

### Пояснительная записка

Исходные таблицы «transaction» и «customer» находятся в первой нормальной форме, поскольку все атрибуты в таблицах – простые, а все ячейки содержат лишь скалярные значения. Эти таблицы не соответствуют второй нормальной форме (и, соответственно, третьей тоже), потому что в них не соблюдается требование об обеспечении зависимости полей от первичного ключа.

Исходной таблице «transaction» свойственна избыточность данных и в ней усложнён процесс выборки из-за одновременного присутствия данных о самой транзакции и о продукте. Так же, по последней причине, в исходной таблице присутствуют транзитивные связи. Например, характеристики продукта (brand, product\_line, product\_class, product\_size) и его цена/себестоимость (list\_price, standard\_cost) не зависят непосредственно от транзакции, а определяются принадлежностью к продукту.

Для исключения транзитивных связей и приведения исходной таблицы «transaction» к третьей нормальной форме она разбивается на две таблицы: «transaction» и «product».

Новая таблица «transaction» состоит из полей, характеризующих транзакцию и относящихся к ней непосредственно: transaction\_id, product\_id, customer\_id, transaction\_date, online\_order, order\_status.

Новая таблица «product» состоит из ключевого поля product\_id и остальных полей, зависящих от него непосредственно (т.е. относящихся только к товару): product\_id, brand, product\_line, product\_class, product\_size, list\_price, standard\_cost. Выделение таблицы «product» не только устраняет транзитивные связи, но и позволяет существенно сократить избыточность данных: если количество transaction\_id составляет 20 тыс., то количество product\_id составляет всего около 100.

Исходная таблица «customer» так же содержит транзитивные связи, поскольку в ней, возможно выделить независимые блоки информации об адресе клиента, работе клиента и его благосостоянии.

Для исключения транзитивных связей и приведения исходной таблицы «customer» к третьей нормальной форме она разбивается на четыре таблицы: «customer», «address», «job» и «wealth».

Новая таблица «customer» состоит из ключевого поля `customer_id` и остальных полей, зависящих от него непосредственно: `address_id`, `job_id`, `wealth_id`, `first_name`, `last_name`, `DOB`, `deceased_indicator`.

В новую таблицу «address» добавляется ключевое поле `address_id`, от которого непосредственно зависят остальные поля таблицы, определяющие адрес клиента: `address`, `postcode`, `state`, `country`.

В новую таблицу «job» добавляется ключевое поле `job_id`, от которого непосредственно зависят остальные поля таблицы, определяющие профессию клиента: `job_title`, `job_industry_category`.

В новую таблицу «wealth» добавляется ключевое поле `wealth_id`, от которого непосредственно зависят остальные поля таблицы, определяющие благосостояние клиента: `wealth_segment`, `owns_car`, `property_valuation`.

При устранении транзитивных связей в таблице «customer» также сокращается избыточность данных. Например, выборке из первых ста значений `customer_id` соответствует количество значений `job_id`, уменьшенное до 84, и количество значений `wealth_id`, уменьшенное до 42.