

Министерство образования и науки Украины
Харьковский национальный университет радиоэлектроники

Дисциплина:
«Организация баз данных и знаний»

для студентов дневной формы обучения
по направлению
6.050101 «Компьютерные науки»

Проектирование схемы базы данных

Основные задачи:

- 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 этапы проектирования баз данных

Концептуальное (инфологическое)
проектирование — построение семантической модели предметной области, то есть информационной модели наиболее

высокого уровня абстракции. Такая модель создаётся без ориентации на какую-либо конкретную СУБД и модель данных. Термины «семантическая модель», «концептуальная модель» и «инфологическая модель» являются синонимами. Кроме того, в этом контексте равноправно могут использоваться слова «модель базы данных» и «модель предметной области» (например, «концептуальная модель базы данных» и «концептуальная модель предметной области»), поскольку такая модель является как образом реальности, так и образом проектируемой базы данных для этой реальности.

Конкретный вид и содержание концептуальной модели базы данных определяется выбранным для этого формальным аппаратом. Обычно используются графические нотации, подобные ER-диаграммам.

Чаще всего концептуальная модель базы данных включает в себя:

- описание информационных объектов или понятий предметной области и связей между ними.
- описание ограничений целостности, т.е. требований к допустимым значениям данных и к связям между ними.

Логическое (дatalogическое) проектирование — создание схемы базы данных на основе конкретной модели данных, например, реляционной модели данных. Для реляционной модели данных дatalogическая модель — набор схем отношений с указанием первичных ключей, а также «связей» между отношениями, представляющих собой внешние ключи.

На этапе логического проектирования учитывается специфика конкретной модели данных и специфика конкретной СУБД.

Физическое проектирование — создание схемы базы данных для конкретной СУБД. Специфика конкретной СУБД может включать в себя ограничения на именование объектов базы данных, ограничения на поддерживаемые типы данных и т.п. Кроме того, специфика конкретной СУБД при физическом проектировании включает выбор решений, связанных с физической средой хранения данных (выбор методов управления дисковой памятью, разделение БД по файлам и устройствам, методов доступа к данным), создание индексов и т.д.

**

Для дальнейшего ознакомления с возможностями конкретной СУБД необходимо вспомнить терминологию БД, основы проектирования БД, а также методы организации реляционных баз данных.

Основы проектирования БД

ЦЕЛЬ ПРОЕКТИРОВАНИЯ – создание рациональной схемы БД, абстрагирование структур данных от конечных пользователей.

РЕЗУЛЬТАТ

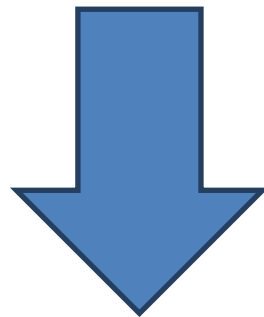


СХЕМА БАЗЫ ДАННЫХ

Схема – представление администратора о логической структуре БД

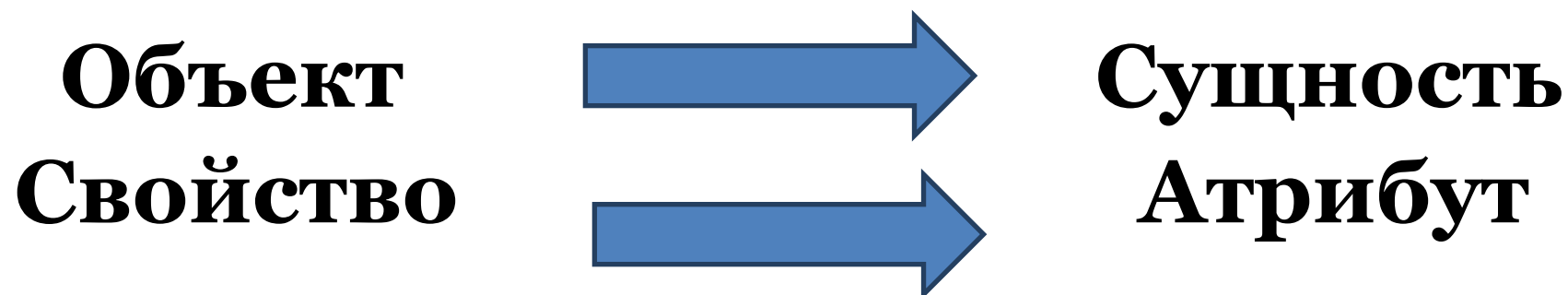
Способы задания схем БД

- графический;
- в терминах языка (SQL, PL/SQL).

Требования к графическому заданию схем:

1. Отображение всех объектов и связей ПрО;
2. Отображение типов информационных связей;
3. Отображение спецификации типов элементов данных.
- 4.

Описание сущностей ПрО

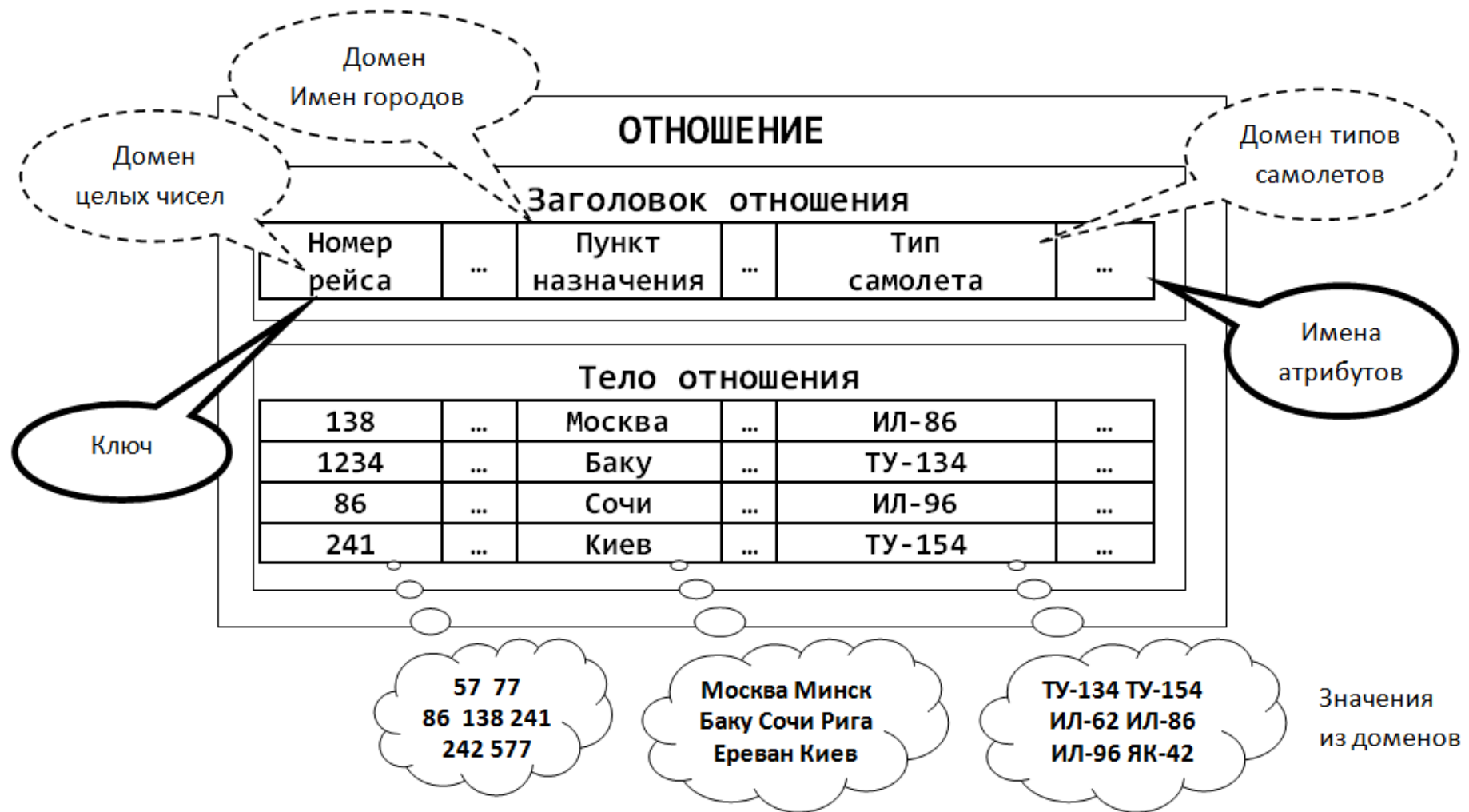


Атрибут – упорядоченная пара $\langle A, V \rangle$, где A – имя атрибута, V – множество его значений.

Домен – множество **допустимых** значений атрибута.

ПОЛ_ЧЕЛОВЕКА
<i>МУЖСКОЙ</i>
<i>ЖЕНСКИЙ</i>

Пример



Сущности БД

Сущность – объект, который может быть идентифицирован определенным способом, который отличает его от других объектов.

Тип сущности – объединение сущностей, которые характеризуются одинаковым набором свойств.

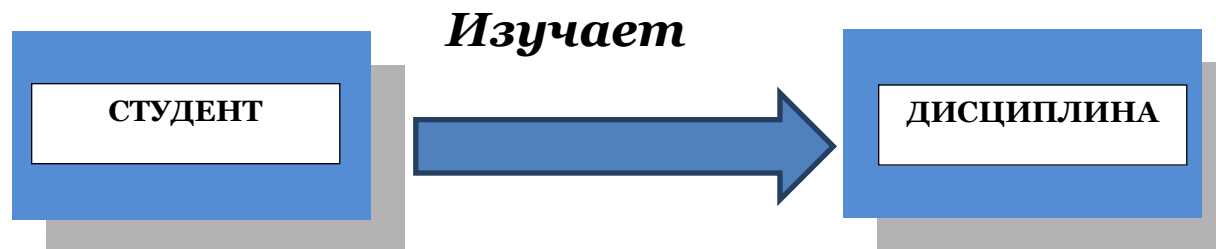
Классификация сущностей



Сильный тип Слабый тип

Связи БД

Связь – ассоциация, установленная между сущностями различных типов.



Типы информационных связей

Информационная связь — свойство предметной области.

1. Один к одному: « $1 \rightarrow 1$ »;
2. Один ко многим: « $1 \rightarrow N$ »;
3. Многие ко многим: « $M \rightarrow N$ ».



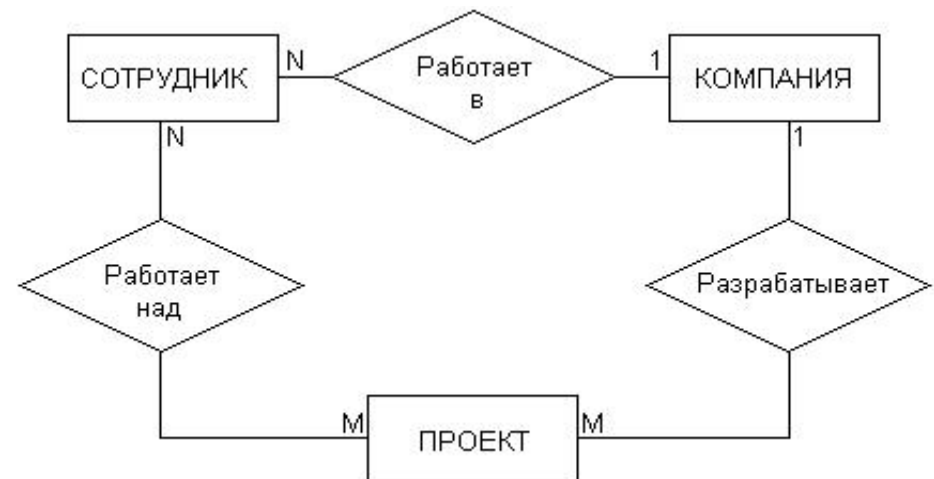
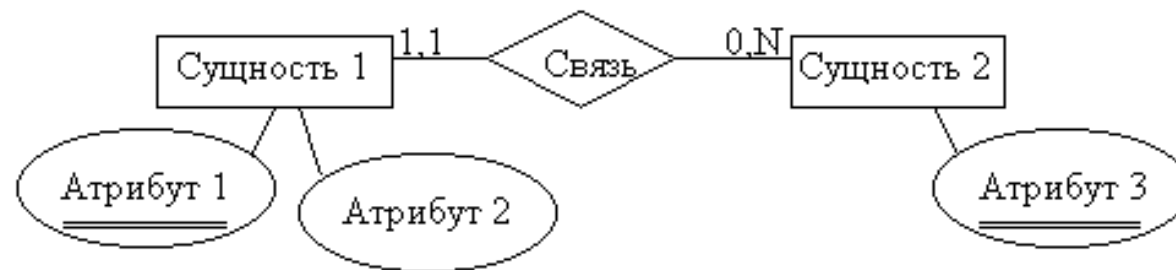
Связи устанавливаются между **СУЩНОСТЯМИ**,
а не их свойствами!



Диаграмма «Сущность-связь»

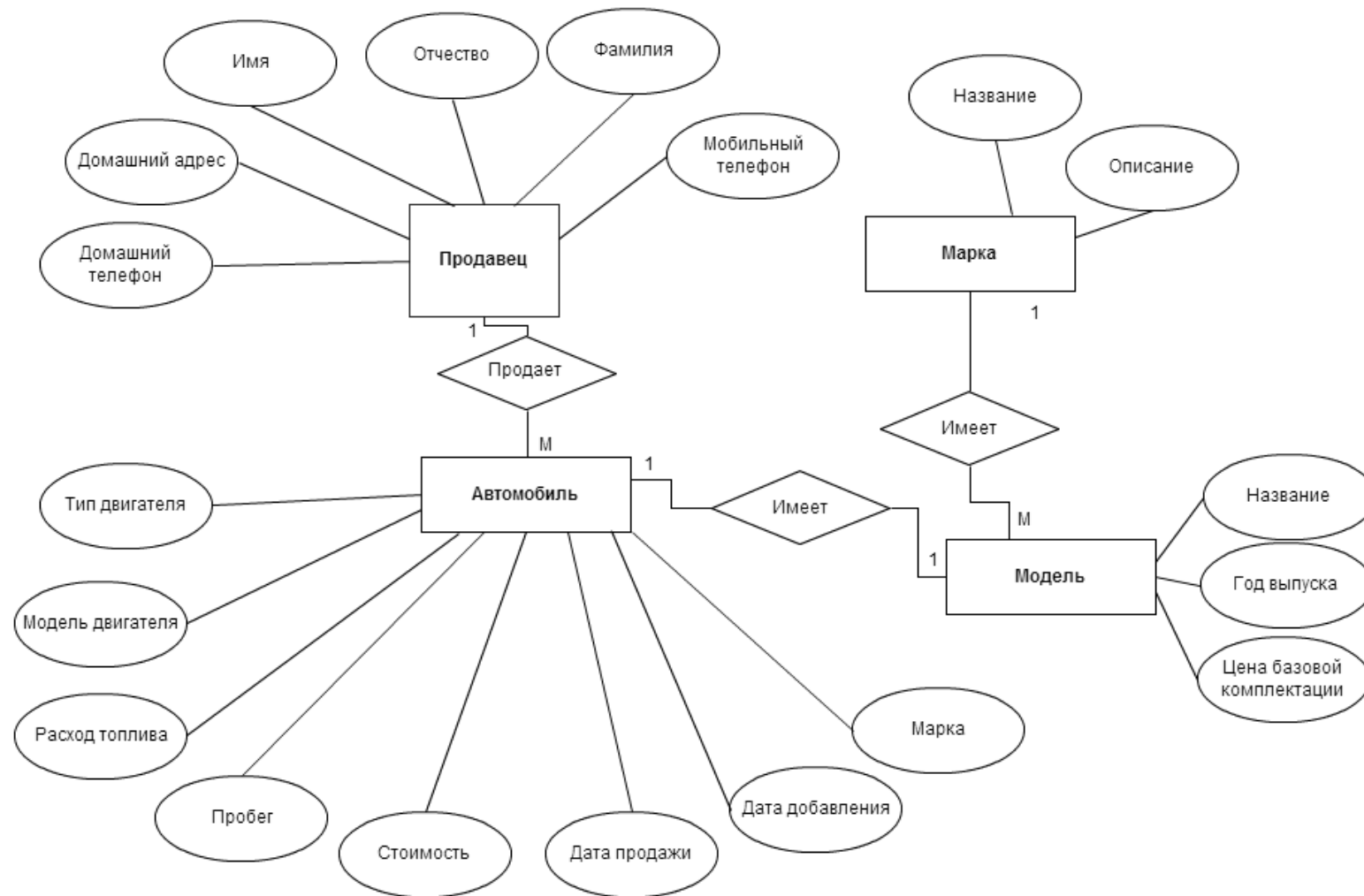
ER-диаграмма (англ. *Entity-relationship*)

Нотация П. Чена



Пример ER-диаграммы в нотации

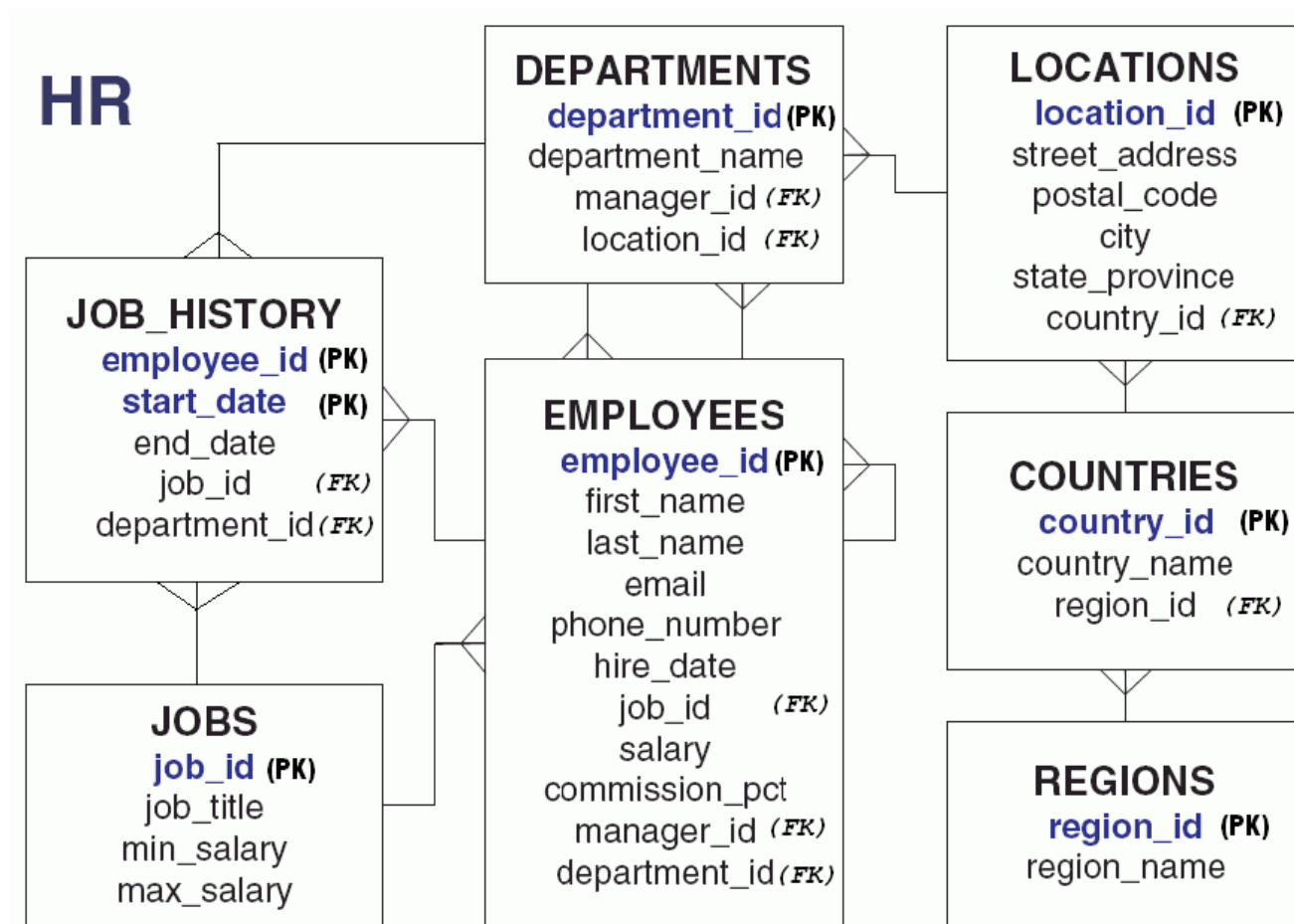
Цена



ER-диаграмма БД

Human Resource (HR)

(тестовая БД СУБД Oracle)



Реляционная модель данных (РМД)

Классическое описание реляционной модели данных содержит три основные функциональные компоненты:

$$РМД = \langle SRM, I, O \rangle$$

SRM - структурная компонента; *I* - ограничения целостности;
O - операционная спецификация.

Структурная компонента реляционной модели данных (*SRM*) – *n*-арное отношение.

$$SRM = \{R, D, A, dom\}$$

где *R* - множество имен отношений;
D - множество доменов;
A - множество имен атрибутов;
dom - отображение из *A* в *D* $dom: A \rightarrow D$

$$R_i = domA_1 \times \dots \times domA_n$$

Сотрудники		
Код	Фамилия	Оклад
2	Петров	2000
3	Сидоров	3000

Ограничения целостности РМД.

Идентификация данных. Ключи

Целостность данных – непротиворечивость.

Идентификация данных – возможность уникального определения каждого отдельного кортежа отношения по значениям его атрибутов.

Реляционная целостность:

1. *целостность сущности* – обеспечивается заданием **первичного** ключа;
2. *ссылочная целостность* – обеспечивается заданием **внешних** ключей .



Ключи отношений БД

Первичный ключ (ПК) – наименьшее подмножество атрибутов, которое уникально идентифицирует кортеж отношения.

<u>И ПРОЕКТА</u>	ПРОЕКТ
100	Космос
200	Климат

Составной ПК – ключ, состоящий из нескольких атрибутов.

<u>СОТРУДНИК</u>	<u>ОТДЕЛ</u>	ДОЛЖНОСТЬ
Иванов	ПФО	Экономист
Иванов	Бухгалтерия	Бухгалтер

Ключи отношений БД (1)

Суррогатный ключ – дополнительный атрибут отношения, значения которого генерируются искусственно.

<u>ID</u>	ФАМИЛИЯ	ИМЯ	ОТЧЕСТВО	ДАТА_РОЖД	ИНН
1	Иванов	Иван	Иванович	12.01.1990	3021506909
2	Иванов	Иван	Иванович	12.01.1990	–

Внешний ключ – атрибут / множество атрибутов одного отношения, значения которых совпадают со значениями первичного ключа другого отношения.

<u>ID ОТДЕЛА</u>	НАИМЕНОВАНИЕ
10	библиотека
20	бухгалтерия
30	канцелярия

<u>ID СОТР</u>	ID_ОТДЕЛА
1	10
2	10
3	30



Реляционная база данных



1. Уникальные имена отношений (таблиц);
2. Уникальные имена атрибутов в таблице;
3. Значения атрибута принадлежат одному домену;
4. Каждый кортеж уникален;
5. Значения атрибутов *атомарны*;
6. Отношения *нормализованы*.



Нормализация отношений БД

Нормализация – процесс дальнейшего усовершенствования схемы БД. Метод создания набора отношений с заданными свойствами на основе требований, которые предъявляются к данным.

Базируется на понятии



Функциональная зависимость (ФЗ): Пусть R – отношение. Множество атрибутов Y функционально зависимо от множества атрибутов X (X функционально определяет Y) тогда и только тогда, когда для любого состояния отношения R для любых кортежей имеющих одинаковые значения атрибутов X , значения атрибутов Y также совпадают в *любом состоянии* отношения R

$$X \rightarrow Y$$

Цель нормализации

Избыточная функциональная зависимость – зависимость, содержащая информацию, которая может быть получена на основе других ФЗ, занесенных в БД.

<u>Н ПРОЕКТА</u>	ПРОЕКТ
100	Космос
200	Климат

<u>ID СОТР</u>	<u>Н ПРОЕКТА</u>	ПРОЕКТ
1	200	Климат
1	100	Космос
2	100	Космос

Цель нормализации:

- устранение избыточных функциональных зависимостей;
- обеспечение целостности данных;
- устранение аномалий.

Аномалии при работе с БД

Аномалии проявляются при выполнении операций, **изменяющих состояние** базы данных.

Различают:

- Аномалии вставки (*INSERT*).
- Аномалии обновления (*UPDATE*).
- Аномалии удаления (*DELETE*).

Н_СОТР	ФАМ	Н_ОТД	ТЕЛ	Н_ПРО	ПРОЕКТ	Н_ЗАДАН
1	Иванов	1	11-22-33	1	Космос	1
1	Иванов	1	11-22-33	2	Климат	1
2	Петров	1	11-22-33	1	Космос	2
3	Сидоров	2	33-22-11	1	Космос	3
3	Сидоров	2	33-22-11	2	Климат	2

ОГРАНИЧЕНИЯ

Н_СОТР,

Н_ОТД,

Н_ПРО



обязательны

1НФ (Первая Нормальная Форма)

1НФ: Отношение находится в первой нормальной форме тогда, и ТОЛЬКО ТОГДА, КОГДА:

- отношение не содержит повторяющихся кортежей;
- кортежи следуют в произвольном порядке;
- атрибуты имеют уникальные имена;
- все значения атрибутов атомарны.

<u>Н СОТР</u>	ФАМ	Н_ОТД	ТЕЛ	<u>Н ПРО</u>	ПРОЕКТ	Н_ЗАДАН
1	Иванов	1	11-22-33	1	Космос	1
1	Иванов	1	11-22-33	2	Климат	1
2	Петров	1	11-22-33	1	Космос	2
3	Сидоров	2	33-22-11	1	Космос	3
3	Сидоров	2	33-22-11	2	Климат	2

Функциональные зависимости в отношении

В отношении «СОТРУДНИКИ_ОТДЕЛЫ_ПРОЕКТЫ»
зависимость атрибутов от потенциального ключа отношения:

{Н СОТР, Н ПРО}:

$\{Н_СОТР, Н_ПРО\} \rightarrow \text{ФАМ}$

$\{Н_СОТР, Н_ПРО\} \rightarrow \text{Н_ОТД}$

$\{Н_СОТР, Н_ПРО\} \rightarrow \text{ТЕЛ}$

$\{Н_СОТР, Н_ПРО\} \rightarrow \text{ПРОЕКТ}$

$\{Н_СОТР, Н_ПРО\} \rightarrow \text{Н_ЗАДАН}$

2НФ (Вторая Нормальная Форма)

2НФ: Отношение находится во второй нормальной форме тогда и только тогда, когда отношение находится в 1НФ и все *неключевые* атрибуты функционально не зависят от части составного ключа.

<u>Н СОТР</u>	ФАМ	Н_ОТД	ТЕЛ	<u>Н ПРО</u>	ПРОЕКТ	Н_ЗАДАН
1	Иванов	1	11-22-33	1	Космос	1
1	Иванов	1	11-22-33	2	Климат	1
2	Петров	1	11-22-33	1	Космос	2
3	Сидоров	2	33-22-11	1	Космос	3
3	Сидоров	2	33-22-11	2	Климат	2

$\text{Н_СОТР} \rightarrow \text{ФАМ}$
 $\text{Н_СОТР} \rightarrow \text{Н_ОТД}$
 $\text{Н_СОТР} \rightarrow \text{ТЕЛ}$

$\text{Н_ПРО} \rightarrow \text{ПРОЕКТ}$

Декомпозиция отношений

<u>Н СОТР</u>	ФАМ	Н_ОТД	ТЕЛ	<u>Н ПРО</u>	ПРОЕКТ	Н_ЗАДАН
1	Иванов	1	11-22-33	1	Космос	1
1	Иванов	1	11-22-33	2	Климат	1
2	Петров	1	11-22-33	1	Космос	2
3	Сидоров	2	33-22-11	1	Космос	3
3	Сидоров	2	33-22-11	2	Климат	2



<u>Н СОТР</u>	ФАМ	Н_ОТД	ТЕЛ
1	Иванов	1	11-22-33
2	Петров	1	11-22-33
3	Сидоров	2	33-22-11



<u>Н ПРО</u>	ПРОЕКТ
1	Космос
2	Климат

3НФ (Третья Нормальная Форма)

3НФ: Отношение находится в третьей нормальной форме тогда, и только тогда, когда отношение находится в 2НФ и все *неключевые* атрибуты взаимно независимы.

<u>Н СОТР</u>	ФАМ	Н_ОТД	ТЕЛ
1	Иванов	1	11-22-33
2	Петров	1	11-22-33
3	Сидоров	2	33-22-11

↓ Н_ОТД → ТЕЛ ↓

Н_СОТР	ФАМ	Н_ОТД
1	Иванов	1
2	Петров	1
3	Сидоров	2

Н_ОТД	ТЕЛ
1	11-22-33
2	33-22-11

Декомпозированные отношения

$\{H_COTR, H_ПРО\} \rightarrow H_ЗАДАН$

H_COTR	H_ПРО	H_ЗАДАН
1	1	1
1	2	1
2	1	2
3	1	3
3	2	2

H_ПРО	ПРОЕКТ
1	Космос
2	Климат

$H_ПРО \rightarrow ПРОЕКТ$

$H_COTR \rightarrow ФАМ$

$H_COTR \rightarrow H_ОТД$

$H_COTR \rightarrow ТЕЛ$

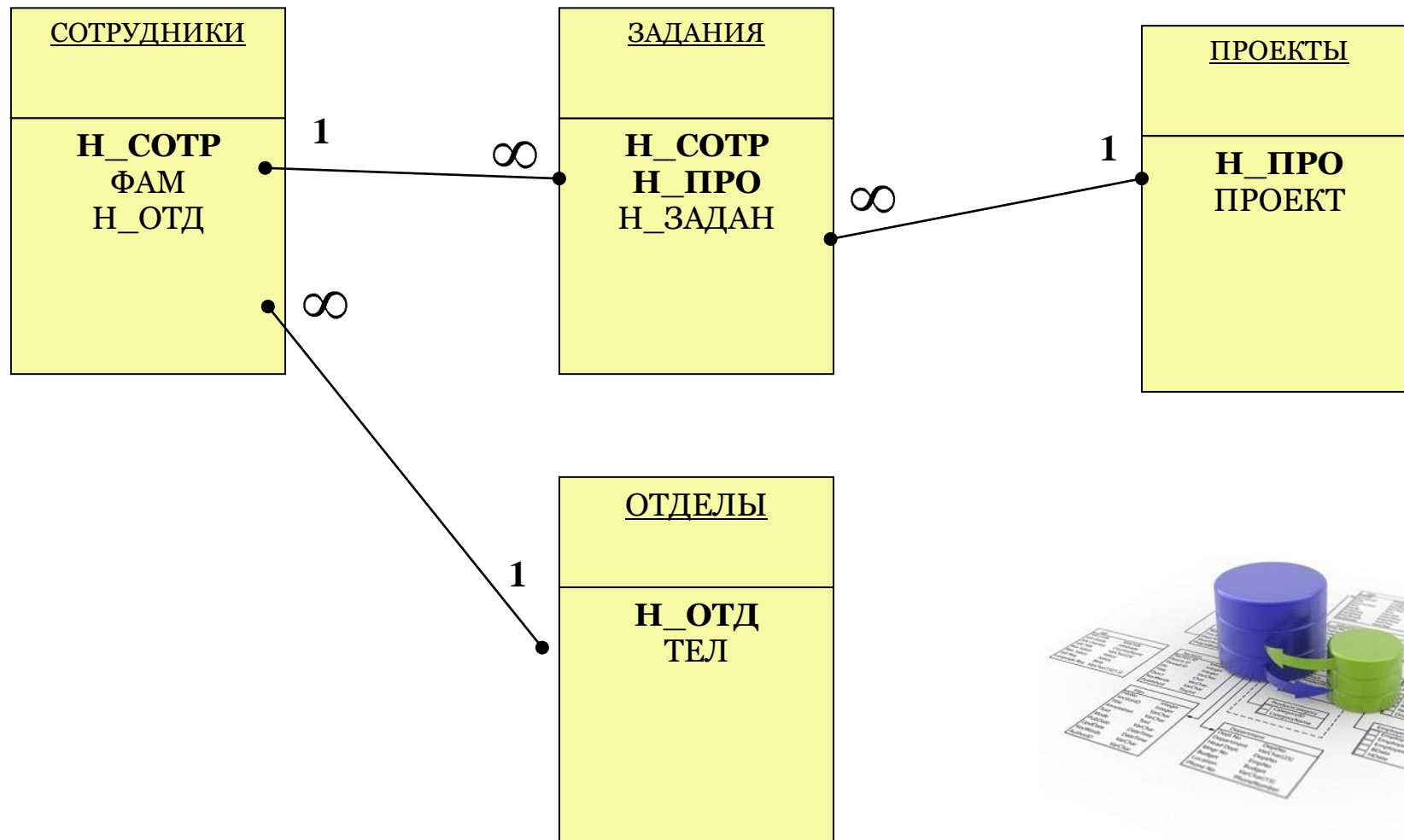
H_COTR	ФАМ	H_ОТД	ТЕЛ
1	Иванов	1	11-22-33
2	Петров	1	11-22-33
3	Сидоров	2	33-22-11

$H_ОТД \rightarrow ТЕЛ$

H_COTR	ФАМ	H_ОТД
1	Иванов	1
2	Петров	1
3	Сидоров	2

H_ОТД	ТЕЛ
1	11-22-33
2	33-22-11

Реляционная схема БД



Рекомендованная литература

1. Дейт, К. Дж. Введение в системы баз данных: пер. с англ. / К. Дж. Дейт. – М.: Вильямс, 2005. – 1328 с.
2. Конноли, Т. М. Базы данных. Проектирование, реализация, сопровождение. Теория и практика : пер. с англ. / Т. М. Конноли, К. Бегг: 3-е изд. – Москва-Санкт-Петербург-Киев, 2003. – 1436 с.
3. Тиори Т., Фрай Дж. Проектирование структур баз данных, Москва, 1985. – 275 с.
4. Кевин Луни, Oracle Database 10g. Полный справочник, 1,2 том. – Изд. «Лори», 2006.
5. Прайс Дж. Oracle 10g SQL. Операторы SQL и программы PL/SQL. Изд. «Лори», 2007. – 565 с.
6. Кайт Т. Oracle для профессионалов. – Изд «DiaSoft», 2003. – 1831 с.
7. Кайт Т., Oracle для профессионалов: архитектура, методики программирования и особенности версий 9i, 10g и 11g., 2 издание – Изд. «Вильямс», 2013. – 848 с.
8. Мишра С., Бьюли А. Секреты Oracle SQL. – Изд. «Символ», 2003. – 366 с.
9. Аллен К. Как писать мощные и гибкие программы на PL/SQL. – Изд. «Лори», 2001.
10. Урман С. Программирование на языке PL/SQL – Изд. «Лори», 2001.