

Декомпозиция и нормализация данных на примере учебного файла

Александр Гоппе

9 февраля 2025 г.

1 Исходная структура данных

Исходные данные представлены в плоской структуре, где атрибуты клиента и его адреса объединены в одной таблице:

title	first_name	last_name	language	birth_date	gender	marital_status	country	postal_code	region	city	street	building
Dr.	Danilo	Ambrosini	IT	1900-01-01	Unknown		IT	21100		Varese	Via Dolomiti	13
Mrs	Deborah	Thomas	EN		Female		GB	RH16 3TQ		Haywards Heath	Cobbetts Mead	31

2 Проблемы исходной структуры

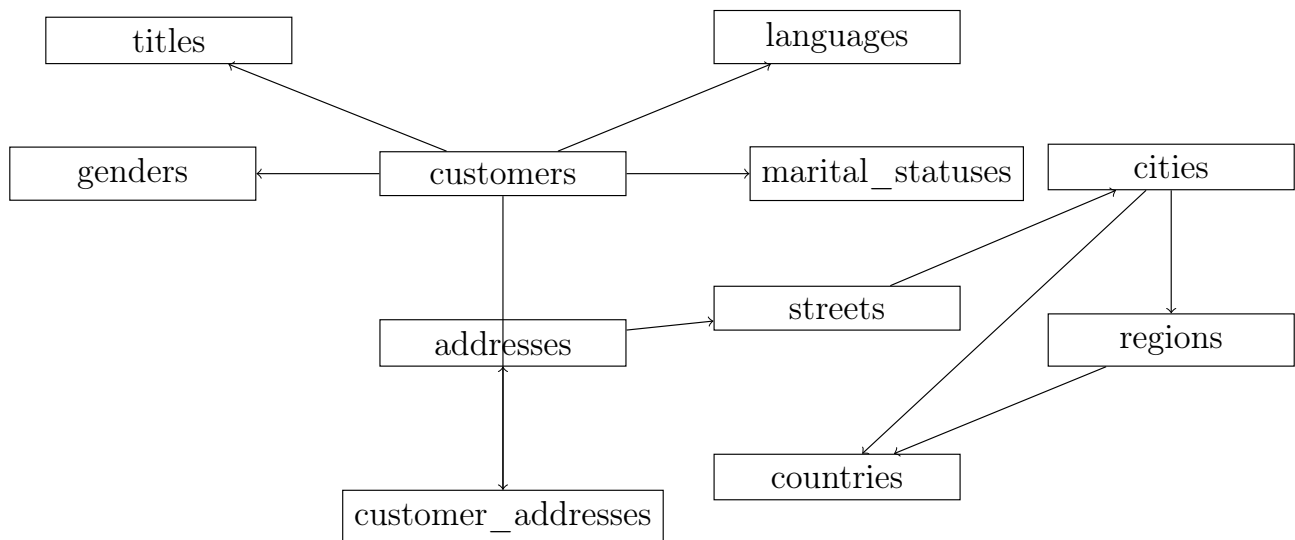
- Отсутствие типизации (все данные хранятся как строки).
- Отсутствие нормализации (данные о клиенте и адресе объединены).
- Повторяющиеся данные (страны, города, улицы, языки, пол).
- Затрудненная модификация и поддержка (например, изменение названия города потребует обновления всех записей).

3 Нормализация и конечная модель данных

В результате декомпозиции были выделены следующие сущности:

- **customers** (id, title_id, first_name, last_name, birth_date, gender_id, marital_status_id, language_id)
- **titles** (id, name)
- **genders** (id, name)
- **languages** (id, name)
- **marital_statuses** (id, name)
- **countries** (id, name, code)
- **cities** (id, name, region_id, country_id)
- **regions** (id, name, country_id)
- **streets** (id, name, city_id)
- **addresses** (id, postal_code, street_id, building_number)
- **customer_addresses** (customer_id, address_id)

3.1 Диаграмма конечной структуры



4 Первичная проверка корректности переноса

Запрос	версия	результат
<code>select count(distinct first_name, last_name) from init_customers where first_name <> '' or last_name <> ''</code>	старая	1122
<code>select count(distinct first_name, last_name) from customers</code>	новая	1122
<code>select count(distinct postal_code) from init_customers</code>	старая	1079
<code>select count(distinct postal_code) from addresses</code>	новая	1079

5 Заключение

Благодаря нормализации удалось:

- Устранить избыточность данных.
- Обеспечить целостность и удобство обновления.
- Улучшить читаемость структуры и её расширяемость.