение чтобы отфильтровать результаты

Фильтровать соединение

DBeaver Sample Database (S)

hw_03 localhost:5432

Базы данных

hw_03

Схемы

public

Таблицы

customer transaction

Внешние таб

Представлен

Мат. предста

Индексы

Функции

Последовате

Типы данных

Агрегатные с

Событийные триг

Расширения

Хранилище

Системные объек

Роли

postgres

test_db

Администрирование

Системные объек

mentoring localhost:5433

postgres localhost:5432

test_db localhost:5432

	123 customer_id	AZ first_name	AZ last_name	AZ gender	AZ dob	AZ job_title	AZ job_industry_category
1	1	Laraine	Medendorp	F	1953-10-12	Executive Secretary	Health
2	2	Eli	Bockman	Male	1980-12-16	Administrative Officer	Financial Services
3	3	Arlin	Dearle	Male	1954-01-20	Recruiting Manager	Property
4	4	Talbot		Male	1961-10-03		IT
5	5	Sheila-kathryn	Calton	Female	1977-05-13	Senior Editor	n/a
6	6	Curr	Duckhouse	Male	1966-09-16		Retail
7	7	Fina	Merali	Female	1976-02-23		Financial Services
8	8	Rod	Inder	Male	1962-03-30	Media Manager I	n/a
9	9	Mala	Lind	Female	1973-03-10	Business Systems Development Analyst	Argiculture
10	10	Fiorenze	Birdall	Female	1988-10-11	Senior Quality Engineer	Financial Services
11	11	Uriah	Bisatt	Male	1954-04-30		Property
12	12	Sawyere	Flattman	Male	1994-07-21	Nuclear Power Engineer	Manufacturing
13	13	Gabriele	Norcross	Male	1955-02-15	Developer I	Financial Services
14	14	Rayshell	Kitteman	Female	1983-03-25	Account Executive	Financial Services
15	15	Erroll	Radage	Male	2000-07-13	Junior Executive	Manufacturing
16	16	Harlin	Parr	Male	1977-02-27	Media Manager IV	n/a
17	17	Heath	Faraday	Male	1962-03-19	Sales Associate	n/a
18	18	Marjie	Neasham	Female	1967-07-06	Professor	n/a
19	19	Sorcha	Keyson	Female	2001-04-15	Geological Engineer	Manufacturing
20	20	Basile	Firth	Male	1980-08-13	Project Manager	Manufacturing
21	21	Mile	Cammocke	Male	1980-09-20	Safety Technician I	Manufacturing
22	22	Deeanne	Durtnell	Female	1962-12-10		IT
23	23	Olav	Polak	Male	1995-02-10		n/a
24	24	Kim	Skpsey	Female	1977-12-03	Research Assistant I	Argiculture
25	25	Geoff	Assaf	Male	1976-12-02	Accounting Assistant III	Financial Services
26	26	Trixi	Ginnelly	Female	1978-06-10	Editor	Financial Services
27	27	Garvin	Klees	Male	1978-09-25	Research Nurse	Health
28	28	Fee	Zellmer	Male	1973-09-30	Senior Quality Engineer	Health

Свойства **Данные** **Diagramma**

Название: customer
Комментарий:

Табличное пространство: pg_default
Секционирование по:
ID объекта: 16516

Название	Колонка	Владелец	Тип	Выражение	Комментарий
customer_pkey	customer_id	customer	PRIMARY KEY	—	—
123 customer_id 123customer_id					

Колонки
Ограничения
Внешние ключи

Свойства **Данные** **Diagramma**

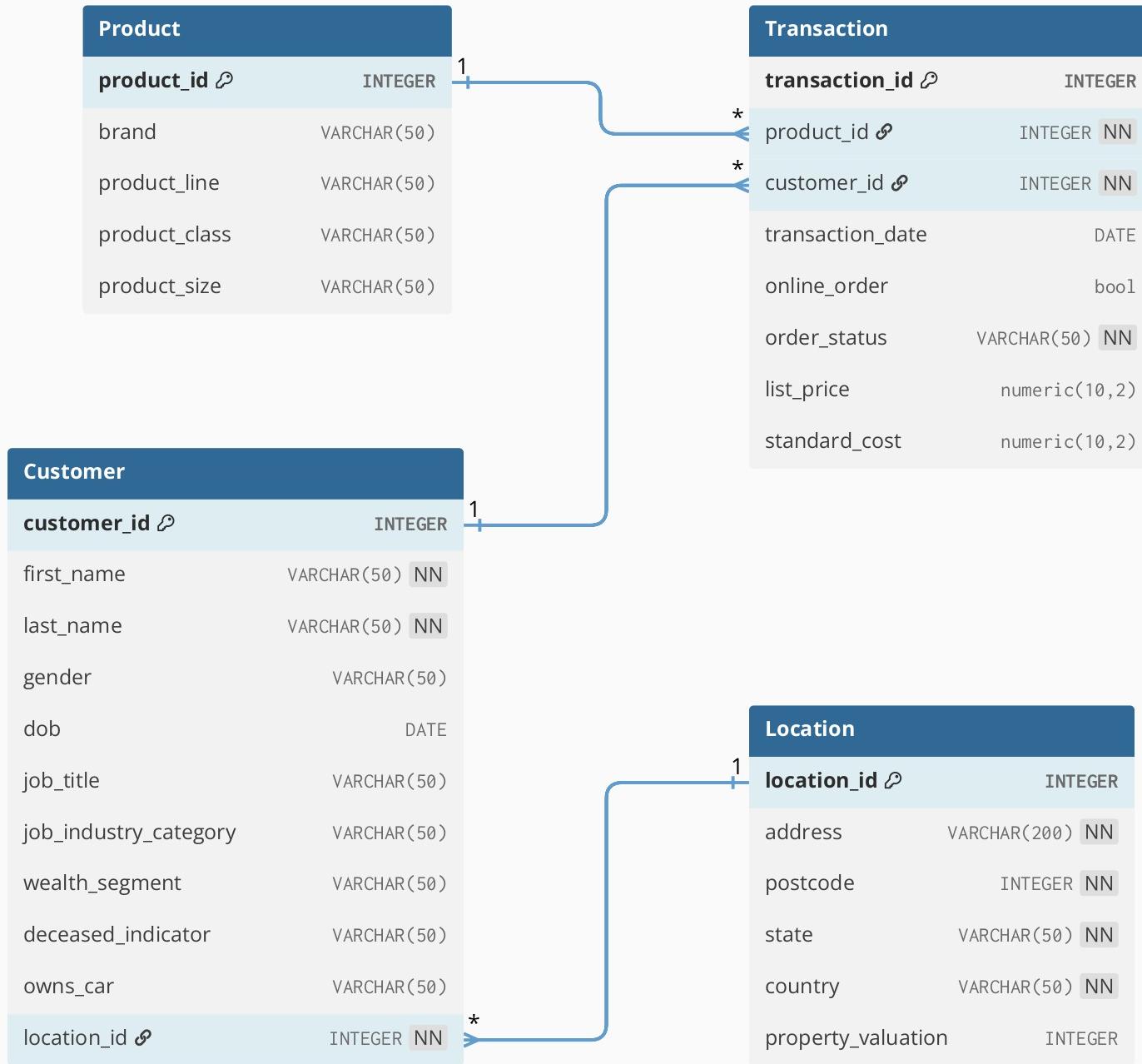
Название: customer
Комментарий:

Табличное пространство: pg_default
Владелец:
Has Row-Level Security Секционирование
Секционирование по:
ID объекта: 16516

Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
123 customer_id	1	int4	—	default	[v]	—	—
AZ first_name	2	varchar(50)	—	default	[v]	—	—
AZ last_name	3	varchar(50)	—	default	[v]	—	—
AZ gender	4	varchar(50)	—	default	[]	—	—
(D) dob	5	date	—	default	[]	—	—
AZ job_title	6	varchar(50)	—	default	[]	—	—
AZ job_industry_category	7	varchar(50)	—	default	[]	—	—
AZ wealth_segment	8	varchar(50)	—	default	[]	—	—
AZ deceased_indicator	9	varchar(50)	—	default	[]	—	—
AZ owns_car	10	varchar(50)	—	default	[]	—	—
AZ address	11	varchar(200)	—	default	[]	—	—
123 postcode	12	int4	—	default	[]	—	—
AZ state	13	varchar(50)	—	default	[]	—	—
AZ country	14	varchar(50)	—	default	[]	—	—
123 property_valuation	15	int4	—	default	[]	—	—

Колонки
Ограничения
Внешние ключи
Индексы
Зависимости
Ссылки
Секции таблиц
Триггеры
Правила
Policies
Statistics
Права доступа
DDL

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.



3НФ:

- Все таблицы в 2НФ
- Устраниены 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:
Комментарий: Дополнит

Has Row-Level Security Секционирование

Секционирование по:

ID объекта: 16542

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			[]		
<input checked="" type="checkbox"/> 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)			[]		

Колонки

- Ограничения
- Внешние ключи
- Индексы
- Зависимости
- Ссылки
- Секции таблиц

Название: product Табличное пространство: pg_default Владелец:
Комментарий: Дополнит

Has Row-Level Security Секционирование

Секционирование по:

ID объекта: 16611

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	[]		

Колонки

- Ограничения
- Внешние ключи
- Индексы
- Зависимости

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:	
Комментарий:		<input type="checkbox"/> Has Row-Level Security <input type="checkbox"/> Секционирование		Дополнительно	
		Секционирование по:			
		ID объекта:	16516		

Колонки

Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
123 customer_id	1	int4			[v]		
A-Z first_name	2	varchar(50)		default	[v]		
A-Z last_name	3	varchar(50)		default	[v]		
A-Z gender	4	varchar(50)		default	[]		
dob	5	date			[]		
A-Z job_title	6	varchar(50)		default	[]		
A-Z job_industry_cate	7	varchar(50)		default	[]		
A-Z wealth_segment	8	varchar(50)		default	[]		
A-Z deceased_indicat	9	varchar(50)		default	[]		
A-Z owns_car	10	varchar(50)		default	[]		
123 location_id	17	int4			[]		

Свойства **Данные** **Диаграмма**

Название:	location	Табличное пространство:	pg_default	Владелец:	
Комментарий:		<input type="checkbox"/> Has Row-Level Security <input type="checkbox"/> Секционирование		Дополнительно	
		Секционирование по:			
		ID объекта:	16647		

Колонки

Название	#	Тип данных	Автоувеличение	Правило сортировки	Not Null	По умолчанию	Комментарий
123 location_id	1	serial4			[v]	nexval('location_id')	
A-Z address	2	varchar(200)		default	[v]		
123 postcode	3	int4			[v]		
A-Z state	4	varchar(50)		default	[v]		
A-Z country	5	varchar(50)		default	[v]		
123 property_valuati	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.