



ПРОЕКТИРОВАНИЕ И НОРМАЛИЗАЦИЯ БАЗ ДАННЫХ

Норма – это то, что встречается лишь изредка.

С. Моэм

Содержание

2

- Понятие проектирования базы данных
- Понятие функциональной зависимости
- Нормальные формы
- Упражнения по проектированию баз данных

Проектирование базы данных

3

Внешний уровень



Пользователи

- ER-диаграммы, диаграммы классов предметной области

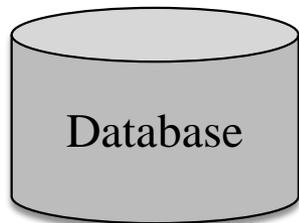
Логический уровень



АБД

- Схема базы данных

Физический уровень



- Настройка СУБД

Функциональная зависимость

4

- Пусть R – переменная отношения, X, Y – атрибуты R . Тогда
 - ▣ Y *функционально зависит* от X , или
 - ▣ X *функционально определяет* Y , или
 - ▣ X является *детерминантом*, а Y является *зависимой* частью
 - ▣ $X \rightarrow Y$

если не существует двух различных кортежей, которые в атрибуте X имеют совпадающие значения, а в атрибуте Y имеют разные значения.

Функциональная зависимость

5

| A | B | C | D | E |
|-----|-----|-----|-----|-----|
| a1 | b1 | c1 | d1 | e1 |
| a2 | b2 | c2 | d2 | e2 |
| a1 | b1 | c1 | d1 | e3 |
| a2 | b2 | c2 | d2 | e4 |
| ... | ... | ... | ... | ... |

$\{A, B\} \rightarrow \{C, D\}$

□ ФЗ подразумевает *все возможные кортежи отношения*, а не только те, которые находятся в отношении в настоящий момент.

□ ФЗ *определяется только семантикой предметной области* и не может быть выведена из текущего состояния базы данных.

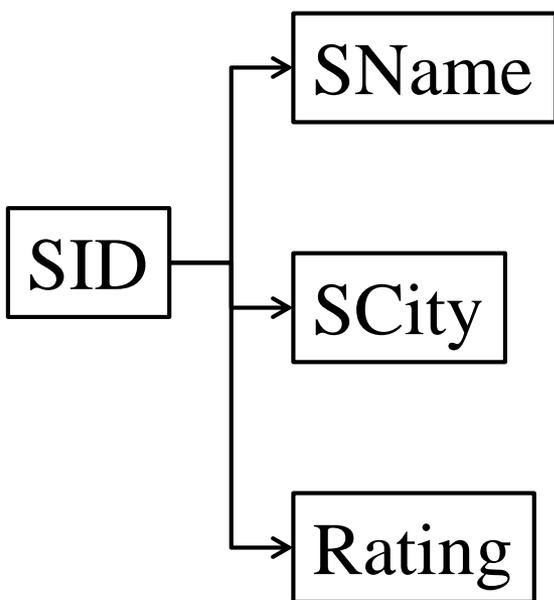
| Код_П | Имя_П | Город_П | Рейтинг |
|-------|--------------|------------|---------|
| S1 | Бендер | Черноморск | 15 |
| S2 | ИТ-монстр | Москва | 29 |
| S3 | Разноэкспорт | Черноморск | 15 |
| S4 | Макрохард | Москва | 29 |
| ... | ... | ... | ... |

$\{\text{Город}\} \rightarrow \{\text{Рейтинг}\}$

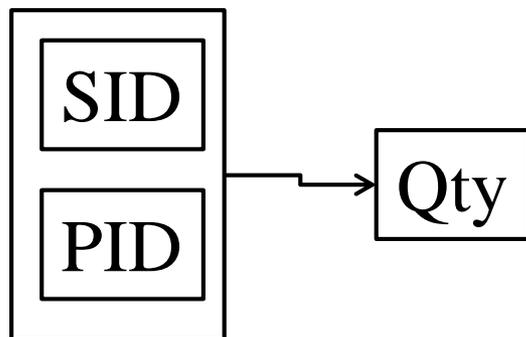
Диаграммы ФЗ

6

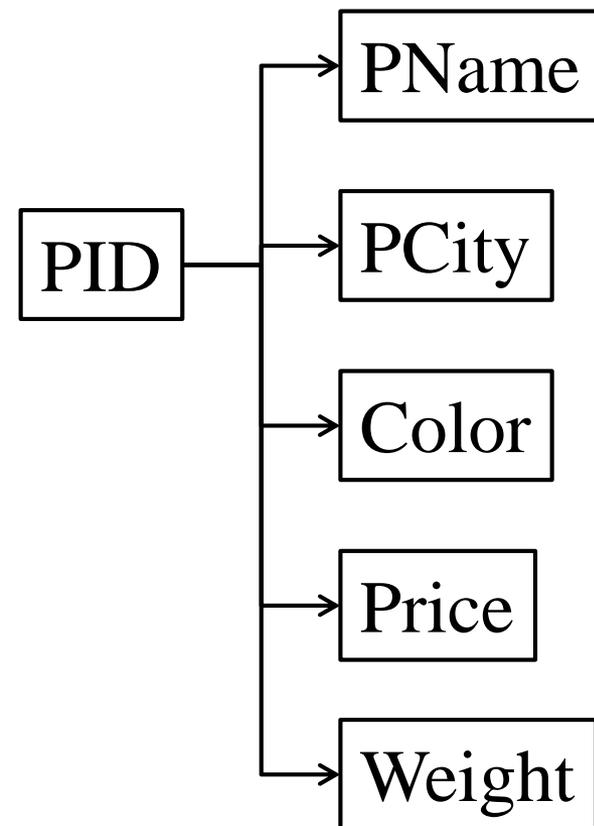
ФЗ отношения S



ФЗ отношения SP



ФЗ отношения P



Количество ФЗ

7

- ФЗ – ограничение целостности, за соблюдением которого должна следить СУБД.
- Меньше ФЗ – легче работа СУБД.
- Как сократить множество ФЗ? (Как получить эквивалентное множество ФЗ меньшей мощности?)

Теорема Хита (Heath)

9

- Пусть R – отношение с заголовком A, B, C (атрибуты, возможно, составные) и выполняется ФЗ $A \rightarrow B$. Тогда $R = \pi_{A,B}(R) \bowtie \pi_{A,C}(R)$.
- Теорема Хита дает возможность выполнить декомпозицию отношений без потерь информации:
 - исходное отношение с ФЗ преобразуется в другие отношения, в каждом из которых атрибуты минимально зависят от первичного ключа.
 - атрибут B минимально зависит от атрибута A , если выполняется минимальная слева ФЗ (т.е. ФЗ с минимальным детерминантом) $A \rightarrow B$.

Теорема Хита (Heath)

10

- Пусть R – отношение с заголовком A, B, C (атрибуты, возможно, составные) и выполняется ФЗ $A \rightarrow B$. Тогда $R = \pi_{A,B}(R) \bowtie \pi_{A,C}(R)$.
- Доказательство
 1. Пусть кортеж $\langle a, b, c \rangle \in R$. Тогда по определению операции проекции $\langle a, b \rangle \in \pi_{A,B}(R)$ и $\langle a, c \rangle \in \pi_{A,C}(R)$. Т.е. $\langle a, b, c \rangle \in (\pi_{A,B}(R) \bowtie \pi_{A,C}(R))$.
 2. Если $\langle a, b, c \rangle \in (\pi_{A,B}(R) \bowtie \pi_{A,C}(R))$, то $\exists \langle a, b \rangle \in \pi_{A,B}(R)$ и $\langle a, c \rangle \in \pi_{A,C}(R)$, что выполнимо $\Leftrightarrow \exists \langle a, b^*, c \rangle \in R$. Т.к. $A \rightarrow B$, то $b = b^*$, т.е. $\langle a, b, c \rangle = \langle a, b^*, c \rangle$ и $\langle a, b, c \rangle \in R$.

Зачем нужна нормализация

11

BAD_S_P_SP

| Код_П | Имя_П | Город_П | Рейтинг | Код_Д | Имя_Д | Город_Д | Вес | Цена | Количество |
|-------|-----------|------------|---------|-------|--------|-----------|-----|------|------------|
| S1 | Бендер | Черноморск | 15 | P1 | Болт | Париж | 15 | 40 | 1000 |
| S34 | ИТ-монстр | Москва | 29 | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| S1 | Бендер | Черноморск | 15 | P2 | Гайка | Челябинск | 20 | 24 | 1240 |
| S34 | ИТ-монстр | Москва | 29 | P4 | Гвоздь | Москва | 10 | 42 | 2700 |

□ *Избыточность* данных

- одно и то же значение атрибутов Имя_П, Город_П, Рейтинг хранится для каждой поставки данного поставщика

□ *Аномалии* работы с данными

■ Аномалии обновления:

- обновление атрибута Рейтинг в одной поставке требует обновления этого атрибута во всех поставках данного поставщика.

■ Аномалии добавления:

- не может быть добавлен поставщик, не поставляющий в данный момент деталей.

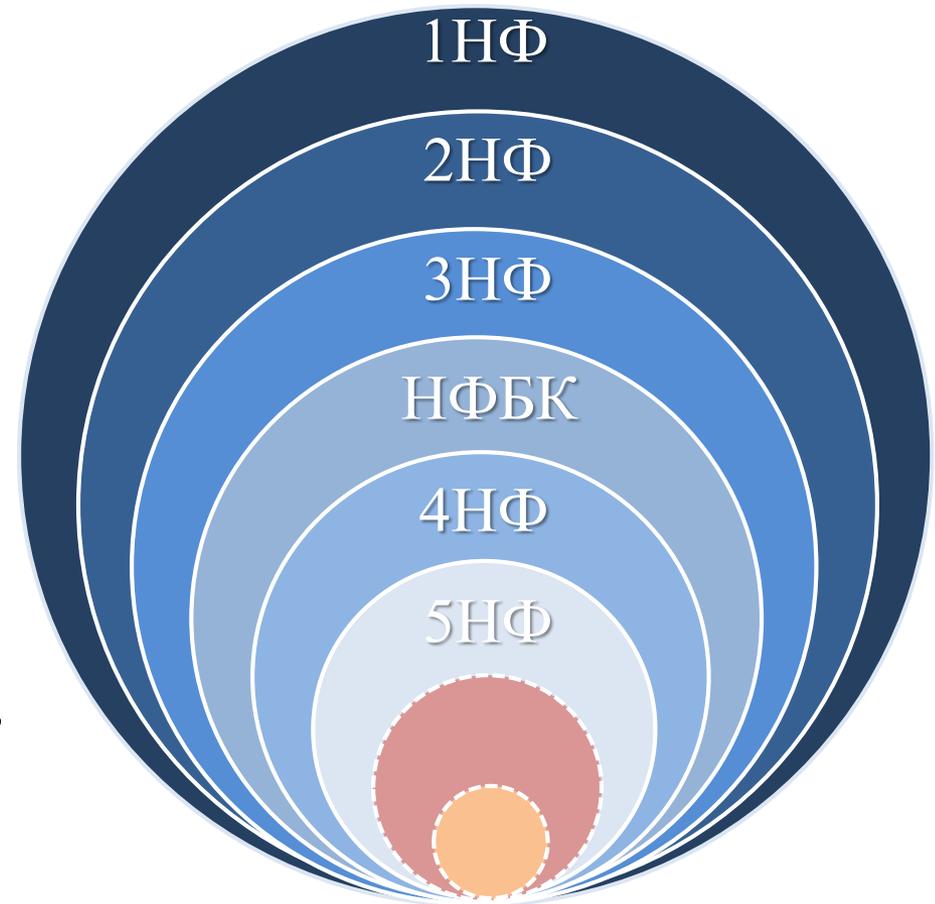
■ Аномалии удаления:

- не может существовать поставщик, не поставляющий в данный момент деталей.

Нормальные формы

12

- *Нормальная форма* – совокупность ограничений, накладываемых на схему отношения. Ограничения исключают аномалии при работе с базой данных.
- *Нормализация* состоит в получении из исходного набора отношений *нормализованных отношений*, каждое из которых находится в НФ как можно более высокого порядка.



1НФ

13

- Любое реляционное отношение находится в 1НФ (по определению реляционного отношения).

NON_1NF_S_P_SP

Приведение к 1НФ

| Код_П | Имя_П | Город_П | Рейтинг | Поставка | | | | | |
|-------|-----------|------------|---------|----------|--------|---------|-----|------|--------|
| S1 | Бендер | Черноморск | 15 | Код_Д | Имя_Д | Город_Д | Вес | Цена | Колич. |
| | | | | P1 | Болт | Париж | 15 | 40 | 1000 |
| | | | | P5 | Шуруп | Москва | 29 | 33 | 1240 |
| S3 | ИТ-монстр | Москва | 29 | Код_Д | Имя_Д | Город_Д | Вес | Цена | Колич. |
| | | | | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| | | | | P4 | Гвоздь | Москва | 10 | 42 | 2700 |

BAD_S_P_SP

| Код_П | Имя_П | Город_П | Рейтинг | Код_Д | Имя_Д | Город_Д | Вес | Цена | Колич. |
|-------|-----------|------------|---------|-------|--------|-----------|-----|------|--------|
| S1 | Бендер | Черноморск | 15 | P1 | Болт | Париж | 15 | 40 | 1000 |
| S3 | ИТ-монстр | Москва | 29 | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| S1 | Бендер | Черноморск | 15 | P2 | Гайка | Челябинск | 20 | 24 | 1240 |
| S3 | ИТ-монстр | Москва | 29 | P4 | Гвоздь | Москва | 10 | 42 | 2700 |

- Отношение *находится* в 2НФ тогда и только тогда, когда
 - ▣ оно находится в 1НФ
 - ▣ каждый неключевой атрибут (не входящий в потенциальный ключ) минимально функционально зависит от первичного ключа.
- Отношение *не находится* в 2НФ, если в отношении существует неключевой атрибут (не входящий в потенциальный ключ), функционально зависящий от собственной части первичного ключа.

Приведение к 2НФ

15

1. Применим теорему Хита к $BAD_S_P_SP$:

▣ Т.к. $Код_П \rightarrow \{Имя_П, Город_П, Рейтинг\}$, то $BAD_S_P_SP =$

$\pi_{Код_П, Имя_П, Город_П, Рейтинг} \bowtie \pi_{Код_П, Код_Д, Имя_Д, Город_Д, Цвет, Вес, Цена, Количество}$

▣ Т.е. $BAD_S_P_SP = S \bowtie BAD_P_SP$.

2. Применим теорему Хита к BAD_P_SP :

▣ Т.к. $Код_Д \rightarrow \{Имя_Д, Город_Д, Цвет, Вес, Цена\}$, то $BAD_P_SP =$

$\pi_{Код_Д, Имя_Д, Город_Д, Цвет, Вес, Цена} \bowtie \pi_{Код_П, Код_Д, Количество}$

▣ Т.е. $BAD_P_SP = P \bowtie SP$.

3. В итоге $BAD_S_P_SP = S \bowtie SP \bowtie P$.

Приведение к 2НФ

16

BAD_S_P_SP

| <u>Код П</u> | <u>Имя П</u> | <u>Город П</u> | <u>Рейтинг</u> | <u>Код Д</u> | <u>Имя Д</u> | <u>Город Д</u> | <u>Вес</u> | <u>Цена</u> | <u>Колич.</u> |
|--------------|--------------|----------------|----------------|--------------|--------------|----------------|------------|-------------|---------------|
| S1 | Бендер | Черноморск | 15 | P1 | Болт | Париж | 15 | 40 | 1000 |
| S34 | ИТ-монстр | Москва | 29 | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| S1 | Бендер | Черноморск | 15 | P2 | Гайка | Челябинск | 20 | 24 | 1240 |
| S34 | ИТ-монстр | Москва | 29 | P4 | Гвоздь | Москва | 10 | 42 | 2700 |



S

BAD_P_SP

| <u>Код П</u> | <u>Имя</u> | <u>Город</u> | <u>Рейтинг</u> |
|--------------|------------|--------------|----------------|
| S1 | Бендер | Черноморск | 15 |
| S34 | ИТ-монстр | Москва | 29 |

| <u>Код П</u> | <u>Код Д</u> | <u>Имя Д</u> | <u>Город Д</u> | <u>Вес</u> | <u>Цена</u> | <u>Колич.</u> |
|--------------|--------------|--------------|----------------|------------|-------------|---------------|
| S1 | P1 | Болт | Париж | 15 | 40 | 1000 |
| S34 | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| S1 | P2 | Гайка | Челябинск | 20 | 24 | 1240 |
| S34 | P4 | Гвоздь | Москва | 10 | 42 | 2700 |

Приведение к 2НФ

17

BAD_P_SP

| <u>Код_П</u> | <u>Код_Д</u> | <u>Имя_Д</u> | <u>Город_Д</u> | <u>Вес</u> | <u>Цена</u> | <u>Колич.</u> |
|--------------|--------------|--------------|----------------|------------|-------------|---------------|
| S1 | P1 | Болт | Париж | 15 | 40 | 1000 |
| S3 | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| S1 | P2 | Гайка | Челябинск | 20 | 24 | 1240 |
| S3 | P4 | Гвоздь | Москва | 10 | 42 | 2700 |



P

SP

| <u>Код_Д</u> | <u>Имя_Д</u> | <u>Город_Д</u> | <u>Вес</u> | <u>Цена</u> |
|--------------|--------------|----------------|------------|-------------|
| P1 | Болт | Париж | 15 | 40 |
| P5 | Шуруп | Москва | 29 | 33 |
| P2 | Гайка | Челябинск | 20 | 24 |
| P4 | Гвоздь | Москва | 10 | 42 |

| <u>Код_П</u> | <u>Код_Д</u> | <u>Колич.</u> |
|--------------|--------------|---------------|
| S1 | P1 | 1000 |
| S3 | P5 | 7600 |
| S1 | P2 | 1240 |
| S3 | P4 | 2700 |

Приведение к 2НФ

18

BAD_S_P_SP

| <u>Код П</u> | <u>Имя П</u> | <u>Город П</u> | <u>Рейтинг</u> | <u>Код Д</u> | <u>Имя Д</u> | <u>Город Д</u> | <u>Вес</u> | <u>Цена</u> | <u>Колич.</u> |
|--------------|--------------|----------------|----------------|--------------|--------------|----------------|------------|-------------|---------------|
| S1 | Бендер | Черноморск | 15 | P1 | Болт | Париж | 15 | 40 | 1000 |
| S3 | ИТ-монстр | Москва | 29 | P5 | Шуруп | Москва | 29 | 33 | 7600 |
| S1 | Бендер | Черноморск | 15 | P2 | Гайка | Челябинск | 20 | 24 | 1240 |
| S3 | ИТ-монстр | Москва | 29 | P4 | Гвоздь | Москва | 10 | 42 | 2700 |



S

P

SP

| <u>Код П</u> | <u>Имя</u> | <u>Город</u> | <u>Рейтинг</u> |
|--------------|------------|--------------|----------------|
| S1 | Бендер | Черноморск | 15 |
| S3 | ИТ-монстр | Москва | 29 |

| <u>Код Д</u> | <u>Имя Д</u> | <u>Город Д</u> | <u>Вес</u> | <u>Цена</u> |
|--------------|--------------|----------------|------------|-------------|
| P1 | Болт | Париж | 15 | 40 |
| P5 | Шуруп | Москва | 29 | 33 |
| P2 | Гайка | Челябинск | 20 | 24 |
| P4 | Гвоздь | Москва | 10 | 42 |

| <u>Код П</u> | <u>Код Д</u> | <u>Колич.</u> |
|--------------|--------------|---------------|
| S1 | P1 | 1000 |
| S3 | P5 | 7600 |
| S1 | P2 | 1240 |
| S3 | P4 | 2700 |

- Отношение *находится в 3НФ* тогда и только тогда, когда
 - ▣ оно находится в 2НФ и
 - ▣ каждый неключевой атрибут нетранзитивно зависит от первичного ключа.
- Отношение *не находится в 3НФ*, если в отношении существуют неключевые атрибуты, транзитивно зависящие от ключа.

Приведение к 3НФ

20

S

| <u>Код_П</u> | Имя_П | Город_П | Рейтинг |
|--------------|--------------|------------|---------|
| S1 | Бендер | Черноморск | 15 |
| S2 | ИТ-монстр | Москва | 29 |
| S3 | Разноэкспорт | Черноморск | 15 |
| S4 | Макрохард | Москва | 29 |



S'

| <u>Код_П</u> | Имя_П | Город_П |
|--------------|--------------|------------|
| S1 | Бендер | Черноморск |
| S2 | ИТ-монстр | Москва |
| S3 | Разноэкспорт | Черноморск |
| S4 | Макрохард | Москва |

CityRating

| <u>Город_П</u> | Рейтинг |
|----------------|---------|
| Черноморск | 15 |
| Москва | 29 |

- Применение теоремы Хита:
Город → Рейтинг

$$S = \pi_{\text{Код_П, Имя_П, Город_П}} \bowtie \pi_{\text{Город_П, Рейтинг}}$$

- Отношение *находится в НФБК* тогда и только тогда, когда
 - ▣ оно находится в 3НФ и
 - ▣ любая выполняемая для этого отношения нетривиальная и минимальная ФЗ имеет в качестве детерминанта потенциальный ключ.
- Отношение *не находится в нормальной форме Бойса-Кодда*, если существует нетривиальная функциональная зависимость некоторого атрибута, входящего в первичный ключ, от атрибутов, не являющихся потенциальным ключом.

Приведение к НФБК

22

Address

| Город | Адрес | Индекс |
|------------|---------------------------------|--------|
| Челябинск | пр. Ленина, д. 76, к. 700 | 454080 |
| Челябинск | пр. Ленина, д. 81, к. 336 | 454080 |
| Челябинск | ул. Российская, д. 52, к. 12 | 454006 |
| Челябинск | ул. Российская, д. 52, к. 43 | 454006 |
| Москва | ул. Тверская-Ямская, д. 3, к. 8 | 128000 |
| Черноморск | ул. Деточкина, д. 1, к. 13 | 123456 |

PostCode

| Город | Индекс |
|------------|--------|
| Челябинск | 454080 |
| Челябинск | 454006 |
| Москва | 128000 |
| Черноморск | 123456 |



Address'

| Адрес | Индекс |
|---------------------------------|--------|
| пр. Ленина, д. 76, к. 700 | 454080 |
| пр. Ленина, д. 81, к. 336 | 454080 |
| ул. Российская, д. 52, к. 12 | 454006 |
| ул. Российская, д. 52, к. 43 | 454006 |
| ул. Тверская-Ямская, д. 3, к. 8 | 128000 |
| ул. Деточкина, д. 1, к. 13 | 123456 |

- Потенциальные ключи:
 - { Город, Адрес }
 - { Индекс, Адрес }
- ФЗ: Индекс → Город

Повышение порядка НФ – не самоцель

23

S

| Код_П | Имя_П | Город_П | Рейтинг |
|-------|--------------|------------|---------|
| S1 | Бендер | Черноморск | 15 |
| S2 | ИТ-монстр | Москва | 29 |
| S3 | Разноэкспорт | Черноморск | 15 |
| S4 | Макрохард | Москва | 29 |

S

| Код_П | Имя_П | Город_П |
|-------|--------------|------------|
| S1 | Бендер | Черноморск |
| S2 | ИТ-монстр | Москва |
| S3 | Разноэкспорт | Черноморск |
| S4 | Макрохард | Москва |

CityRating

| Город_П | Рейтинг |
|------------|---------|
| Черноморск | 15 |
| Москва | 29 |

- Запросы на выборку города и рейтинга вместе с другими данными поставщиков часты.
- Запросы на обновление рейтинга городов редки, не требуется хранение рейтинга городов, в которых нет поставщиков.

Повышение порядка НФ – не самоцель

24

Address

| Город | Адрес | Индекс |
|-----------|---------------------------------|--------|
| Челябинск | пр. Ленина, д. 76, к. 700 | 454080 |
| Челябинск | пр. Ленина, д. 81, к. 336 | 454080 |
| Челябинск | ул. Российская, д. 52, к. 12 | 454006 |
| Челябинск | ул. Российская, д. 52, к. 43 | 454006 |
| Москва | ул. Тверская-Ямская, д. 3, к. 8 | 128000 |

Address

| Адрес | Индекс |
|---------------------------------|--------|
| пр. Ленина, д. 76, к. 700 | 454080 |
| пр. Ленина, д. 81, к. 336 | 454080 |
| ул. Российская, д. 52, к. 12 | 454006 |
| ул. Российская, д. 52, к. 43 | 454006 |
| ул. Тверская-Ямская, д. 3, к. 8 | 128000 |

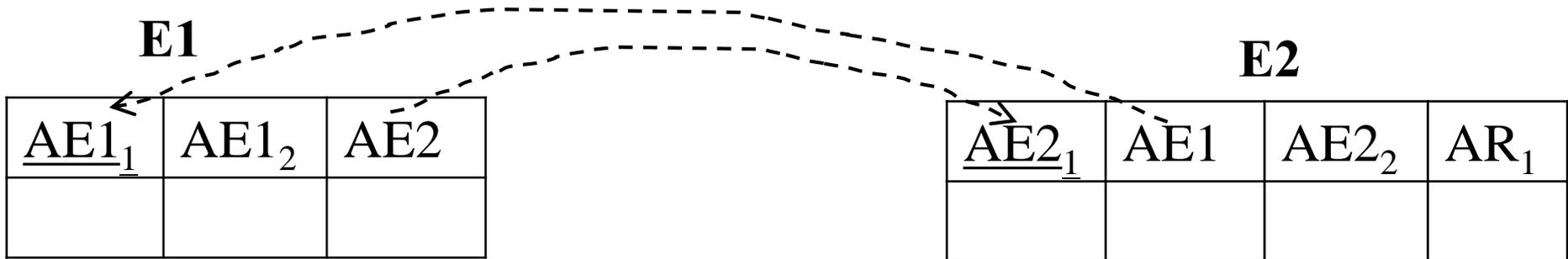
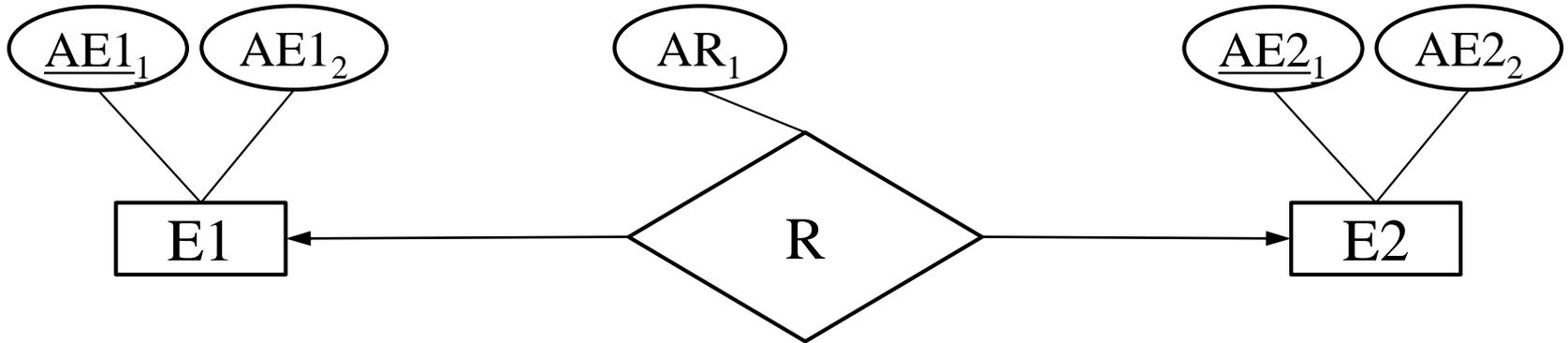
PostCode

| Город | Индекс |
|-----------|--------|
| Челябинск | 454080 |
| Челябинск | 454006 |
| Москва | 128000 |

- Запросы на выборку города и индекса вместе с адресом часты.
- Запросы на обновление индекса городов редки, не требуется хранение индекса городов, в которых нет адресатов.

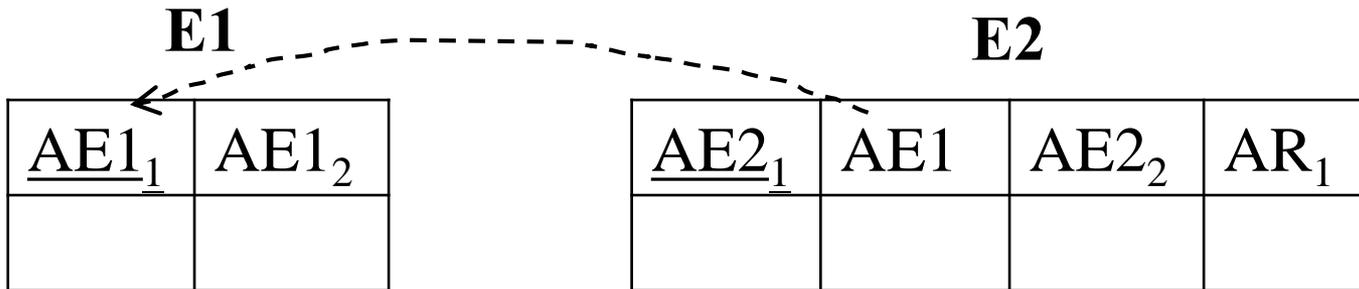
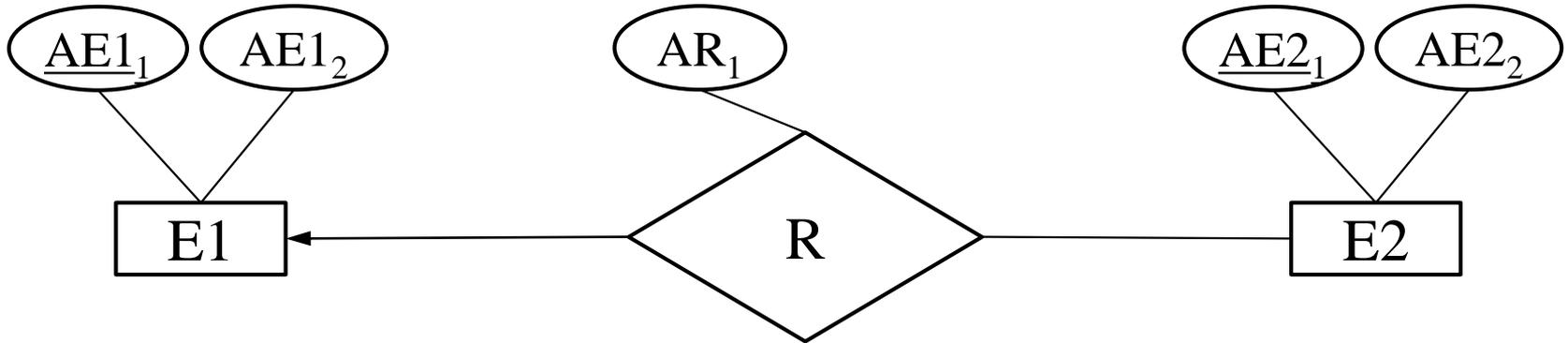
Нормализация и E/R-модель

25



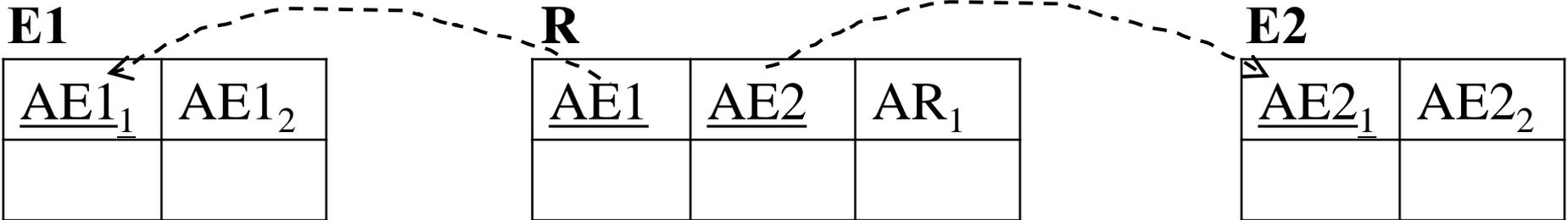
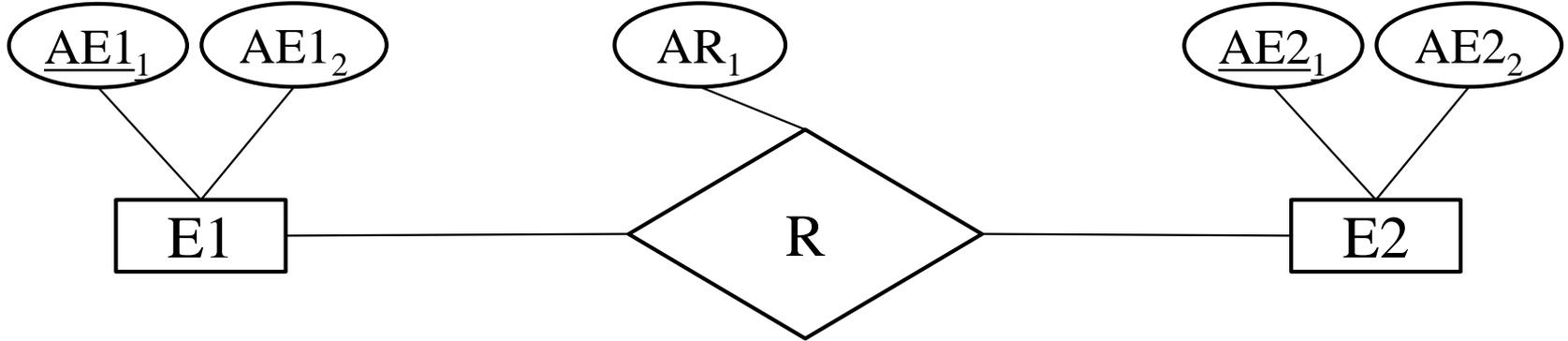
Нормализация и E/R-модель

26



Нормализация и E/R-модель

27



Задача

28

□ Сущности

- ▣ *Шахматисты играют партии в рамках турниров, проводимых организаторами.*
- ▣ Шахматист – ФИО, пол, возраст.
- ▣ Партия – игравший белыми, игравший черными, результат игры.
- ▣ Турнир – название, сроки.
- ▣ Организатор – название, адрес.

□ Связи

- ▣ В турнире участвуют несколько шахматистов. Шахматист может участвовать в нескольких турнирах.
- ▣ У турнира может быть много организаторов. Организатор может организовать много турниров.

Задача

29

□ Сущности

- Футболисты играют *клубах*, участвуя в *матчах чемпионатов*.
- У клуба есть *президент*, *фан-клубы* и *спонсоры*.
- Спонсоры делают *пожертвования* в пользу клуба.
- Фан-клуб состоит из *фанатов*.
- Во время чемпионата клубы могут производить *трансфер* футболистов в другой клуб.

□ Связи

- В клубе играет несколько футболистов. Футболист играет в точности за один клуб.
- В чемпионате принимает участие несколько команд. Команда может принимать участие в нескольких чемпионатах.
- У клуба может быть несколько фан-клубов. Фан-клуб болеет в точности за один футбольный клуб.
- Фан-клуб включает в себя несколько фанатов. Фанат может состоять в нескольких фан-клубах.
- У клуба может быть несколько спонсоров. Спонсор может делать пожертвования для нескольких клубов.
- У клуба не более одного президента. Один и тот же человек может являться президентом нескольких клубов.

Заключение

- Логическое проектирование базы данных – отображение сущностей предметной области в набор реляционных отношений.
- Нормальная форма – совокупность ограничений, накладываемых на схему отношения. Ограничения исключают аномалии при работе с базой данных.
- Нормализация состоит в получении из исходного набора отношений нормализованных отношений, каждое из которых находится в НФ как можно более высокого порядка.