

Лекция 6. Часть 2

**Аномалии в данных.**

**Нормализация реляционной БД**

# Почему схема БД может быть плохой?

**Аномалия** - противоречие между моделью предметной области и моделью данных.

Аномалии включения, обновления (противоречивость), удаления

Блюдо	Вид	Дата	Продукт	Калорийность	Вес (г)	Поставщик	Город	Страна	Цена (\$)
Лобио	Закуска	01.09.2012	Фасоль	307	200	"Хуанхэ"	Пекин	Китай	0.37
Лобио	Закуска	01.09.2012	Лук	45	40	"Наталка"	Киев	Украина	0.52
Лобио	Закуска	01.09.2012	Масло	742	30	"Лайма"	Рига	Латвия	1.55
Лобио	Закуска	01.09.2012	Зелень	18	10	"Даугава"	Рига	Латвия	0.99
Борщ	Суп	01.09.2012	Мясо	166	80	"Наталка"	Киев	Украина	2.18
Борщ	Суп	01.09.2012	Лук	45	30	"Наталка"	Киев	Украина	0.52
Борщ	Суп	01.09.2012	Томаты	24	40	"Полесье"	Киев	Украина	0.45
Борщ	Суп	01.09.2012	Рис	334	50	"Хуанхэ"	Пекин	Китай	0.44
Борщ	Суп	01.09.2012	Масло	742	15	"Полесье"	Киев	Украина	1.62
Борщ	Суп	01.09.2012	Зелень	18	15	"Наталка"	Киев	Украина	0.88

**Нормализация** - формальная процедура, в ходе которой одна таблица разбивается на две или несколько по определенным правилам.

### **Задачи нормализации:**

- Исключение из таблиц повторяющейся информации – борьба с аномалиями.
- Создание структуры, в которой предусмотрена возможность ее будущих изменений и влияние таких изменений на приложения минимально.

# Дублирование и избыточность данных

Книга	Студент
Книга 1	Иванов
Книга 2	Петров
Книга 3	Иванов

Данные  
**дублированные,**  
но не избыточные

Книга	Студент	Телефон
Книга 1	Иванов	9631440779
Книга 2	Петров	9170550011
Книга 3	Иванов	9631440779

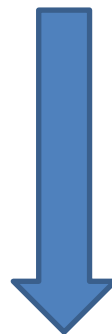
Данные по телефонам  
**избыточные**

В чем опасность избыточности данных в БД?

# Дублирование и избыточность данных

Книга	Студент	Телефон
Книга 1	Иванов	9631440779
Книга 2	Петров	9170550011
Книга 3	Иванов	9631440779

Данные по телефонам  
избыточные



**Разбиваем на две таблицы:**

Книга	Студент
Книга 1	Иванов
Книга 2	Петров
Книга 3	Иванов

Студент	Телефон
Иванов	9631440779
Петров	9170550011

# Нормальные формы

**Первая НФ:** атомарные атрибуты.

Требует, чтобы каждое поле таблицы:

- было неделимым (не должно делиться на более мелкие значения)
- не содержало повторяющихся групп.

**Вторая НФ:** отсутствие зависимостей частичного ключа.

Все поля таблицы должны зависеть от первичного ключа, т.е. первичный ключ однозначно определяет запись и не является избыточным.

Те поля, которые зависят только от части первичного ключа, должны быть выделены в отдельные таблицы.

# Нормальные формы

**Третья НФ:** **устранение транзитивных зависимостей.**

В таблице не должно быть транзитивных зависимостей между неключевыми полями, т.е. чтобы значение любого поля таблицы, не входящего в первичный ключ, не зависело от значения другого поля, не входящего в первичный ключ.

# Первая НФ

Студент	Телефон
Иванов	9631440779; 9780231200
Петров	9170550011; 9051231234



# Первая НФ

Студент	Телефон
Иванов	9631440779; 9780231200
Петров	9170550011; 9051231234

Есть повторяющиеся группы –  
нарушение 1 НФ

# Первая НФ

Студент	Телефон
Иванов	9631440779; 9780231200
Петров	9170550011; 9051231234

Есть повторяющиеся группы –  
нарушение 1 НФ

Студент	Телефон
Иванов	9631440779
Иванов	9780231200
Петров	9051231234
Петров	9170550011

Таблица приведена к 1 НФ

# Пример

Спроектировать БД для автоматизации отпуска товаров со склада

Накладная № 5

Дата: 03.12.2016 Покупатель: ООО "Друг" Адрес: г. Саранск

Наименование	Кол-во	Ед. изм.	Цена	Сумма
Тушенка	1000	Банка	50	50000
Сахар	100	Кг	30	3000

Итого: 53000

# Одна таблица «Отпуск\_товара»

---

Номер	Дата	Покупатель	Адрес	Наименование	Ед_изм	Цена	Кол-во	Сумма
5	03.12.2016	ООО Друг	Саранск	Тушенка	Банка	50	1000	50000
5	03.12.2016	ООО Друг	Саранск	Сахар	Кг	30	100	3000

Отпуск_товаров
Номер
Дата
Покупатель
Адрес
Наименование
Ед_изм
Цена
Количество
Сумма

# Одна таблица «Отпуск\_товара»

Номер	Дата	Покупатель	Адрес	Наименование	Ед_изм	Цена	Кол-во	Сумма
5	03.12.2016	ООО Друг	Саранск	Тушенка	Банка	50	1000	50000
5	03.12.2016	ООО Друг	Саранск	Сахар	Кг	30	100	3000

Отпуск_товаров
Номер
Дата
Покупатель
Адрес
Наименование
Ед_изм
Цена
Количество
Сумма

Таблица приведена к 1 НФ ?

# Одна таблица «Отпуск\_товара»

Номер	Дата	Покупатель	Адрес	Наименование	Ед_изм	Цена	Кол-во	Сумма
5	03.12.2016	ООО Друг	Саранск	Тушенка	Банка	50	1000	50000
5	03.12.2016	ООО Друг	Саранск	Сахар	Кг	30	100	3000

Отпуск_товаров
Номер
Дата
Покупатель
Адрес
Наименование
Ед_изм
Цена
Количество
Сумма

Таблица приведена к 1 НФ ? Да

Первичный ключ?

# Выделение таблицы «Товары»

---

Отпуск_товаров
Номер
Наименование
Дата
Покупатель
Адрес
Ед_изм
Цена
Количество
Сумма

← Первичный ключ

Какие поля зависят только от части  
первичного ключа?

# Выделение таблицы «Товары»

---

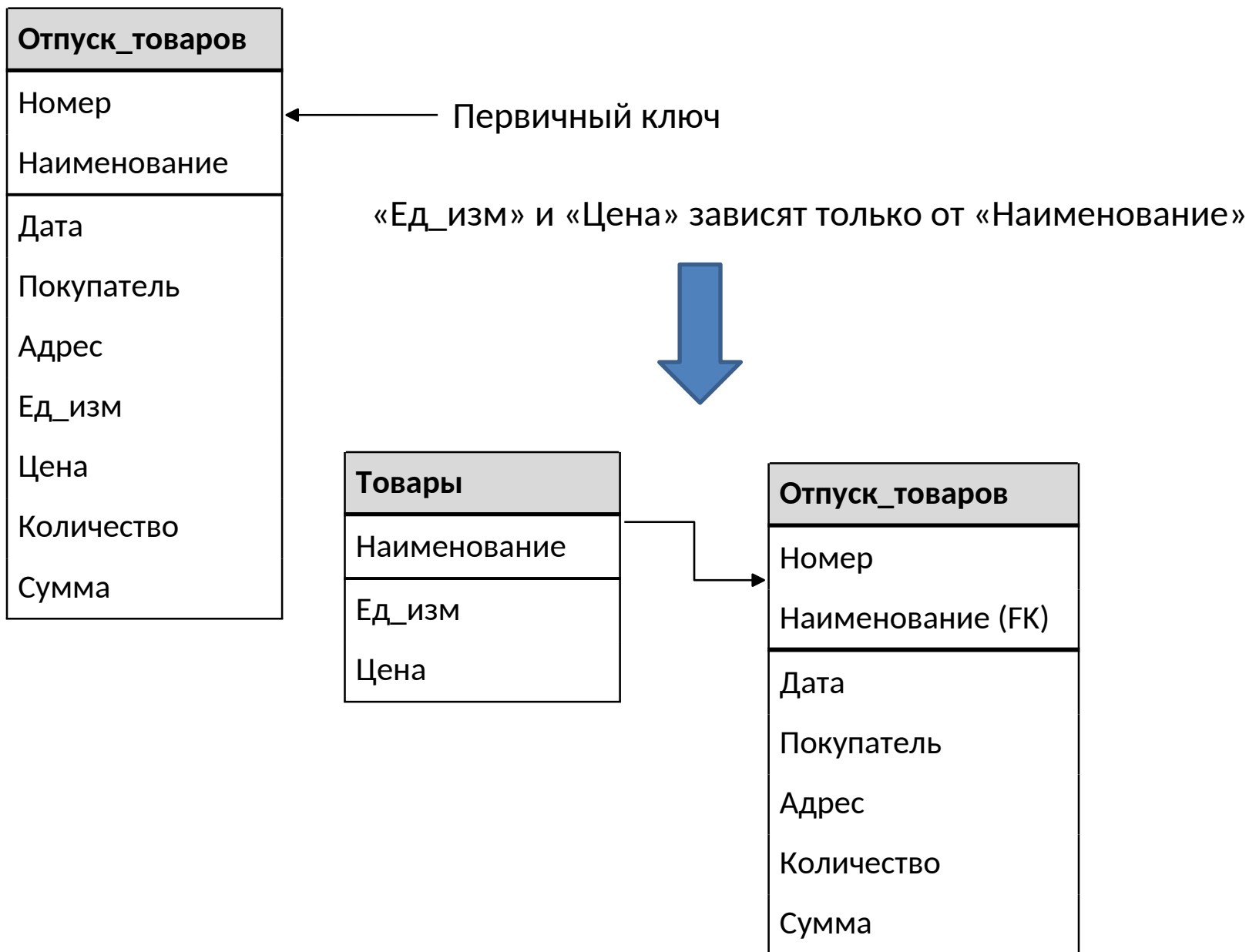
Отпуск_товаров
Номер
Наименование
Дата
Покупатель
Адрес
Ед_изм
Цена
Количество
Сумма

← Первичный ключ

«Ед\_изм» и «Цена» зависят только от «Наименование»



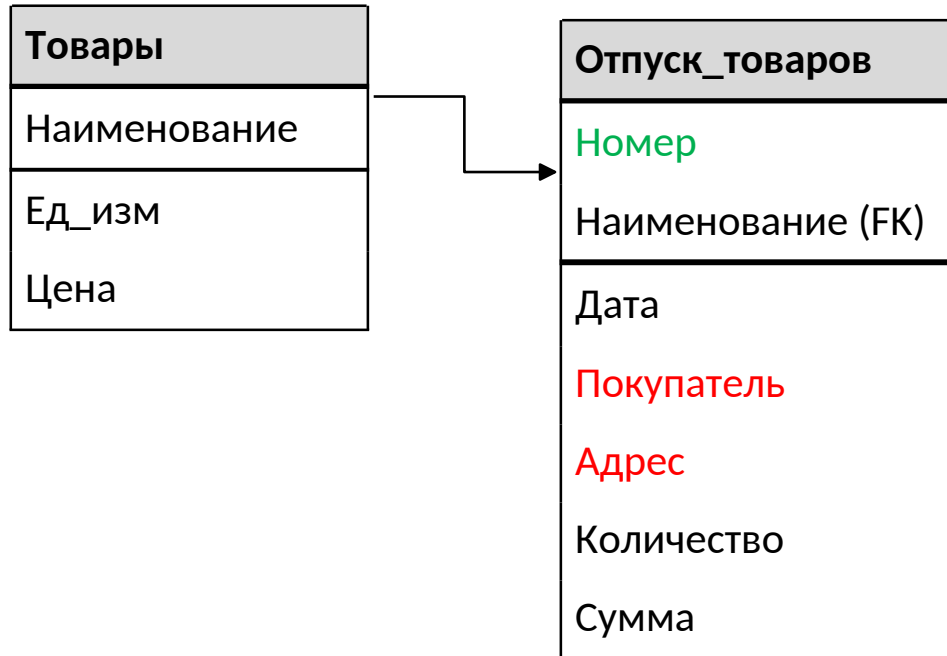
# Выделение таблицы «Товары»



# Выделение таблицы «Покупатели»

---

«Покупатель» и «Адрес» зависят только от номера накладной



# Выделение таблицы «Покупатели»

---

«Покупатель» и «Адрес» зависят только от номера накладной



# Выделение таблицы «Накладные»

---

«Дата» зависит только от номера накладной

Товары
Наименование
Ед_изм
Цена

Отпуск_товаров
Номер
Наименование (FK)
Дата
Количество
Сумма

Покупатели
Покупатель
Адрес

# Выделение таблицы «Накладные»

---

«Дата» зависит только от номера накладной



Товары
Наименование
Ед_изм
Цена

Отпуск_товаров
Номер
Наименование (FK)
Количество
Сумма

Покупатели
Покупатель
Адрес

Накладные
Номер
Дата

# Устанавливаем связи



**Приведена ли БД к 2-й НФ?**

# Устанавливаем связи



БД приведена к 2-й НФ

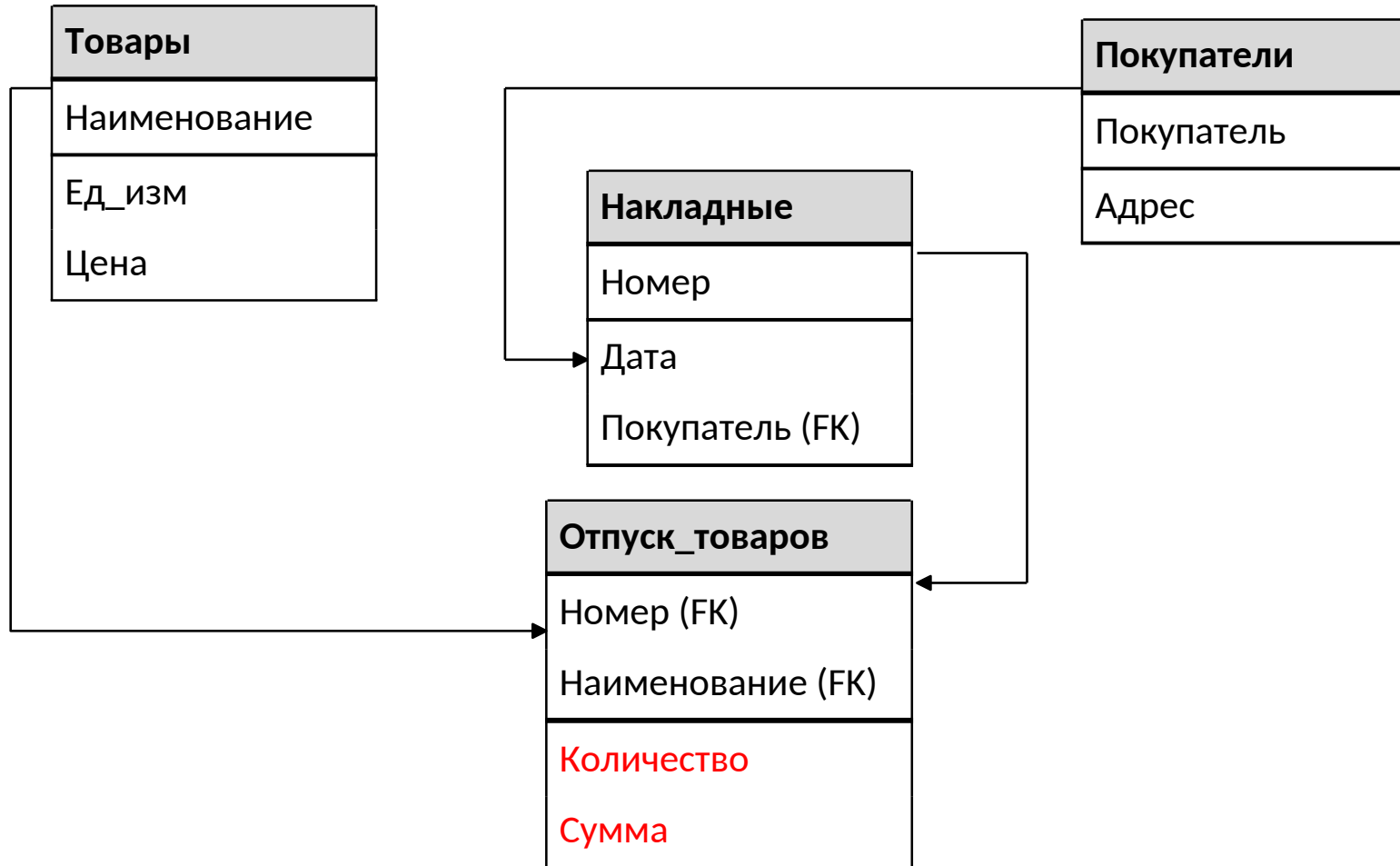
# Устанавливаем связи



**Приведена ли БД к 3-й НФ?**



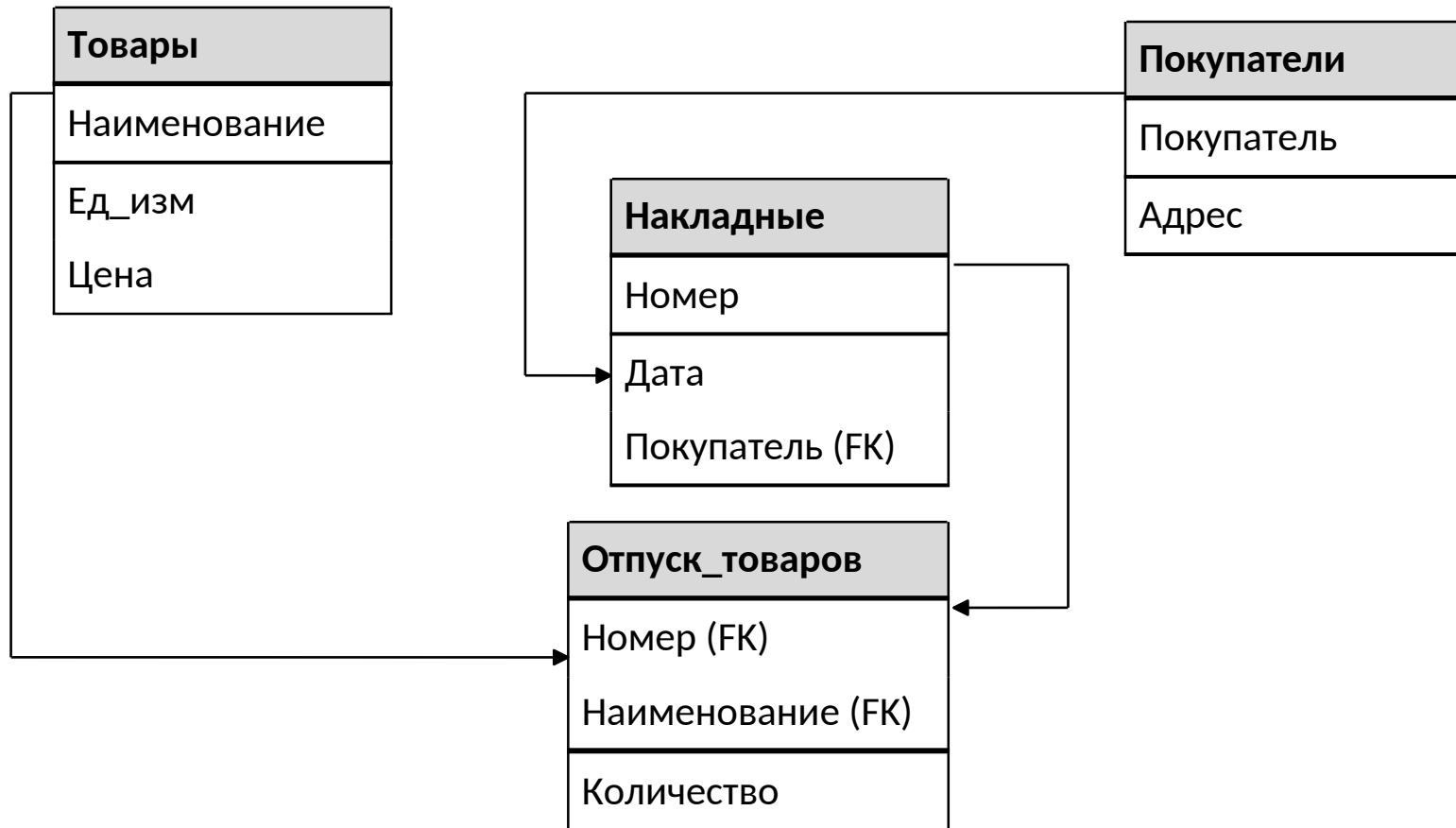
# Устанавливаем связи



«Сумма» зависит от «количества»

# Удаляем лишние поля

"Отпуск\_товаров"."Сумма"="Отпуск\_товаров"."Количество"\*"Товары"."Цена"



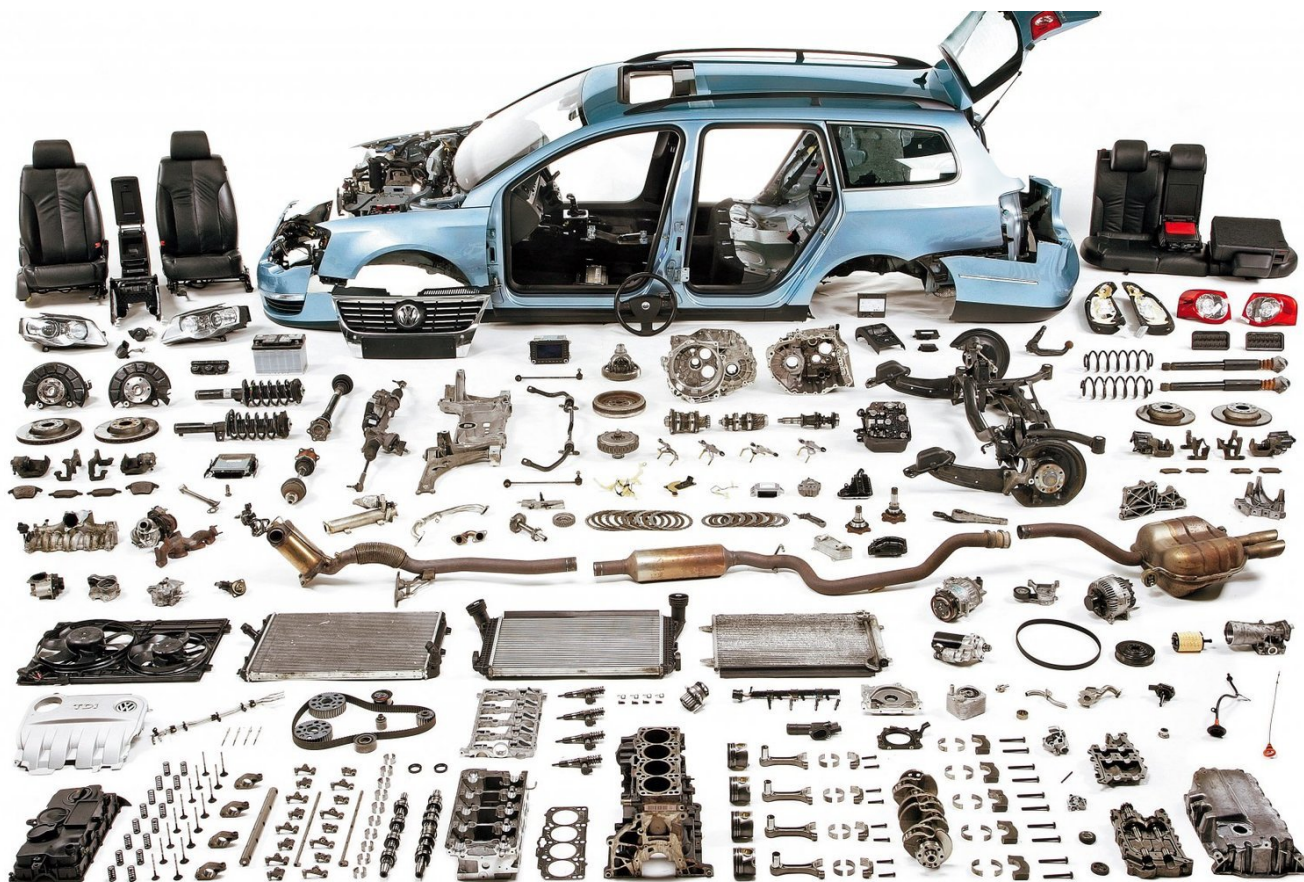
**БД приведена к 3-й НФ**

# Результаты нормализации

**Плюс.** Таблицы содержат только один элемент избыточных данных – это поле связи, присутствующее одновременно у родительской и дочерних таблиц.

**Минус.** Сильно увеличивается число таблиц:

- Теряется восприятие системы в целом.
- Усложняются запросы, приходится объединять данные из разных таблиц (скорость выполнения падает)



**Нормализованный автомобиль**