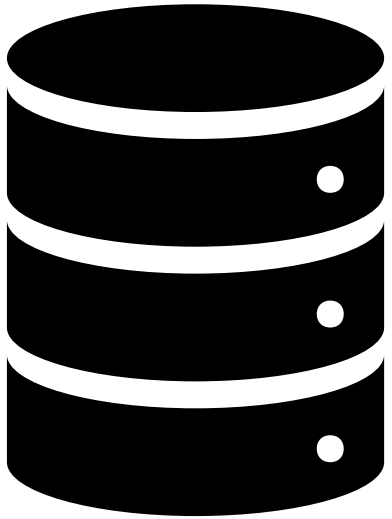


Базы данных

Лекция 4.

Проектирование баз данных (Часть 1)

Меркурьева Надежда
✉ merkurievanad@gmail.com
@merkurievanad

Проектирование базы данных

- Процесс создания детализированной модели данных* БД, а также необходимых ограничений целостности

- Модель данных – абстрактная модель, которая:
  - Организует элементы данных
  - Описывает, как они взаимодействуют друг с другом
  - Описывает, как они взаимодействуют с объектами внешнего мира

Основные задачи проектирования БД

- Обеспечение хранения в БД всей необходимой информации
- Обеспечение возможности получения данных по всем необходимым запросам
- Сокращение избыточности и дублирования данных
- Обеспечение целостности базы данных

Основные шаги проектирования

1. Определение данных, которые будут храниться в БД
2. Определение взаимосвязей между различными элементами данных
3. Наложение логической структуры на данные на основе определенных соотношений
4. Создание спроектированной БД с использованием СУБД

Этапы проектирования БД

Концептуальное (инфологическое) проектирование

Логическое (даталогическое) проектирование

Физическое проектирование

Шаг 1. Определение данных

- Ожидание:
  - Найти эксперта по проектированию БД
  - Найти эксперта в предметной области БД
  - Делегировать задачу этим людям

- Реальность:
  - Разобраться во всем самостоятельно

Шаг 1. Определение данных

- Какую предметную область описываем?
  - Полная база товаров магазинов «Пятерочка»
  - Полная клиентская база банка «Сбербанк»
  - Учебные процессы «МФТИ»
  - И т.д.

Шаг 1. Определение данных

- Как будем описывать? Как будем детализировать?
  - Полная база товаров магазина «Пятерочка»:
    - Каталог товаров
    - Каталог складов
    - Актуальные характеристики наполнения склада
    - Планы поставок
    - И т.д.

Шаг 1. Определение данных

- Как будем описывать? Как будем детализировать?
  - Полная клиентская база банка «Сбербанк»:
    - Клиент
    - Дебетовые продукты клиента
    - Кредитные продукты клиента
    - Транзакции
    - И т.д.

Шаг 1. Определение данных

- Как будем описывать? Как будем детализировать?
  - Учебные процессы «МФТИ»
    - Студент
    - Группа
    - Преподаватель
    - Расписание пар
    - Расписание экзаменов
    - Распределение аудиторий
    - И т.д.

Концептуальное проектирование

- **Концептуальная модель данных** – описание основных объектов и отношений между ними

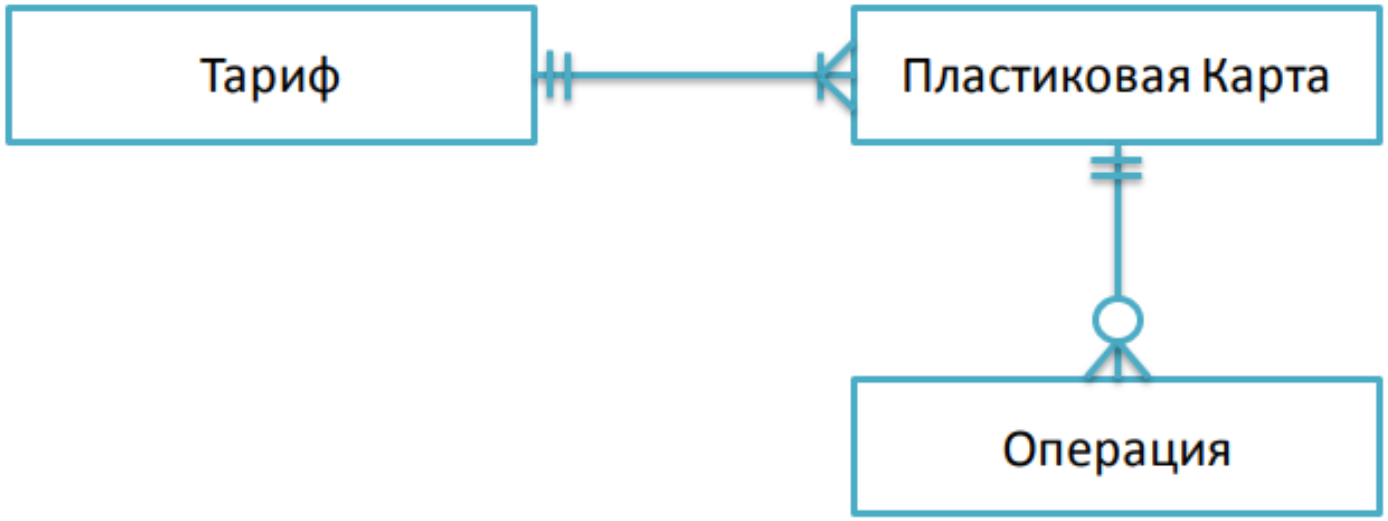
- Про описание объектов поговорили. Что насчет связей?

ER-модель

- ER-модель (entity-relationship model) – модель данных, позволяющая описывать концептуальные и логические схемы

- Задаются:
  - Сущности
  - Связи

ER-модель

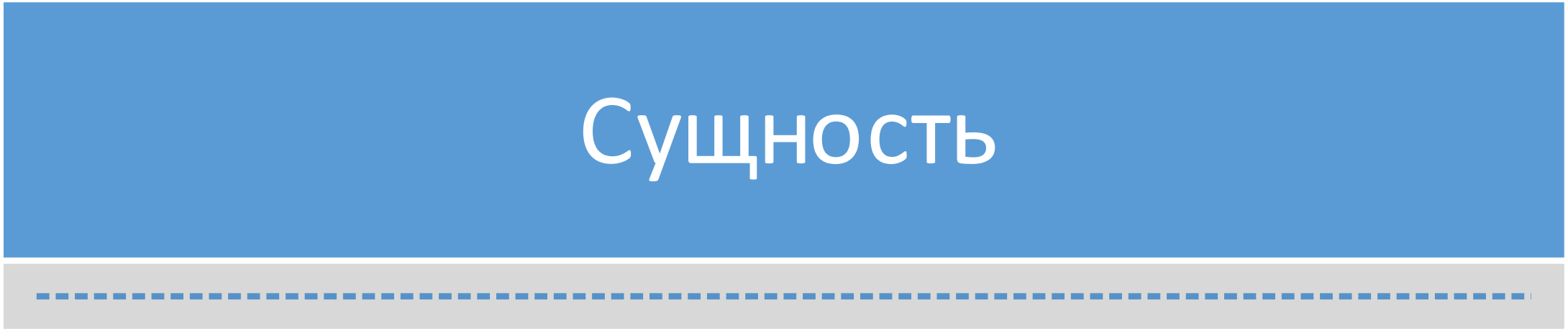


Шаг 2. Определение взаимосвязей

- Определили предметную область и детализировали разбиение на сущности

- Необходимо установить взаимосвязи

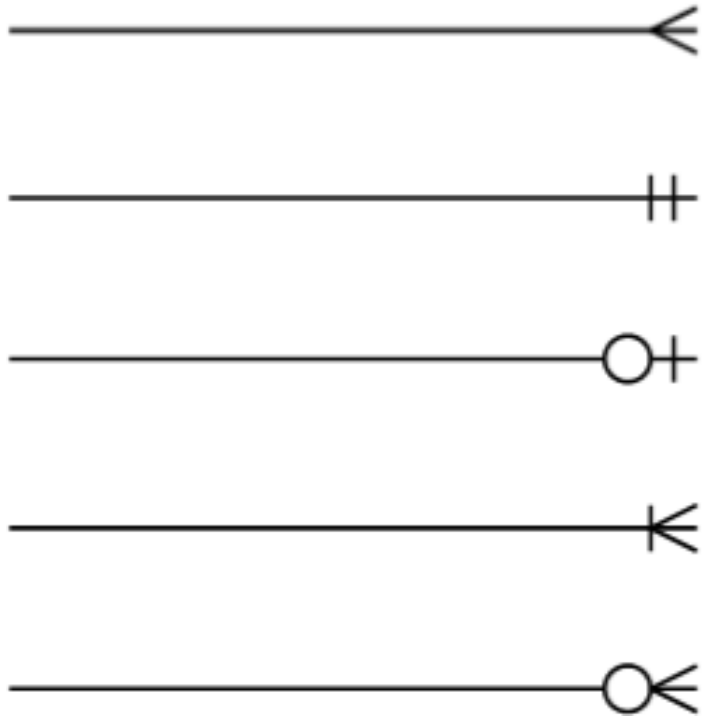
Нотация «Сущность»



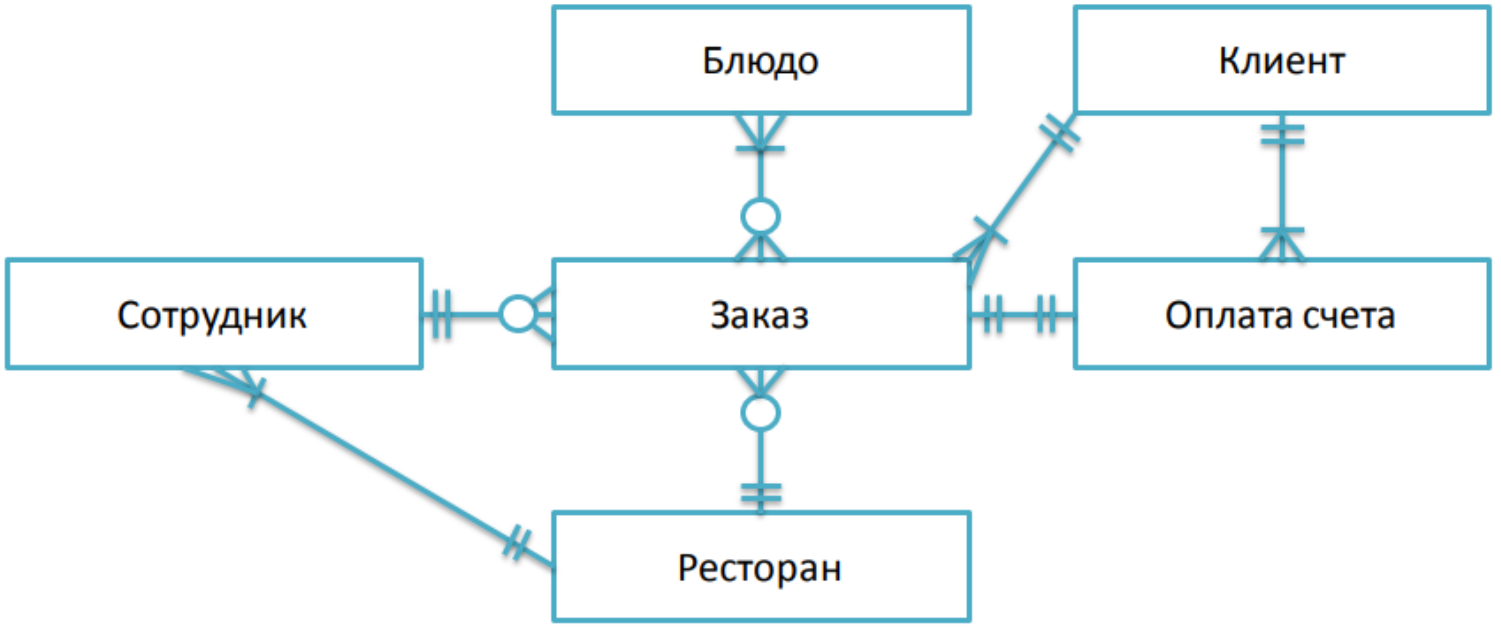
Нотация «Воронья лапка»

- В оригинале Crow’s Foot notation

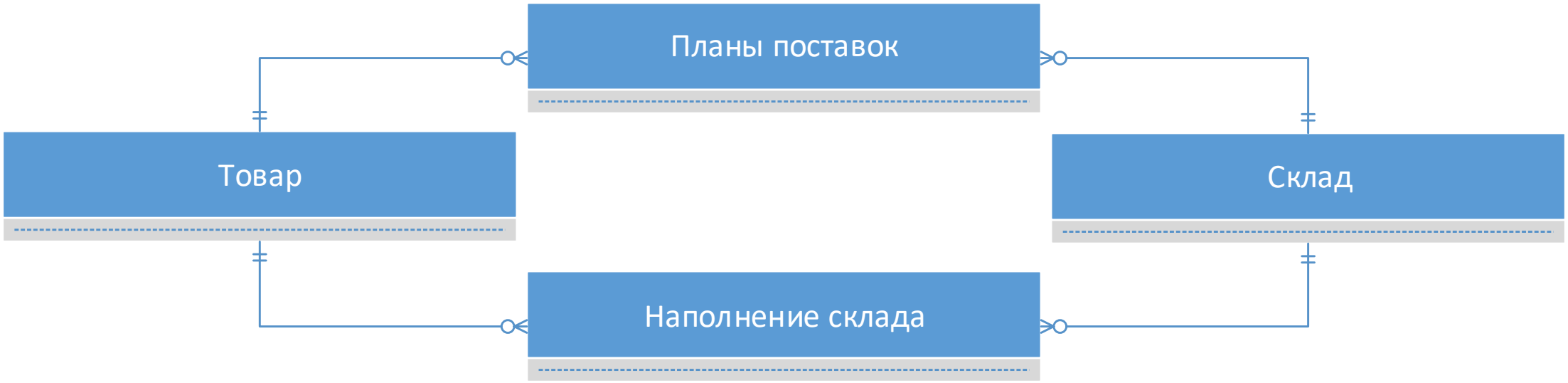
  - ✓ Многие (строго больше одного)
  - ✓ Один и только один
  - ✓ Один или ноль
  - ✓ Многие или один
  - ✓ Многие или один или ноль



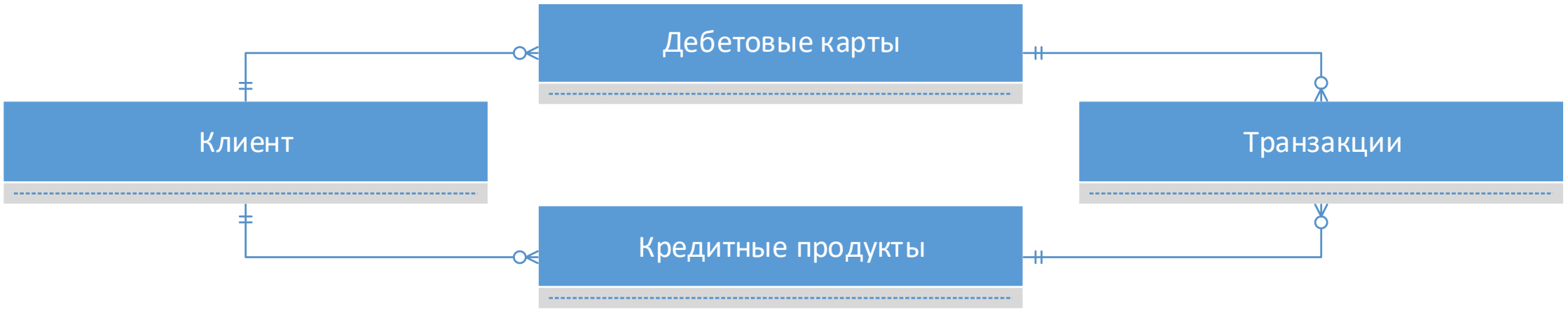
Концептуальная модель БД сети ресторанов



Примеры ER-диаграмм: «Пятерочка»



Примеры ER-диаграмм: «Сбербанк»



Шаг 3. Создание логической структуры

- Уже имеем концептуальную схему

- Хотим получить больший уровень детализации:
  - Уточнение атрибутивного состава
  - Уточнение ограничений, накладываемых на атрибутивный состав
  - Нормализация отношений
  - В некоторых случаях допускается уточнение типа атрибута

Потенциальный ключ

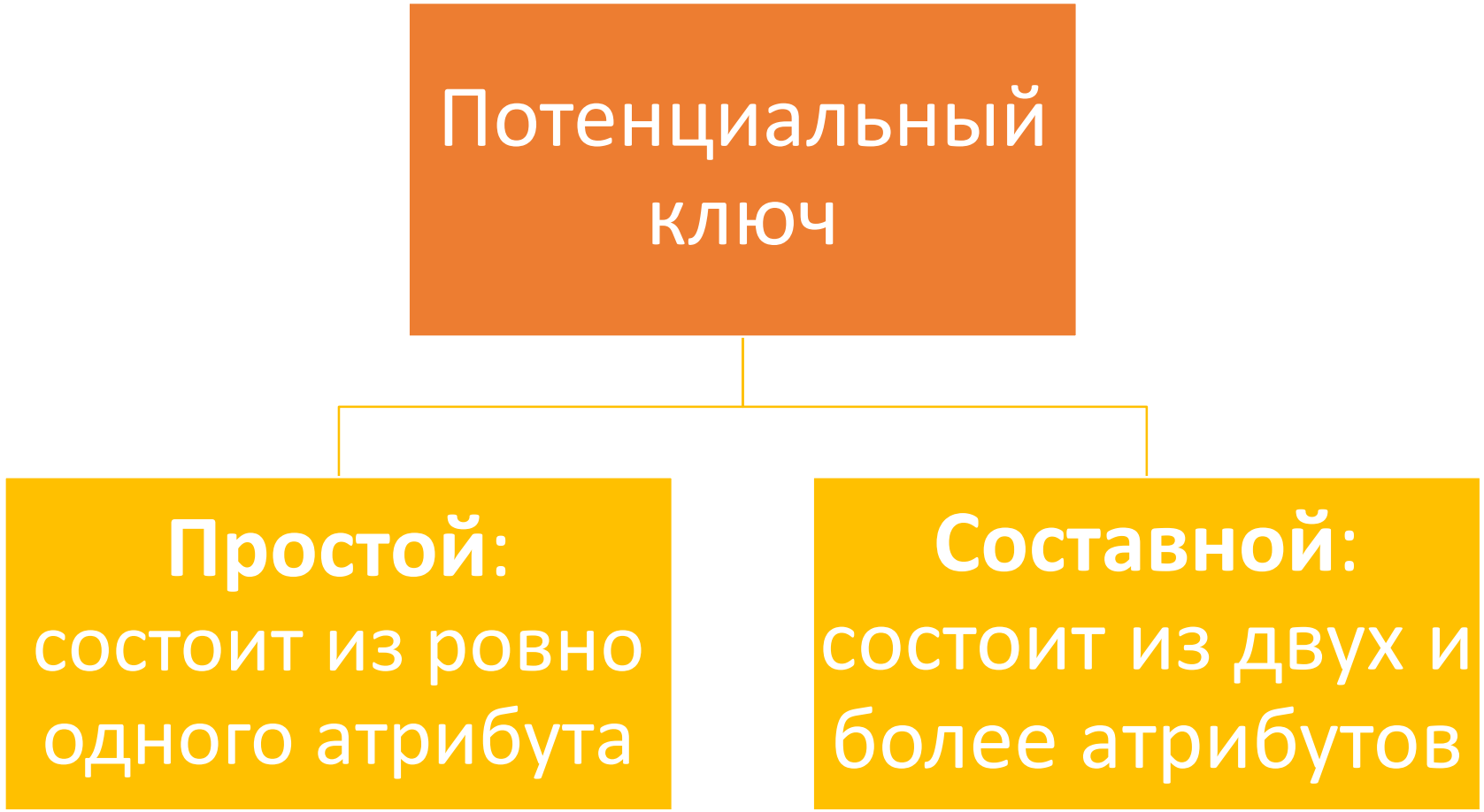
– это подмножество атрибутов отношения, удовлетворяющее требованиям уникальности и минимальности

- *Уникальность*: нет и не может быть двух кортежей данного отношения, в которых значения этого подмножества атрибутов совпадают
- *Минимальность*: в составе потенциального ключа отсутствует меньшее подмножество атрибутов, удовлетворяющее условию уникальности

Потенциальный ключ

- Из свойства отношения: потенциальный ключ существует **всегда**, даже если он включает **все** атрибуты отношения
- Допустимо наличие нескольких потенциальных ключей в отношении

Классификация по признаку общности



Первичный ключ

– это один из потенциальных ключей отношения, выбранный в качестве основного (Primary key, PK)

- Если в отношении имеется лишь один потенциальный ключ, он и будет первичным ключом
- Если потенциальных ключей несколько, то:
  - Один из них выбирается в качестве первичного
  - Остальные ключи называются ***альтернативными***
- Недопустимо отсутствие значения

Первичный ключ

- Теоретически, все потенциальные ключи одинаково пригодны для использования в качестве первичного ключа
- На практике, используют тот потенциальный ключ, который:
  - Занимает меньше места при хранении
  - Не утратит свою уникальность со временем

Суррогатный ключ

– это дополнительное служебное поле, которое добавляется к уже имеющимся информационным полям таблицы, единственное предназначение которое – служить первичным ключом

- Значение такого поля генерируется искусственно
- Не стоит искать в нем какой-то глубинный смысл

- Ключ, который основан на уже существующем поле, называется ***естественным***
- Ключ, который основан на естественном ключе путем добавления дополнительного поля, называется ***интеллектуальным***

Суррогатный ключ

- Обычно суррогатный ключ – числовое поле
- Значения суррогатного ключа генерируется арифметической прогрессией с шагом 1
- В ряде СУБД существует специальный тип данных, автоматически генерирующий такую последовательность:
  - PostgreSQL – SERIAL
  - MySQL – AUTO_INCREMENT

Достоинства суррогатных ключей

- **Неизменность**: заполнили одним значением раз и навсегда (кроме экстраординарных ситуаций)
- **Гарантированная уникальность**: т.к. значение генерируется автоинкрементом, повторение значений исключено
- **Гибкость**: т.к. такой ключ не несет никакой информативной нагрузки, его можно свободно заменить
- **Проще программировать**: позволяет не завязывать на структуре конкретной БД. Особенно удобно для языков со статической типизацией
- **Эффективность**: удобнее при создании ссылок на другие таблицы

Недостатки суррогатных ключей

- **Уязвимость генераторов**: по номерам ключей возможно узнать число новых записей за определенный период времени
- **Неинформативность**: усложняется ручная проверка БД
- **Склоняет администратора пропустить нормализацию**: вместо того, чтобы разбить отношение на несколько отношений и аккуратно учесть все связи, велик соблазн просто создать суррогатный ключ
- **Вопросы оптимизации**: необходимость поддержания и суррогатного, и естественного ключей (об этом поговорим на лекции 6)
- **Невольная привязка разработчика к поведению генератора ключей в конкретной СУБД**

Внешний ключ

Пусть R_1 и R_2 – две переменные отношения, не обязательно различные. ***Внешним ключом FK*** в R_2 является подмножество атрибутов переменной R_2 такое, что выполняются следующие требования:

- В переменной отношения R_1 имеется потенциальный ключ PK такой, что PK и FK совпадают с точностью до переименования атрибутов
- В любой момент времени каждое значение FK в текущем значении R_2 идентично значению PK в некотором кортеже в текущем значении R_1. Иными словами, в любой момент времени множество всех значений FK в R_2 является подмножеством значений PK в R_1.

Внешний ключ

- Отношение R_1, содержащее потенциальный ключ, называется ***главным***, ***целевым*** или ***родительским***
- Отношение R_2, содержащее внешний ключ, называется ***подчиненным*** или ***дочерним***

Внешний ключ

ID (PK)	CITY
1	Москва
2	Санкт-Петербург
3	Владивосток

ID (PK)	STREET	CITY_ID (FK)
181	Малая Бронная	1
182	Тверской бульвар	1
183	Невский проспект	2
184	Пушкинская	2
185	Светланская	3
186	Пушкинская	3

Ссылочная целостность

– это необходимое качество реляционной базы данных, заключающееся в отсутствии в любом её отношении внешних ключей, ссылающихся на несуществующие кортежи

- База данных обладает свойством ссылочной целостности, когда для любой пары связанных внешним ключом отношений в ней условие ссылочной целостности выполняется

Поддержание ссылочной целостности

- **CASCADE**
  - При удалении / изменении строки главной таблицы соответствующая запись дочерней таблицы также будет удалена / изменена
- **RESTRICT**
  - Строка не может быть удалена / изменена, если на нее имеется ссылка
  - Значение не может быть удалено / изменено, если на него есть ссылка
- **NO ACTION**
  - Похож на RESTRICT, только проверка происходит при операции ALTER TABLE, а не при UPDATE или DELETE
- **SET NULL**
  - При удалении записи главной таблицы, соответствующее значение дочерней таблицы становится NULL
- **SET DEFAULT**
  - Аналогично SET NULL, только вместо значения NULL устанавливается некоторое значение по умолчанию

Нормальная форма

- **Нормальная форма** — свойство отношения в реляционной модели данных, характеризующее его с точки зрения избыточности, потенциально приводящей к логически ошибочным результатам выборки или изменения данных

- Нормальная форма определяется как совокупность требований, которым должно удовлетворять отношение

- Приведение БД к нормальной форме – нормализация
  - Каждая следующая форма включает в себя ограничения предыдущих

Нормализация БД

- Предназначена:
  - Минимизация логической избыточности
  - Уменьшение потенциальной противоречивости

- Не предназначена:
  - Уменьшение / увеличение производительности БД
  - Уменьшение / увеличение физического объема БД

Нормализация БД

- Исключение некоторых типов избыточности
- Устранение некоторых аномалий* обновления
- Разработка проекта БД, который является:
  - Качественным представление реального мира
  - Интуитивно понятен
  - Легко расширяем в дальнейшем
- Упрощение процедуры применения необходимых ограничений целостности

Аномалии

- Ситуация в таблице БД такая, что:
  - Существенно осложнена работа с БД
  - В БД присутствует противоречия

- Причина:
  - Излишнее дублирование данных в таблице

Аномалии модификации

- Изменение данных одной записи влечет за собой необходимость изменения аналогичных данных еще некоторых записей

Номер поставки (PK)	Название товара (PK)	Цена товара	Количество	Дата поставки	Название поставщика	Адрес поставщика
1	Карандаш	15	10000	12.10.2017	Поставщик_1	Адрес_1
2	Клей	30	1500	03.03.2018	Поставщик_1	Адрес_1
2	Тетрадь	5	10000	03.03.2018	Поставщик_2	Адрес_2
3	Ручка	5	13000	05.03.2018	Поставщик_1	Адрес_1
3	Блокнот	50	20000	05.03.2018	Поставщик_1	Адрес_1
3	Альбом	100	25000	05.03.2018	Поставщик_2	Адрес_2

- Хотим изменить адрес поставщика 1:
  - Придется менять адрес во всех строках

Аномалии удаления

- Удаление определенных записей несет потерю информации, которую удалять не хотели

Номер поставки (PK)	Название товара (PK)	Цена товара	Количество	Дата поставки	Название поставщика	Адрес поставщика
1	Карандаш	15	10000	12.10.2017	Поставщик_1	Адрес_1
2	Клей	30	1500	03.03.2018	Поставщик_1	Адрес_1
2	Тетрадь	5	10000	03.03.2018	Поставщик_2	Адрес_2
3	Ручка	5	13000	05.03.2018	Поставщик_1	Адрес_1
3	Блокнот	50	20000	05.03.2018	Поставщик_1	Адрес_1
3	Альбом	100	25000	05.03.2018	Поставщик_2	Адрес_2

- Хотим удалить записи о поставках от поставщика 2:
  - Теряем всю информации о поставщике 2, включая его адрес

Аномалии добавления

- Не можем добавить новую запись, если неизвестны значения первичных ключей

Номер поставки (PK)	Название товара (PK)	Цена товара	Количество	Дата поставки	Название поставщика	Адрес поставщика
1	Карандаш	15	10000	12.10.2017	Поставщик_1	Адрес_1
2	Клей	30	1500	03.03.2018	Поставщик_1	Адрес_1
2	Тетрадь	5	10000	03.03.2018	Поставщик_2	Адрес_2
3	Ручка	5	13000	05.03.2018	Поставщик_1	Адрес_1
3	Блокнот	50	20000	05.03.2018	Поставщик_1	Адрес_1
3	Альбом	100	25000	05.03.2018	Поставщик_2	Адрес_2

- Заключили контракт с поставщиком 3:
  - Не можем добавить информацию о нем в таблицу, т.к. еще не было поставок

Нормализация БД

- Нормализация БД производится за счет декомпозиции отношения

- Декомпозиция – разложение исходной переменной отношения на несколько эквивалентных
- Декомпозиция обратна соединению

- Декомпозиция называется ***декомпозицией без потерь*** или ***правильной***, если она обратима

Нормальные формы

- Первая нормальная форма (1NF)
- Вторая нормальная форма (2NF)
- Третья нормальная форма (3NF)
- Нормальная форма Бойса-Кодда (BCNF)
- Четвертая нормальная форма (4NF)
- Пятая нормальная форма / Нормальная форма проекции-соединения (5NF / PJNF)
- Доменно-ключевая нормальная форма (DKNF)
- Шестая нормальная форма (6NF)

Нормальные формы

- **Первая нормальная форма (1NF)**
- **Вторая нормальная форма (2NF)**
- **Третья нормальная форма (3NF)**
- Нормальная форма Бойса-Кодда (BCNF)
- Четвертая нормальная форма (4NF)
- Пятая нормальная форма / Нормальная форма проекции-соединения (5NF / PJNF)
- Доменно-ключевая нормальная форма (DKNF)
- Шестая нормальная форма (6NF)

Первая нормальная форма

- Переменная отношения находится в ***первой нормальной форме*** тогда и только тогда, когда значения всех атрибутов отношения атомарны

Первая нормальная форма

- Переменная отношения находится в ***первой нормальной форме*** тогда и только тогда, когда значения всех атрибутов отношения атомарны

- Отношение находится в 1NF, если все его атрибуты являются простыми. Все используемые домены содержат только скалярные значения

Первая нормальная форма

Группа	Студент
291	Шехтер
293	Гусев, Медведева, Меркурьева, Шапошников
298	Мавлютов

Таблица не находится в 1NF, т.к. студенты группы 293 записаны единым списком в одной строке

Первая нормальная форма

Приведем таблицу к 1NF:

Группа	Студент
291	Шехтер
293	Гусев
293	Медведева
293	Меркурьева
293	Шапошников
298	Мавлютов

Вторая нормальная форма

- Переменная отношения находится во ***второй нормальной форме*** тогда и только тогда, когда она находится в первой нормальной форме, и каждый неключевой атрибут минимально функционально зависит от потенциального ключа
- ***Функциональная зависимость*** между множествами атрибутов X и Y означает, что для любого допустимого набора кортежей в данном отношении верно следующее: если два кортежа совпадают по значению X, то они совпадают по значению Y
- ***Минимальная функциональная зависимость*** означает, что в составе первичного ключа отсутствует меньшее подмножество атрибутов, от которого можно также вывести данную функциональную зависимость

Вторая нормальная форма

- Любая переменная отношения, находящаяся в 1NF, но не находящаяся в 2NF, может быть приведена к набору переменных отношений, находящихся в 2NF

- В результате декомпозиции получим набор проекций исходной переменной отношения, причем обратимый

Вторая нормальная форма

Название группы	Название CD-диска	Название песни	Автор слов	Композитор
Scorpions	World Wide Live	Countdown	Klaus Meine	Matthias Jabs
Scorpions	World Wide Live	Coming Home	Rudolf Schenker	Klaus Meine
Scorpions	World Wide Live	Blackout	Rudolf Schenker	Klaus Meine
Scorpions	Blackout	Blackout	Rudolf Schenker	Klaus Meine
The Big City	Blackout	Blackout	Rudolf Schenker	Klaus Meine

Вторая нормальная форма

Название группы	Название CD-диска	Название песни	Автор слов	Композитор
Scorpions	World Wide Live	Countdown	Klaus Meine	Matthias Jabs
Scorpions	World Wide Live	Coming Home	Rudolf Schenker	Klaus Meine
Scorpions	World Wide Live	Blackout	Rudolf Schenker	Klaus Meine
Scorpions	Blackout	Blackout	Rudolf Schenker	Klaus Meine
The Big City	Blackout	Blackout	Rudolf Schenker	Klaus Meine

Данная таблица находится в 1NF, но не во 2NF, т.к. автор слов и композитор зависят только от полей «Название группы» и «Название песни», т.е. от того, что песня включена на другой CD-диск, значения этих полей никак не изменятся

Вторая нормальная форма

Название группы	Название CD-диска	Название песни
Scorpions	World Wide Live	Countdown
Scorpions	World Wide Live	Coming Home
Scorpions	World Wide Live	Blackout
Scorpions	Blackout	Blackout
The Big City	Blackout	Blackout

Название группы	Название песни	Автор слов	Композитор
Scorpions	Countdown	Klaus Meine	Matthias Jabs
Scorpions	Coming Home	Rudolf Schenker	Klaus Meine
Scorpions	Blackout	Rudolf Schenker	Klaus Meine
Scorpions	Blackout	Rudolf Schenker	Klaus Meine
The Big City	Blackout	Rudolf Schenker	Klaus Meine

Вторая нормальная форма

Название группы	Название CD-диска	Название песни
Scorpions	World Wide Live	Countdown
Scorpions	World Wide Live	Coming Home
Scorpions	World Wide Live	Blackout
Scorpions	Blackout	Blackout
The Big City	Blackout	Blackout

Название группы	Название песни	Автор слов	Композитор
Scorpions	Countdown	Klaus Meine	Matthias Jabs
Scorpions	Coming Home	Rudolf Schenker	Klaus Meine
Scorpions	Blackout	Rudolf Schenker	Klaus Meine
The Big City	Blackout	Rudolf Schenker	Klaus Meine

Третья нормальная форма

- Переменная отношения находится в ***третьей нормальной форме*** (3NF) в том и только в том случае, когда она находится во второй нормальной форме, и каждый неключевой атрибут нетранзитивно функционально зависит от первичного ключа

- Иными словами, каждое неключевое поле должно содержать информацию о ключе, полном ключе и ни о чём, кроме ключа

Третья нормальная форма

Сотрудник	Отдел	Телефон
Иванов	Бухгалтерия	11-22-334
Петров	Бухгалтерия	11-22-334
Сидоров	Снабжение	22-33-445

Данная таблица находится в 2NF, но не во 3NF, т.к. поле «Телефон» зависит от поля «Отдел»

Функциональные зависимости:

- Сотрудник -> Отдел
- Отдел -> Телефон
- Сотрудник -> Телефон

Третья нормальная форма

Отдел	Телефон
Бухгалтерия	11-22-334
Снабжение	22-33-445

Сотрудник	Отдел
Иванов	Бухгалтерия
Петров	Бухгалтерия
Сидоров	Снабжение

Нормальная форма Бойса-Кодда

- Переменная отношения находится в **нормальной форме Бойса-Кодда (BCNF)** тогда и только тогда, когда каждая её нетривиальная и неприводимая слева функциональная зависимость имеет в качестве своего детерминанта некоторый потенциальный ключ
- Переменная отношения находится в BCNF тогда и только тогда, когда она находится в 3NF, и при этом не только любой неключевой атрибут полностью функционально зависит от любого ключа, но и любой ключевой атрибут должен полностью функционально зависеть от любого ключа

Нормальная форма Бойса-Кодда

Номер корта	Время начала	Время окончания	Тариф
1	09:30	10:30	«Корт 1 для членов клуба»
1	11:00	12:00	«Корт 1 для членов клуба»
1	14:00	15:30	«Корт 1 для не членов клуба»
2	10:00	12:00	«Корт 2 для не членов клуба»
2	11:30	13:30	«Корт 2 для не членов клуба»
2	15:00	16:30	«Корт 2 для членов клуба»

Потенциальные ключи:

- (Номер корта, Время начала)
- (Номер корта, Время окончания)
- (Тариф, Время начала)
- (Тариф, Время окончания)

Нормальная форма Бойса-Кодда

Тариф	Номер корта	Для членов клуба
«Корт 1 для членов клуба»	1	Да
«Корт 1 для не членов клуба»	1	Нет
«Корт 2 для членов клуба»	2	Да
«Корт 2 для не членов клуба»	2	Нет

Тариф	Время начала	Время окончания
«Корт 1 для членов клуба»	09:30	10:30
«Корт 1 для членов клуба»	11:00	12:00
«Корт 1 для не членов клуба»	14:00	15:30
«Корт 2 для не членов клуба»	10:00	12:00
«Корт 2 для не членов клуба»	11:30	13:30
«Корт 2 для членов клуба»	15:00	16:30

Четвертая нормальная форма

- Переменная отношения находится в **четвёртой нормальной форме (4NF)**, если она находится в 3НФ и все многозначные зависимости фактически являются функциональными зависимостями от её потенциальных ключей

- 5NF, DKNF, 6NF – для любителей формализовывать все и вся

Зачем вообще это нужно?

- На первых порах позволяет проектировать неплохие БД
- Помогает избегать типичных ошибок при проектировании БД людям без опыта
- Формирует привычку делать _нормально_ сразу, не оставляя на потом
- Постепенно вырабатывает навык проектирования, без прочной завязки на нормальных формах

Шаг 3. Создание логической структуры

- Уже имеем концептуальную схему

- Хотим получить больший уровень детализации:
  - Уточнение атрибутивного состава
  - Уточнение ограничений, накладываемых на атрибутивный состав
  - В некоторых случаях допускается уточнение типа атрибута

Логическая модель данных

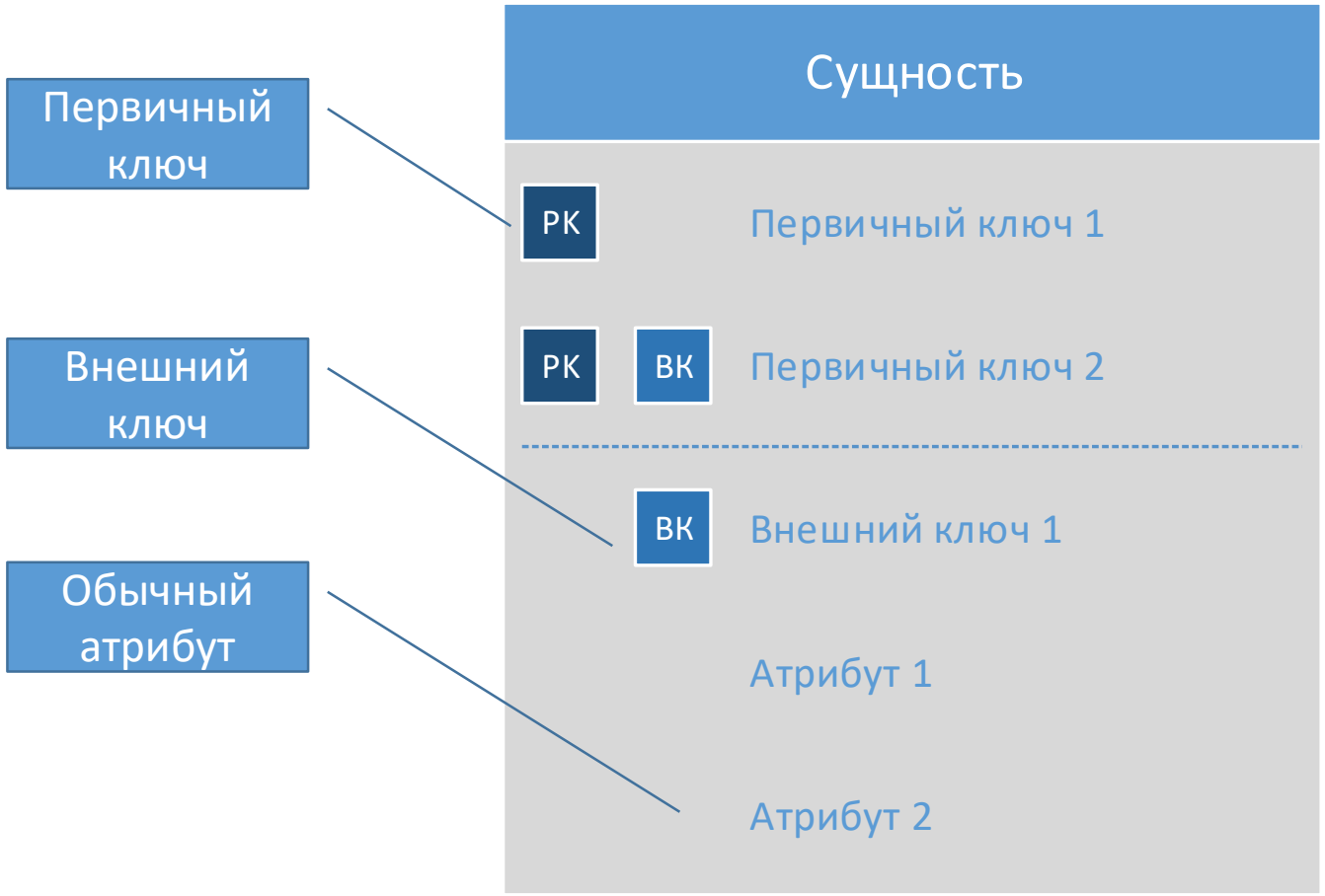
- **Логическая модель данных** – расширение концептуальной модели данных путем определения для сущностей их атрибутов, описаний и ограничений, частично уточняет состав сущностей и взаимосвязи между ними

Логическая модель данных

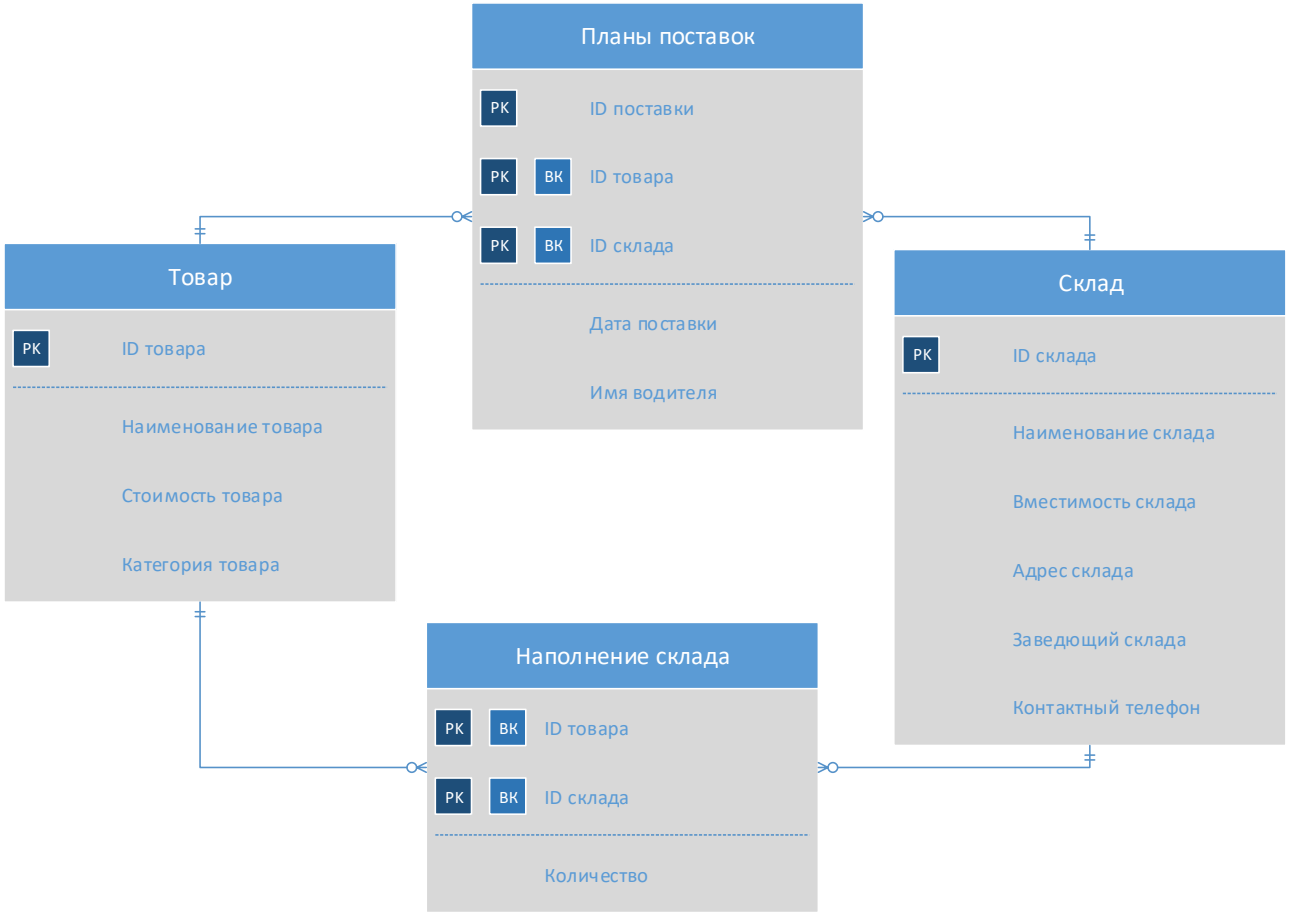
- Разрешает выход за рамки концептуальной модели
- Является прототипом будущей физической модели
- Не учитывает специфику какой-либо конкретной СУБД

- Для иллюстрации также используется ER-нотация

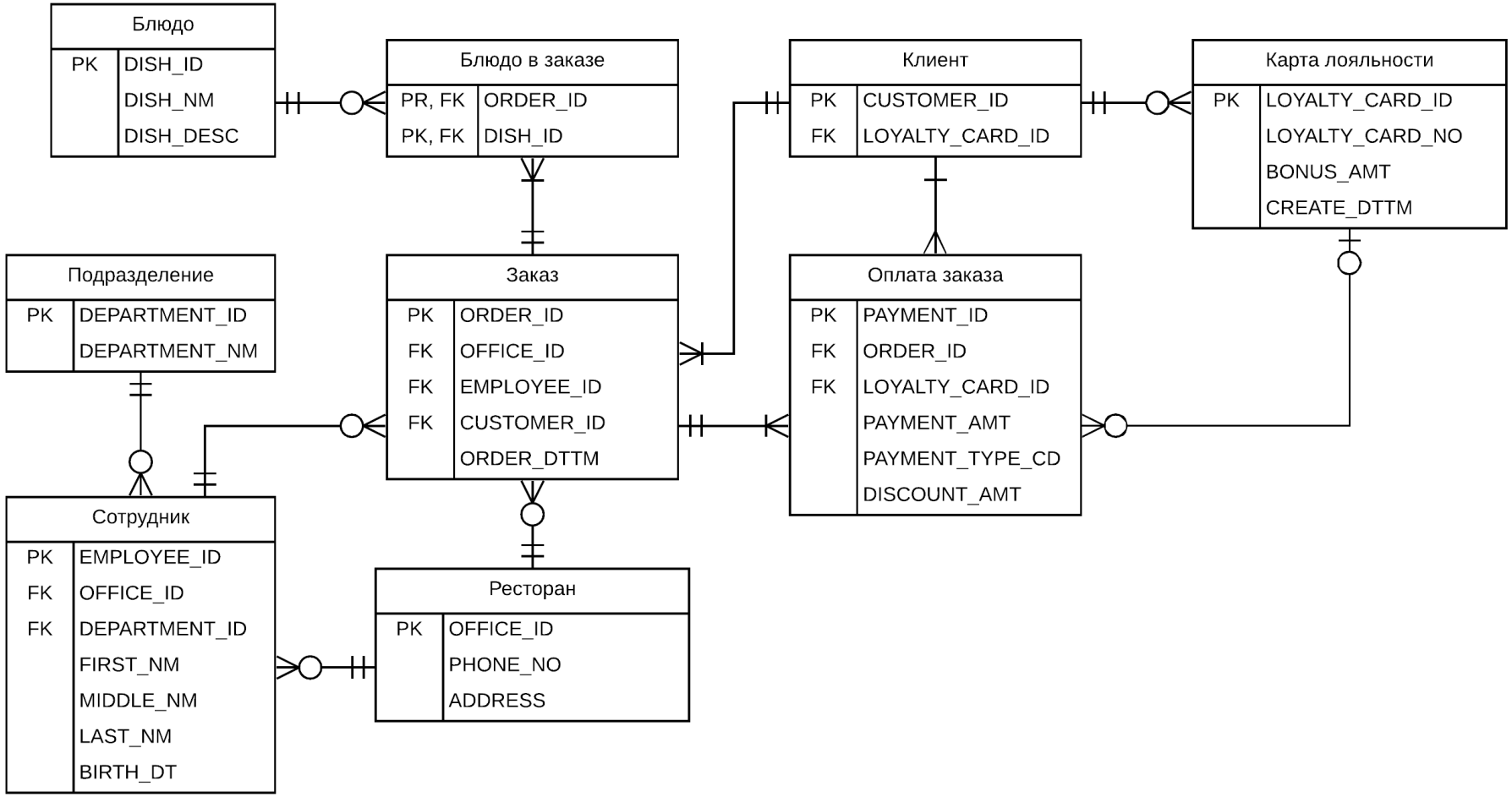
ER-нотация «Сущность»



Примеры ER-диаграмм: «Пятерочка»



Пример логической модели данных



Физическое проектирование

- На следующей лекции