

## Нормализация таблиц

- Приводим исходные таблицы к форме 1НФ, для этого разбиваем колонку *address* содержащую номер дома и название улицы на две отдельные колонки *house* и *street*. Теперь все значения в колонках атомарны.

The diagram illustrates the normalization process. On the left, table K contains a single column 'address' with multiple rows of address data. A large blue arrow points to the right, where the data is split into two tables: P and Q. Table P contains the 'house' column with values like 60, 6, 0, 17979, 9, 4, 49, 97736, 93405, and 44339. Table Q contains the 'street' column with values like Morning Avenue, Meadow Vale Court, Holy Cross Court, Del Mar Point, Oakridge Court, Delaware Trail, Londonderry Lane, 7th Trail, Ludington Park, and Golden Leaf Alley.

K	P	Q
address	house	street
060 Morning Avenue	60	Morning Avenue
6 Meadow Vale Court	6	Meadow Vale Court
0 Holy Cross Court	0	Holy Cross Court
17979 Del Mar Point	17979	Del Mar Point
9 Oakridge Court	9	Oakridge Court
4 Delaware Trail	4	Delaware Trail
49 Londonderry Lane	49	Londonderry Lane
97736 7th Trail	97736	7th Trail
93405 Ludington Park	93405	Ludington Park
44339 Golden Leaf Alley	44339	Golden Leaf Alley

- Для приведения таблиц к форме 2НФ выделим колонки связанные с продуктом (*product\_id, brand, product\_line, product\_class, product\_size, list\_price, standard\_cost*) из таблицы *transactions* в отдельную таблицу *products*, и создадим ключ *prod\_id*, т.к. данные по товару не зависят от *product\_id* (например встречается набор из разных товаров под одним номером *product\_id*). Добавляем ключ *prod\_id* в таблицы *product* и *transactions*.

Наименование товара и цена  
совпадают при разных  
*product\_id*

A screenshot of a database table showing multiple rows for the same product. The table has columns: product\_id, brand, product\_line, product\_class, product\_size, list\_price, and standard\_cost. The 'brand' column consistently shows 'Norco Bicycles'. The 'list\_price' and 'standard\_cost' columns are also consistently filled with the value '381,10'. However, the 'product\_id' column contains various values such as 4, 31, 356, 396, 402, 520, 834, 868, 1011, 1019, 1023, 1038, 1138, 1165, 1212, 1342, 1545, and 1598, indicating that different product IDs are used for the same product model.

product_id	brand	product_line	product_class	product_size	list_price	standard_cost
4	88	Norco Bicycles	Standard	medium	1198,46	381,10
31	88	Norco Bicycles	Standard	medium	1198,46	381,10
356	88	Norco Bicycles	Standard	medium	1198,46	381,10
396	88	Norco Bicycles	Standard	medium	1198,46	381,10
402	88	Norco Bicycles	Standard	medium	1198,46	381,10
520	88	Norco Bicycles	Standard	medium	1198,46	381,10
834	88	Norco Bicycles	Standard	medium	1198,46	381,10
868	88	Norco Bicycles	Standard	medium	1198,46	381,10
1011	88	Norco Bicycles	Standard	medium	1198,46	381,10
1019	88	Norco Bicycles	Standard	medium	1198,46	381,10
1023	23	Norco Bicycles	Standard	medium	1198,46	381,10
1038	88	Norco Bicycles	Standard	medium	1198,46	381,10
1138	88	Norco Bicycles	Standard	medium	1198,46	381,10
1165	88	Norco Bicycles	Standard	medium	1198,46	381,10
1212	23	Norco Bicycles	Standard	medium	1198,46	381,10
1342	88	Norco Bicycles	Standard	medium	1198,46	381,10
1545	23	Norco Bicycles	Standard	medium	1198,46	381,10
1598	88	Norco Bicycles	Standard	medium	1198,46	381,10

- Для приведения таблиц к форме 3НФ выделяем в отдельную таблицу все данные в колонках на листе *customer*, убирая транзитивные связи, которые были в изначальном датасете.

Итоговая структура таблиц с типами данных показана на следующем слайде

## Структура таблиц

