Лекция 9

Проектирование базы данных

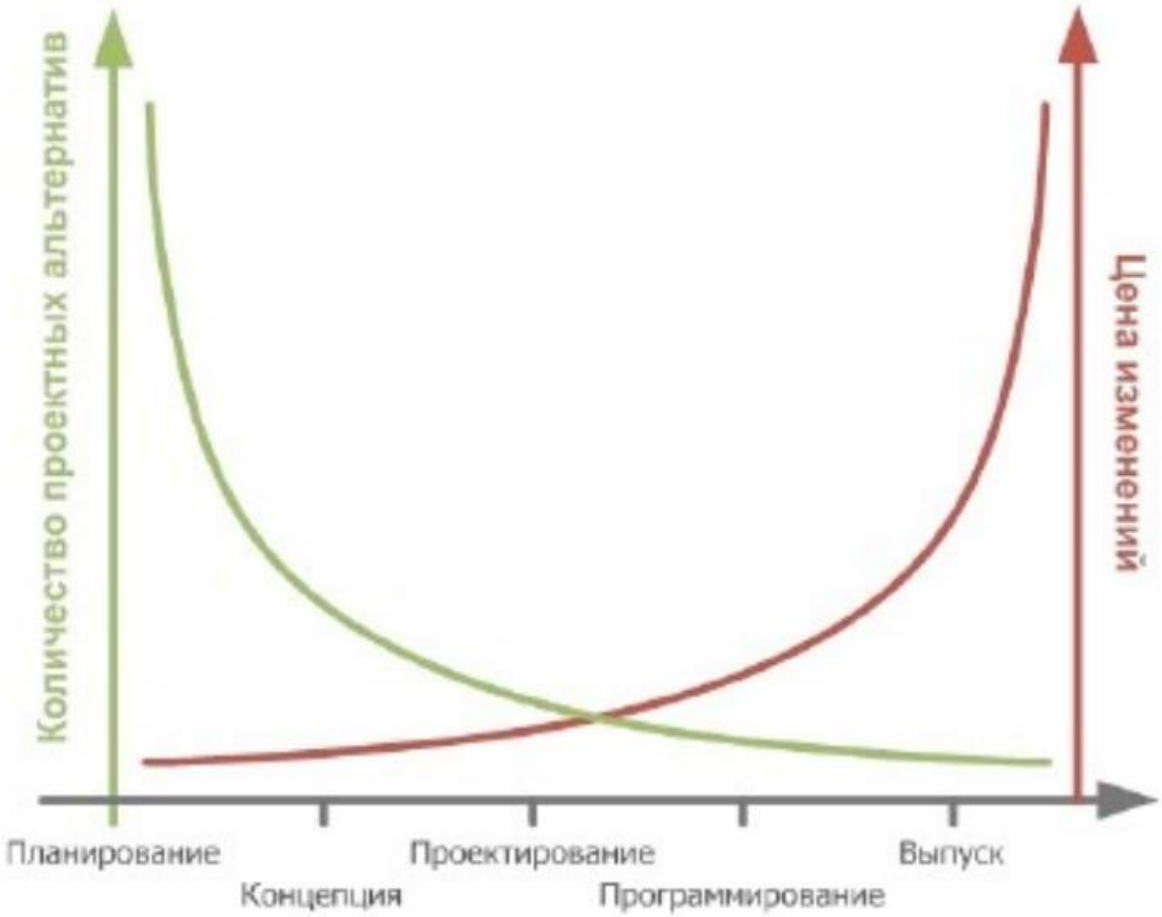
Проектирование БД – процесс создания схемы базы данных и определения необходимых ограничений целостности.

Основные задачи:

- Хранение в БД всей необходимой информации.
- Получение данных по всем необходимым запросам.
- Сокращение избыточности и дублирования данных.
- Обеспечение целостности данных (исключение противоречий, предотвращение потерь и т.д.).

Строим без плана

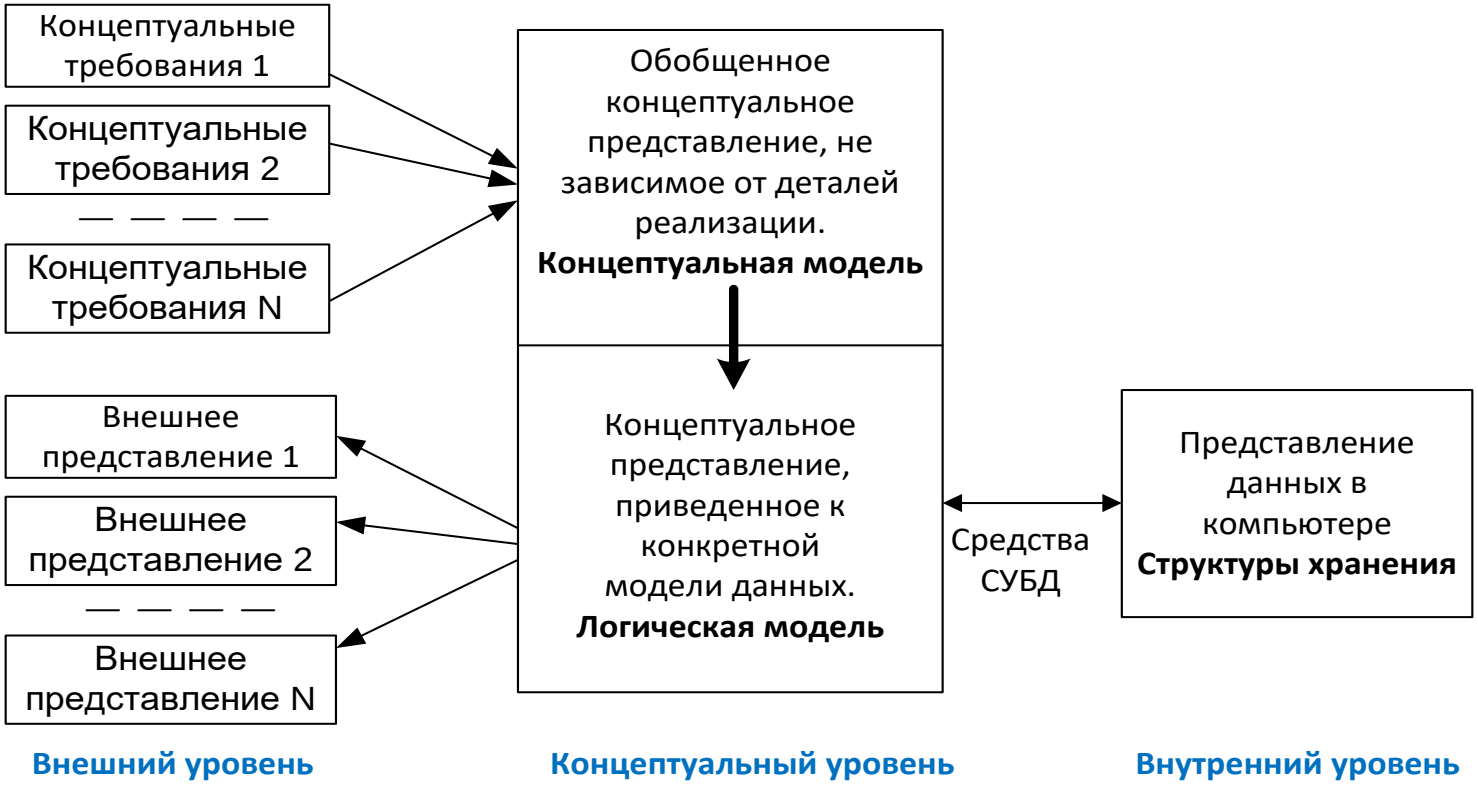
Проектирование БД



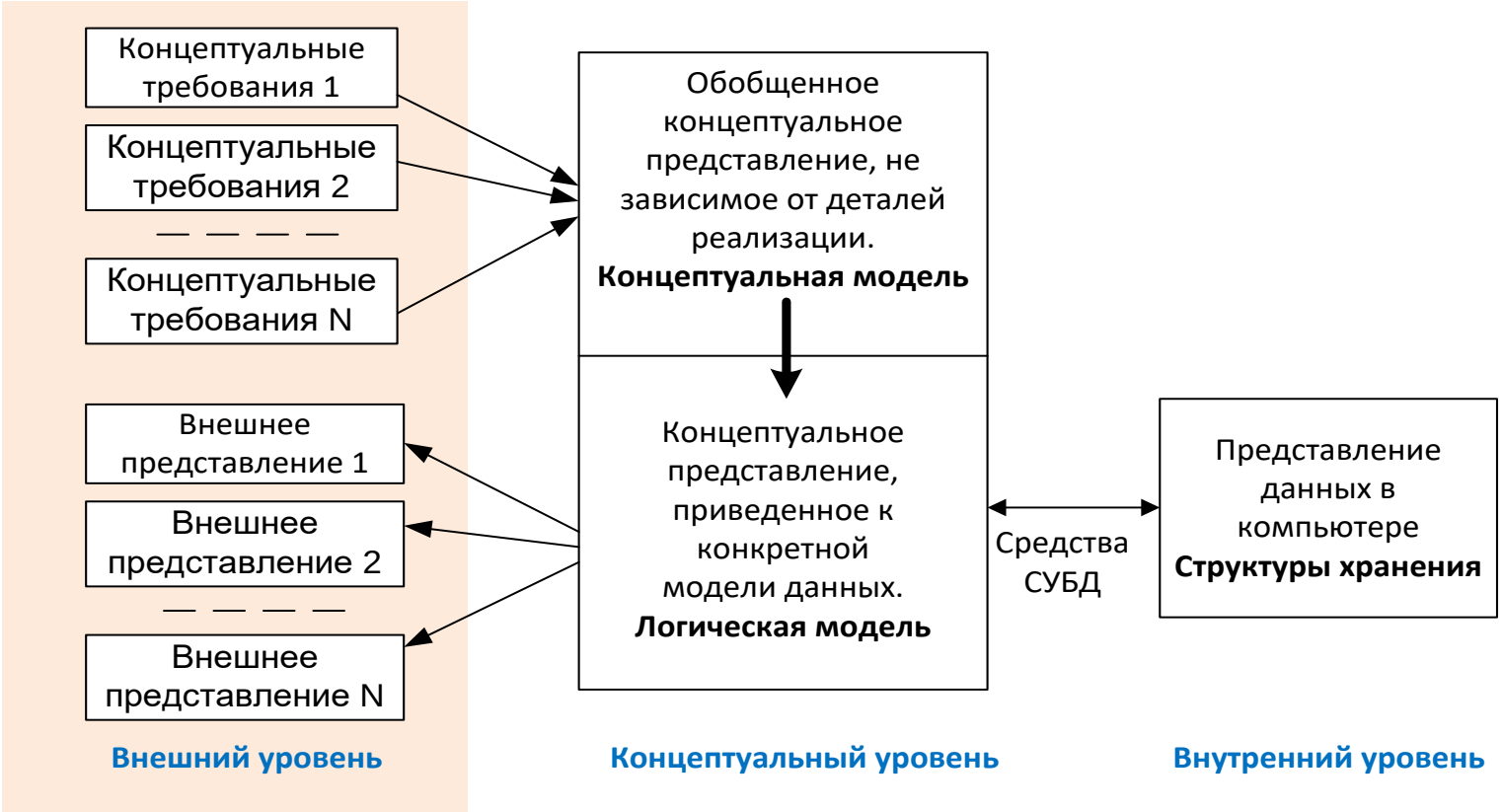
Этапы проектирования

- **Сбор данных**
- **Концептуальное проектирование** - создание концептуальной модели данных, исходя из представлений пользователей о предметной области
- **Логическое проектирование** - преобразование концептуальной модели на основе выбранной модели данных в логическую модель, не зависящую от особенностей СУБД
- **Физическое проектирование** - описание конкретной реализации базы данных

Три уровня данных

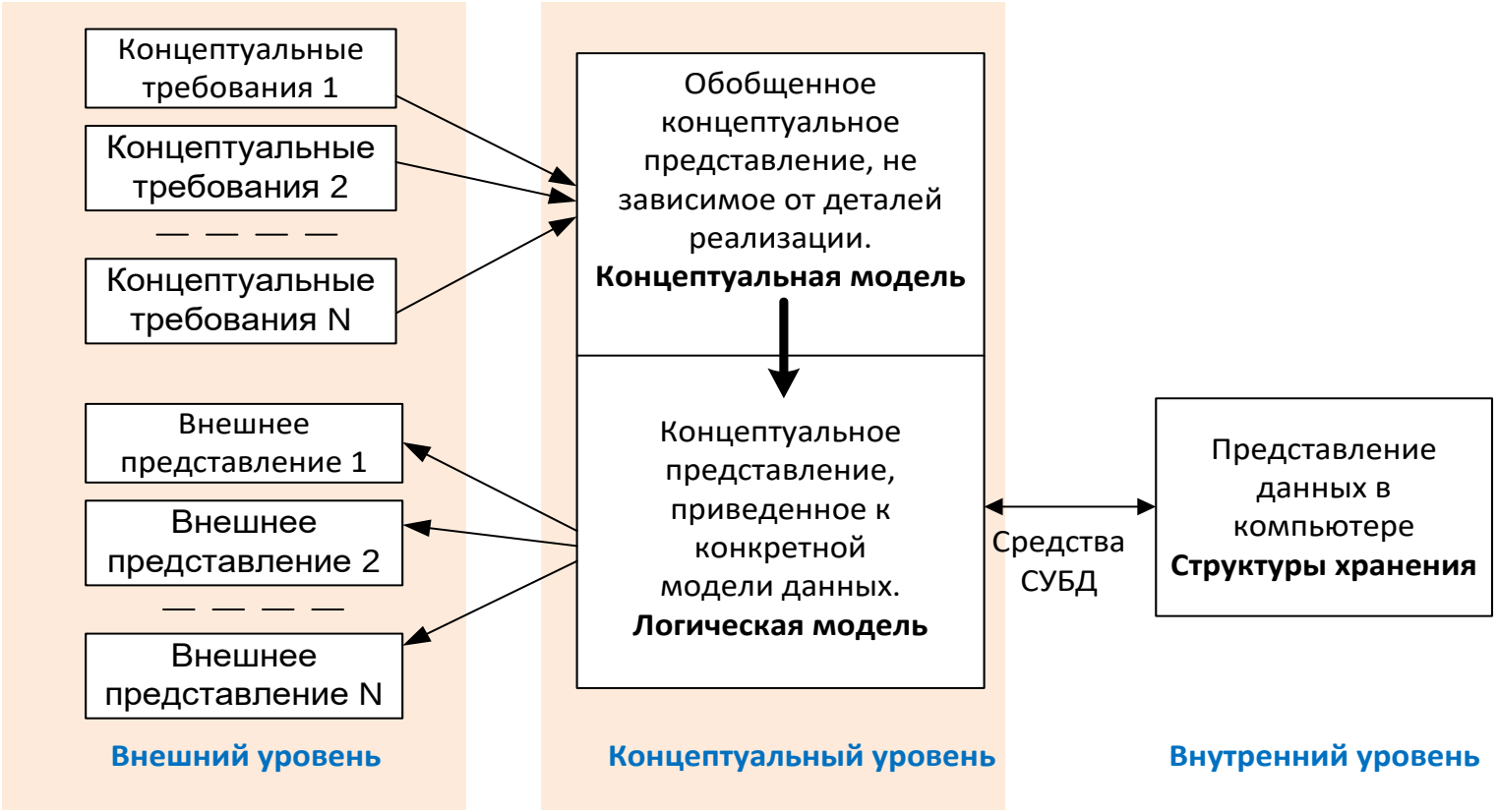


Основная цель - отделение пользовательского представления о данных от их физического представления.



Внешнее представление – для специалиста предметной области (пользователя). Работают только с интерфейсом, с элементами управления.

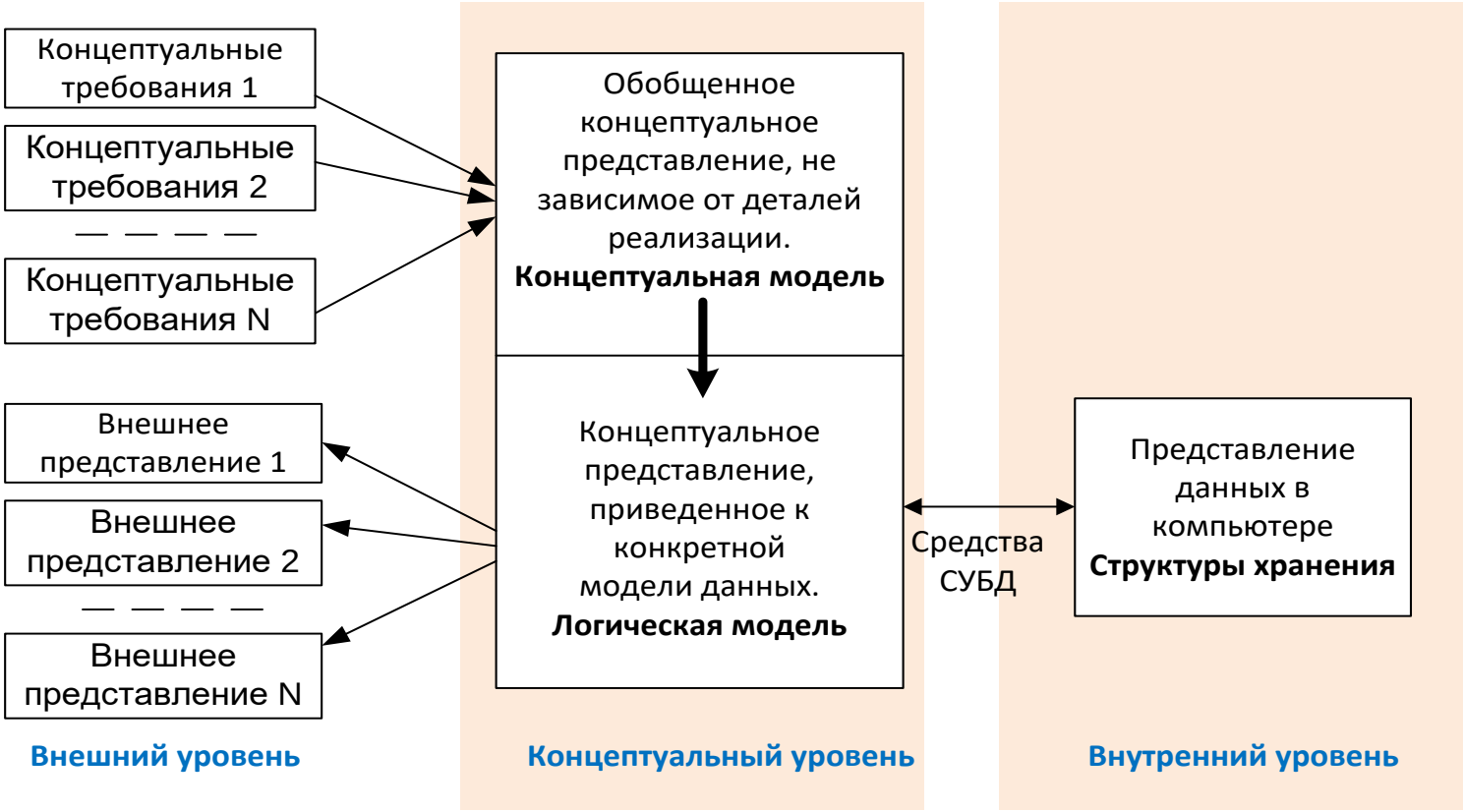
Пользовательский интерфейс должен быть удобным, понятным, связанным с предметной областью.



Внешнее представление и логическая модель – для прикладного программиста.

Программисты разрабатывают:

- пользовательский интерфейс и функционал приложения (это связано с предметной областью),
- взаимодействие с логическими структурами данных (абстракция, не связанная с предметной областью).



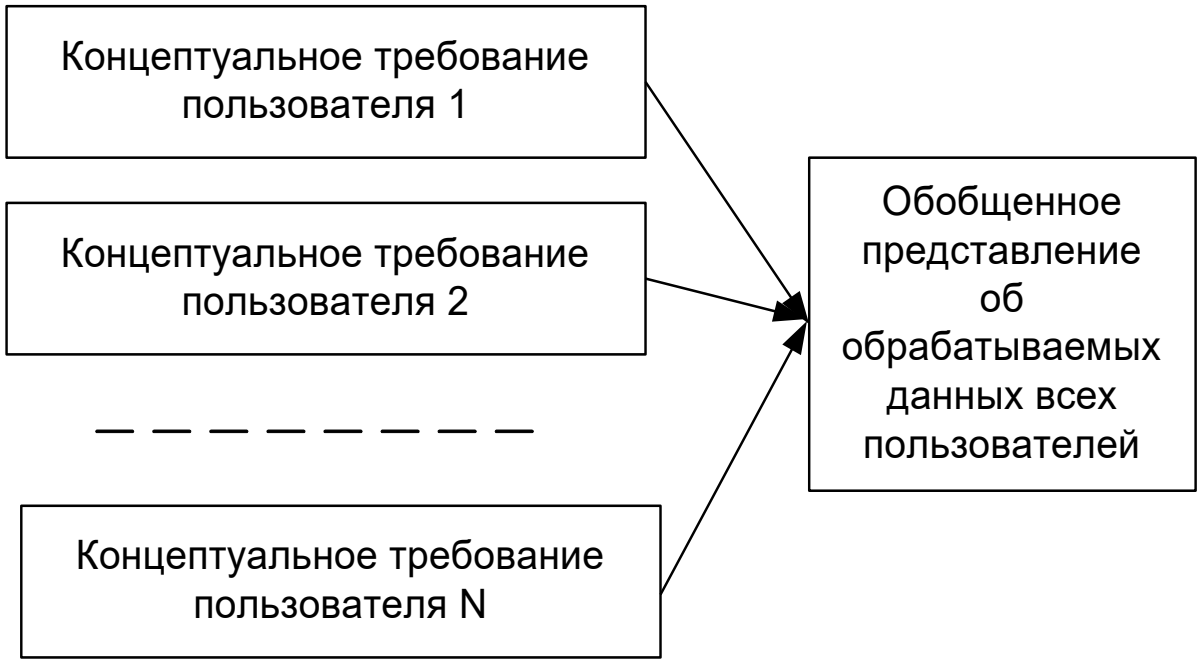
Логическая модель и внутреннее представление – для администратора базы данных.

Администратор БД может быть полностью абстрагирован от предметной области.

Концептуальная модель. ER-диаграммы

Концептуальная модель базы данных

Информационное описание предметной области с учетом логических взаимосвязей.



Терминология

Сущность (Entity) или объект – то, о чем будет накапливаться информация в информационной системе.

Правильная (сильная) сущность не зависит от существования других объектов. **Слабая сущность** не может существовать при отсутствии некоторой другой сущности.

Атрибут – поименованное свойство (характеристика) сущности. Совокупность атрибутов = **запись об объекте**.

Конкретная сущность = **экземпляр сущностей**. Экземпляры сущностей должны однозначно идентифицироваться по одному или нескольким атрибутам.

Связь (отношение) описывает имеющуюся в предметной области логическую связь между сущностями.

Сущность: **СТУДЕНТ**

Атрибуты:

- номер зачетной книжки,
- фамилия,
- дата рождения,
- место рождения

Экземпляр сущностей: {123456, Иванов, 01.01.2000, Саранск}

1 этап - концептуальное проектирование

- Определение сущностей
- Определение связей между сущностями
- Создание ER-модели предметной области
- Определение атрибутов (имя, тип, значение по умолчанию, NULL)
- Определение наборов допустимых значений атрибутов
- Определение первичных ключей для сущностей

Определение сущностей:

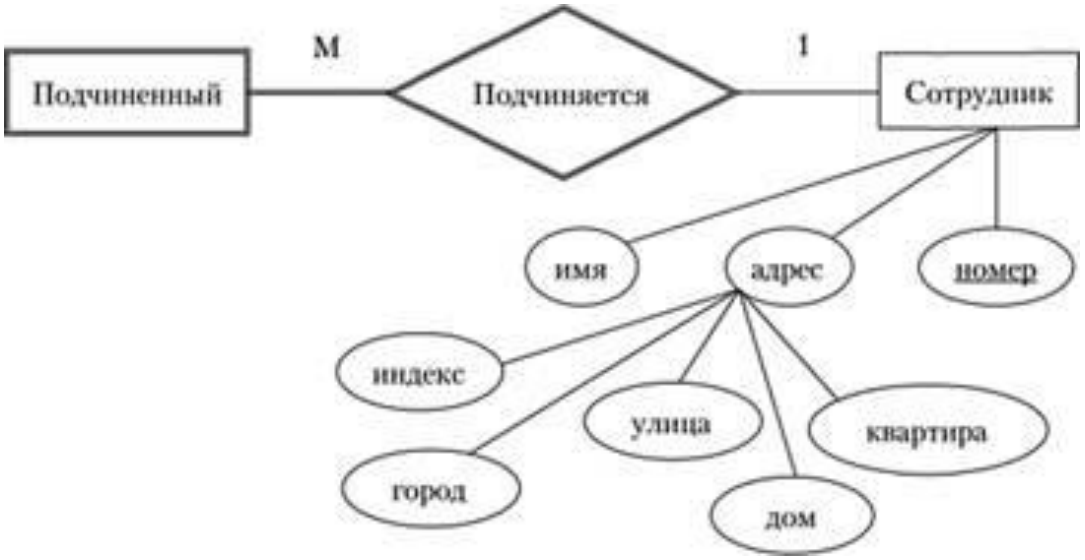
- Понимание, какая информация должна храниться и обрабатываться в виде сущности.
- Присвоение сущности имени.
- Выявление атрибутов сущности и присвоение им имени.
- Исключение синонимов и ононимов имен.
- Определение уникального идентификатора сущности.

Определение связей:

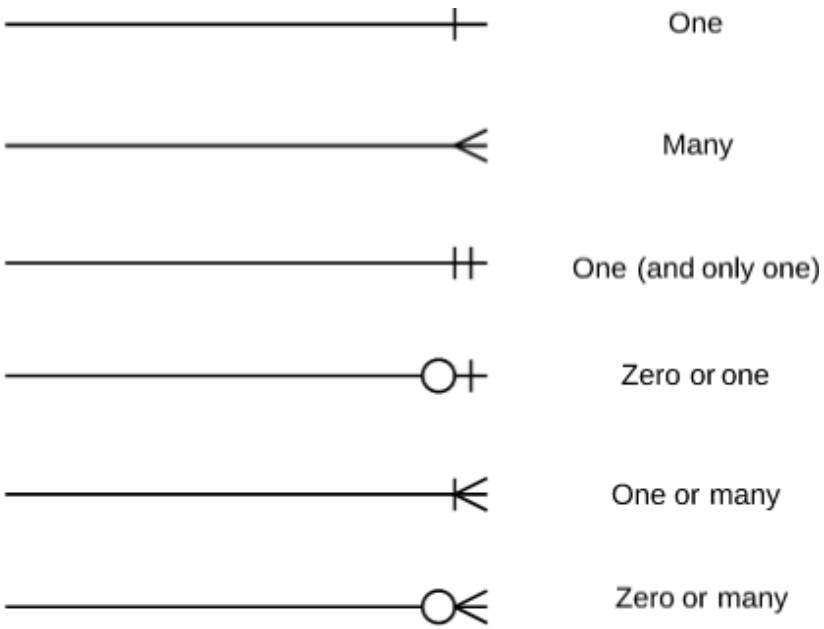
- Выявление «естественных» связей экземпляров одной сущности с экземплярами другой.
- Установка связей, позволяющих ответить на все возможные запросы пользователей.
- Присвоение имен связям и определение их типа.

Диаграмма сущностей–связей (Entity–Relationship)

Нотация Чена



Нотация Мартина (Crow’s Foot)



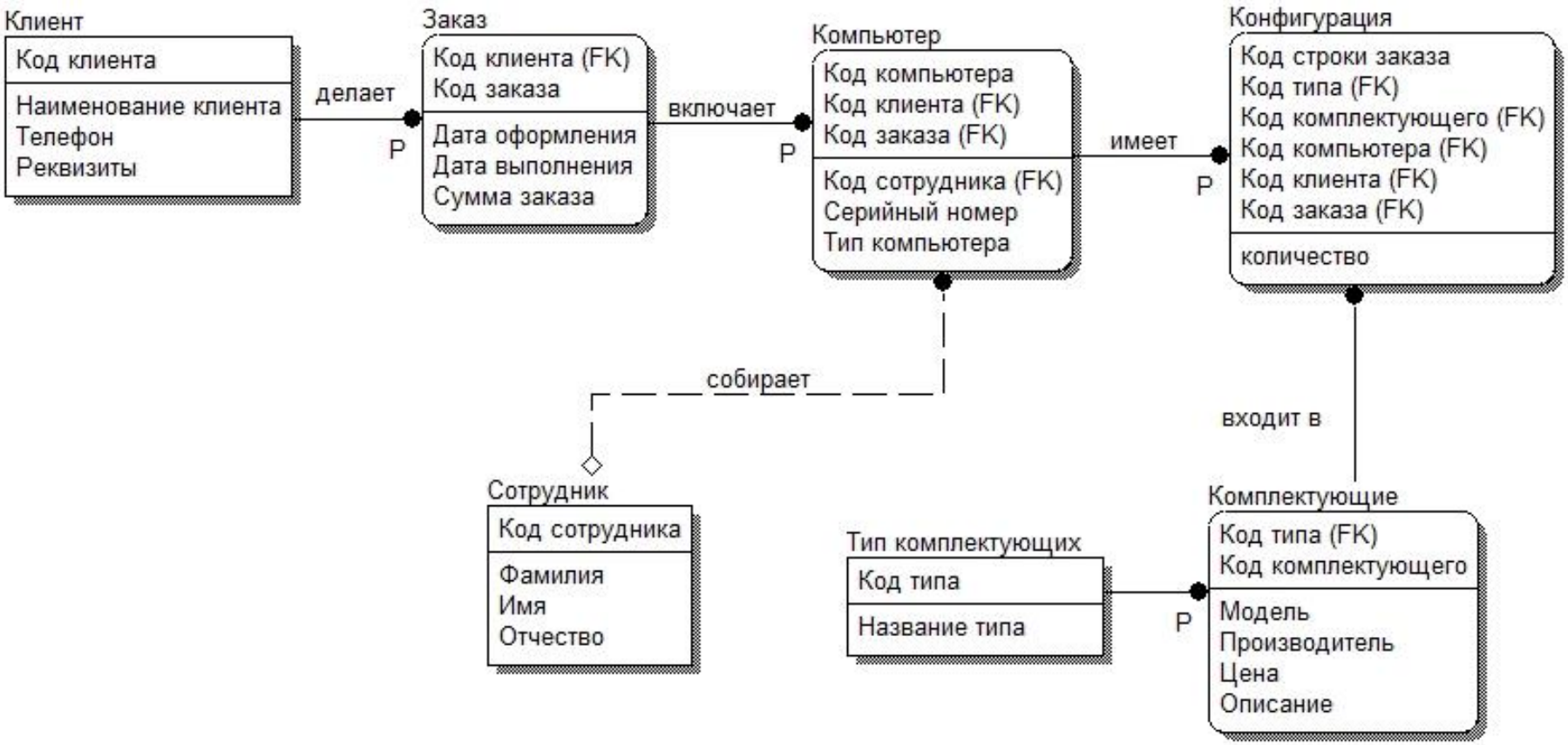
2 этап - логическое проектирование БД

- Выбор модели данных (реляционная, иерархическая, сетевая)
- Определение наборов структурированных данных. Нормализация таблиц (для реляционной модели)
- Проверка модели данных на предмет выполнения всех транзакций бизнес-процессов
- Определение требований поддержки целостности данных

Логическое проектирование БД

- **Сущности** концептуальной модели описываются как записи, состоящие из полей.
- Каждый **атрибут** описывается как поле с типом и характеристиками, возможными в выбранной модели данных.
- **Связи** концептуальной модели описываются в понятиях, соответствующих выбранной модели данных.
- Определяется порядок реализации запросов пользователей к базе данных с помощью типовых операций СУБД, поддерживающих выбранную модель данных, и т.д.

Нотация IDEF1X



3 этап - физическое проектирование БД

- Проектирование таблиц средствами выбранной СУБД
- Реализация бизнес-правил в среде выбранной СУБД
- Проектирование физической организации СУБД
- Разработка стратегии защиты БД
- Организация мониторинга функционирования БД и ее настройка

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»

Отпуск_товаров
Номер
Наименование
Дата
Покупатель
Адрес
Ед_изм
Цена
Количество
Сумма

Первичный ключ

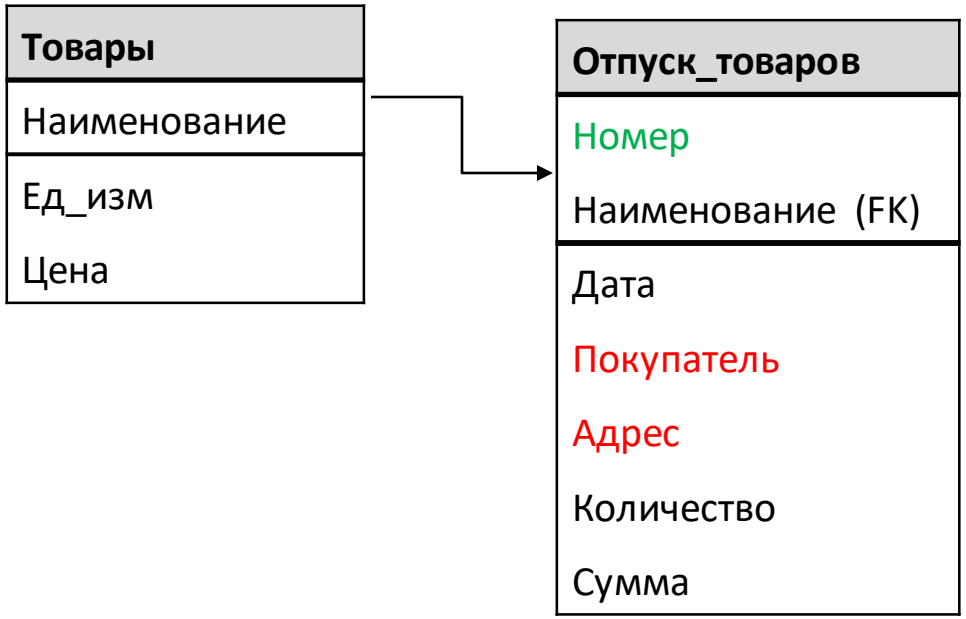
«Ед_изм» и «Цена» зависят только от «Наименование»

Товары
Наименование
Ед_изм
Цена

Отпуск_товаров
Номер
Наименование (FK)
Дата
Покупатель
Адрес
Количество
Сумма

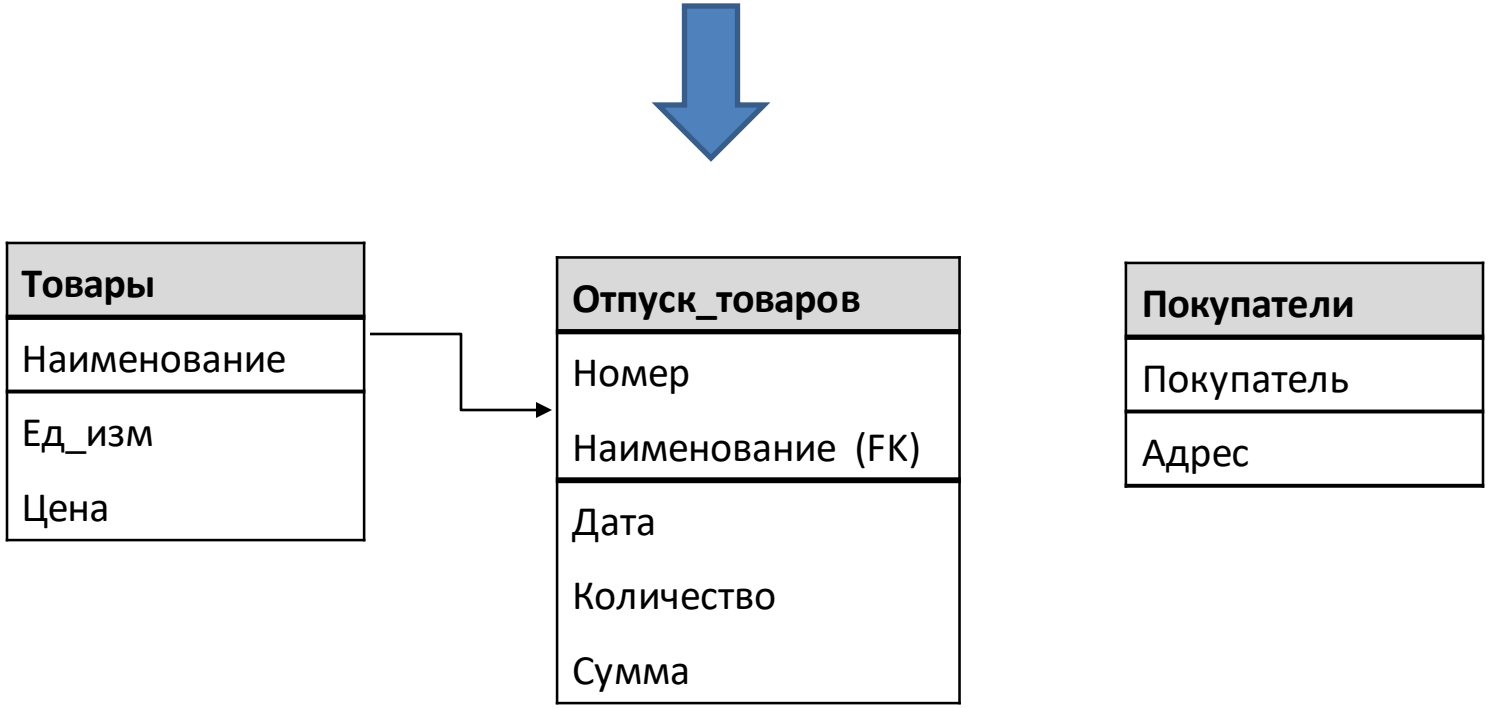
Выделение таблицы «Покупатели»

«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

Выделение таблицы «Накладные»

«Дата» зависит только от номера накладной

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"

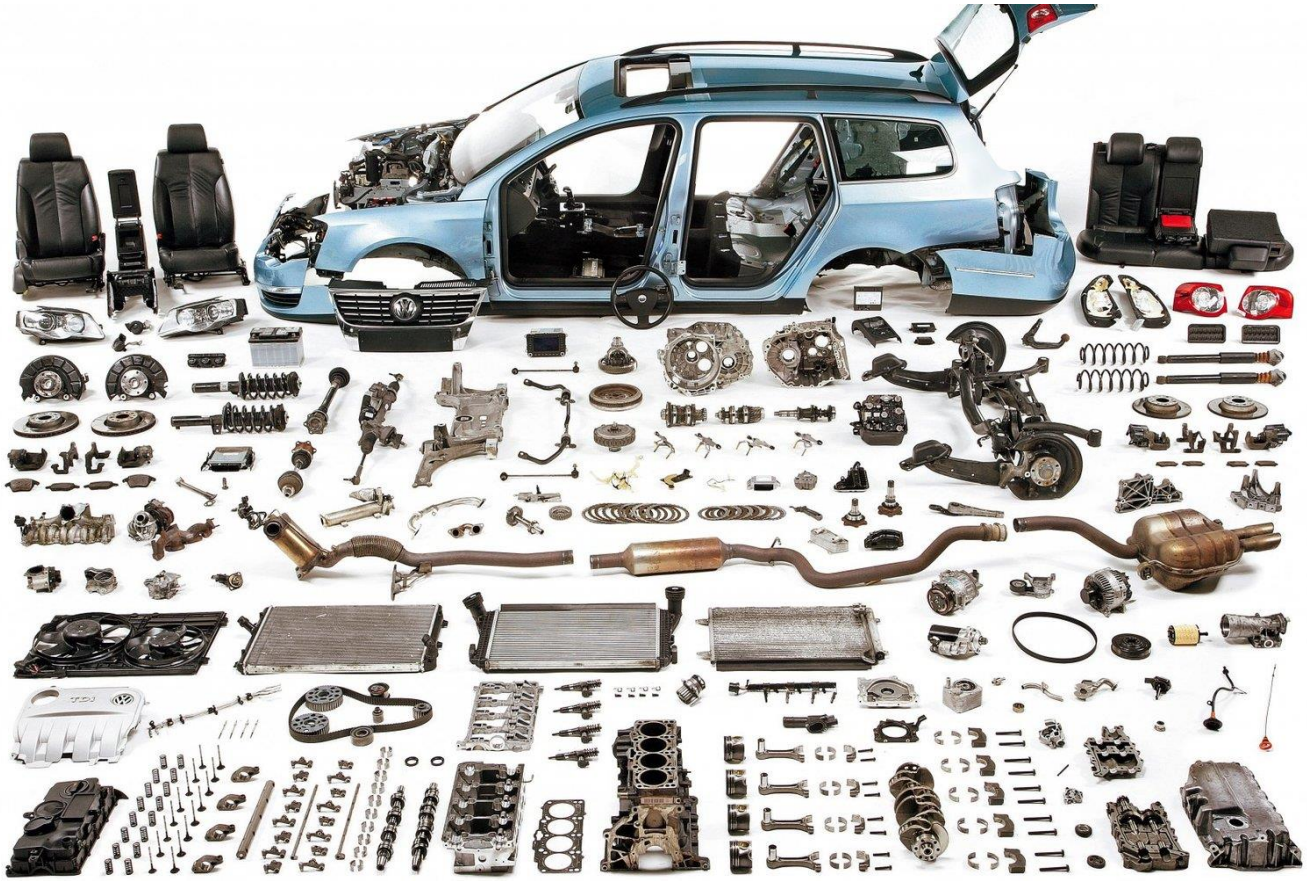
БД приведена к 3-й НФ

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