

Нормализация схемы отношений

Базы данных

Виноградова М.В.

МГТУ им. Н.Э. Баумана, ИУ5

Ограничения реляционной модели

- Внутренние ограничения модели:
 - неделимые (атомарные) значения
 - Каждая таблица должна иметь первичный ключ
 - Возможные ключи
 - Нет дубликатов картежей (строк)
 - Домены (строковые, числовые и т.д.)
- Явные ограничения
 - Описывают предметную область
 - Функциональные зависимости
 - Многозначные зависимости

Условные обозначения

- A, B, C, \dots — одиночные атрибуты,
- X, Y, Z — множества атрибутов,
- a, b, c, x, y, z — значения атрибутов.
- Объединение множеств атрибутов, например, $(A \text{ и } Y)$ — AY .

- U, R, S — отношения.

- Предположим, что существует универсальное отношение U , в схеме которого каждый атрибут имеет уникальное имя. Причем множество атрибутов любого другого отношения есть некоторое подмножество атрибутов схемы отношения U .

Функциональная зависимость

- Пусть X и Y — атрибуты отношения R .
- **Атрибут Y отношения R функционально зависит от атрибута X отношения R , если в каждый момент времени каждому значению x соответствует одно и то же значение y .**
- Функциональная зависимость f атрибута X от атрибута Y такова:

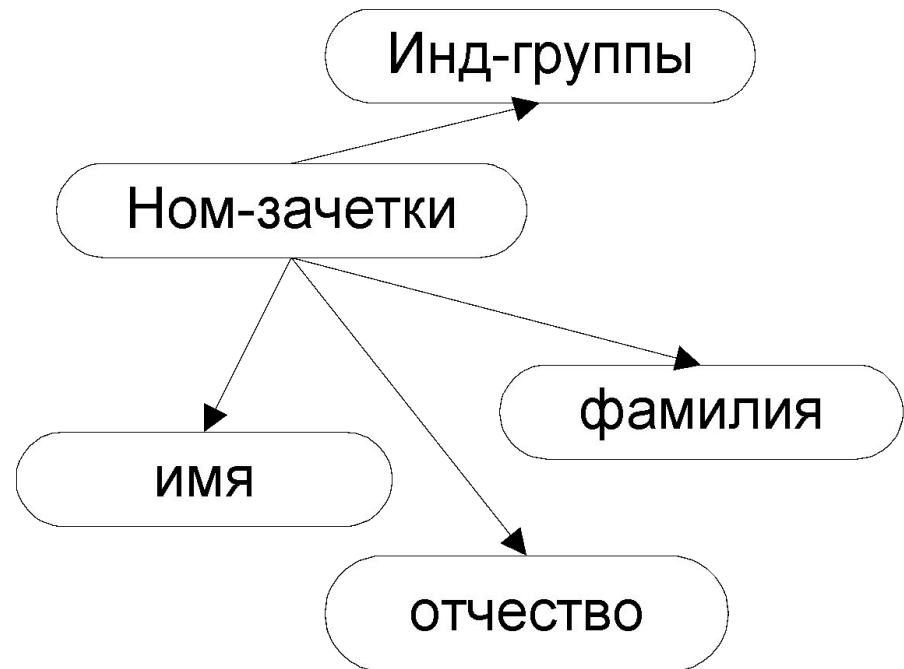
$$f : X \rightarrow Y$$

Свойства ФЗ

- Обозначение $X \neq Y$ показывает, что между X и Y нет функциональной зависимости.
- Если $A \rightarrow Y$ и $Y \rightarrow A$, то между A и Y существует взаимно однозначное соответствие.

Пример функциональной зависимости

Студент
ном-зачетки (PK)
фамилия *
имя *
отчество
инд-группы*



Описание ФЗ на примере

- На отношении $R(A_1, A_2, \dots, A_n)$ определено множество функциональных зависимостей $F = \{ A_1 \rightarrow A_2, A_1 \rightarrow A_3, \dots \}$

- На отношении

Студент(ном-зачетки, фамилия, имя, отчество, инд-группы)

определено мн-во ФЗ

$F = \{ \text{Ном-зачетки} \rightarrow \text{Фамилия}, \text{Ном-зачетки} \rightarrow \text{Имя}, \text{Ном-зачетки} \rightarrow \text{Отчество}, \text{Ном-зачетки} \rightarrow \text{Инд-группы} \}$

Многозначные зависимости

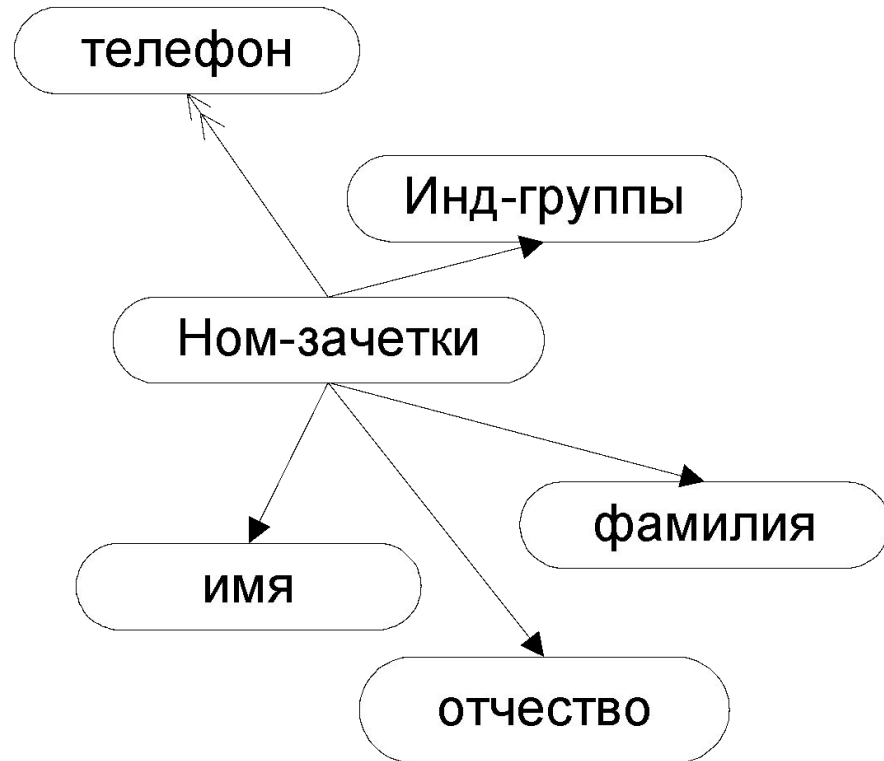
- Многозначная зависимость m Y от X :

$$m: X \twoheadrightarrow Y.$$

- Многозначная зависимость существует, если при заданных значениях атрибутов из X существует множество, состоящее из нуля или более взаимосвязанных значений атрибутов из Y ,
- причем множество значений Y не связано со значениями атрибутов в отношении « $R-X-Y$ », где R — все множество атрибутов отношения.

Пример МЗ

Студент
ном-зачетки (РК)
фамилия *
имя *
отчество
инд-группы*
телефон



Описание МЗ в примере

- На отношении $R(A_1, A_2, \dots, A_n)$ определено множество функциональных и многозначных зависимостей $F = \{ A_1 \rightarrow A_2, A_1 \rightarrow A_3, \dots, A_1 \rightarrow \dots \rightarrow A_n \}$

- На отношении

Студент(ном-зачетки, фамилия, имя, отчество, инд-группы)

определено мн-во ФЗ

$F = \{ \text{Ном-зачетки} \rightarrow \text{Фамилия}, \text{Ном-зачетки} \rightarrow \text{Имя}, \text{Ном-зачетки} \rightarrow \text{Отчество}, \text{Ном-зачетки} \rightarrow \text{Инд-группы}, \text{Ном-зачетки} \rightarrow \dots \rightarrow \text{Телефон} \}$

Проверка наличия МЗ в ОТНОШЕНИИ

- Если в R имеет место зависимость $X \twoheadrightarrow Y$, то для двух произвольных кортежей t и s , таких, что $t[X] = s[X]$ (т. е. t и s совпадают по значениям атрибутов X), отношение обязательно содержит кортежи u и v , такие, что выполняются условия:

$$u[X] = v[X] = t[X] = s[X]$$

$$u[Y] = t[Y]$$

$$u[R-X-Y] = s[R-X-Y]$$

$$v[Y] = s[Y]$$

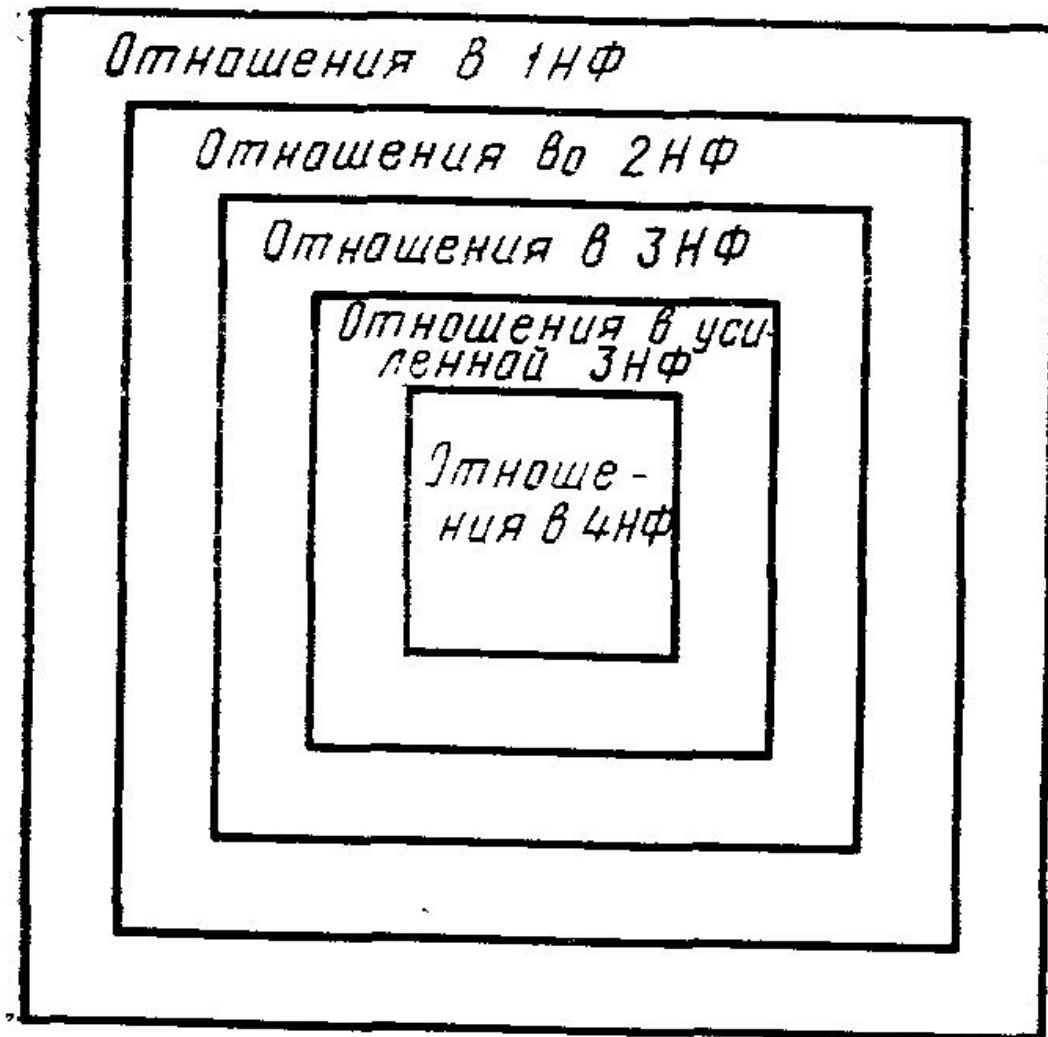
$$v[R-X-Y] = t[R-X-Y]$$

- Если поменять местами значения атрибутов Y в кортежах t и s , то можно получить два кортежа u и v , которые также должны, принадлежать рассматриваемому отношению.

Пример проверки МЗ

ФИО	Город	Улица	Организация	должность
Иванов ИИ	Клин	Ленина	ООО	Секретарь
Иванов ИИ	Клин	Ленина	МММ	Директор
Петров ПП	Москва	Янгеля	МГТУ	Доцент
Петров ПП	Москва	Янгеля	ЛГУ	Декан
Петров ПП	СПб	Невский	МГТУ	Доцент
Петров ПП	СПб	Невский	ЛГУ	Декан

Нормализация схемы БД



1-ая нормальная форма (1ПФ)

- Схема отношения R находится в первой нормальной форме тогда и только тогда, когда все входящие в нее атрибуты являются атомарными
- т. е. значения соответствующих доменов рассматриваются как неделимые, а не как множества или кортежи из более элементарных доменов.

2-ая нормальная форма (2НФ)

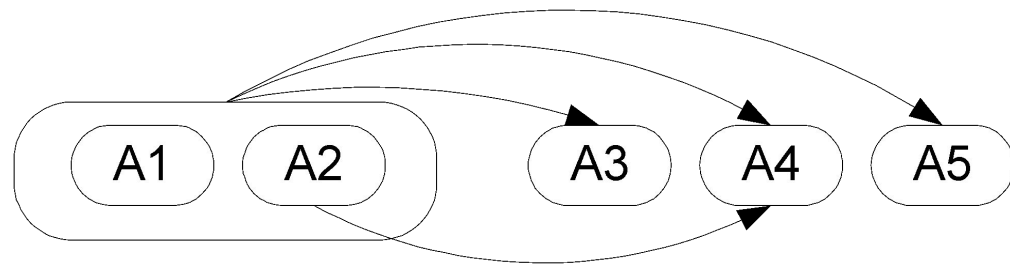
- Схема отношения R находится во второй нормальной форме, если она находится в 1НФ и каждый ее непервичный атрибут функционально полно зависит от первичного ключа.

Свойство 2НФ

- Если X — ключ отношения R , $Y \leq X$,
- A — первичный (ключевой) атрибут отношения R ,
- то в отношении R имеет место частичная зависимость (неполная функциональная зависимость), когда
$$X \rightarrow A \text{ и } Y \rightarrow A.$$
- Если это условие не выполняется, то атрибут A функционально полно зависит от X в отношении R .
- Отношение, находящееся во 2НФ может обладать аномалиями для выполнения операций включения, удаления и модификации.

Пример 2НФ

Исходное отношение $R(\underline{A1}, \underline{A2}, A3, A4, A5)$ – нарушает 2НФ

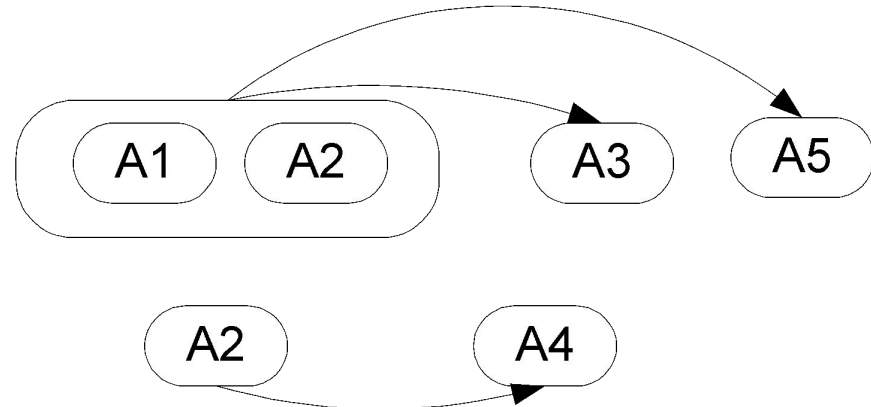


Результат
декомпозиции
исходного
отношения:

$R1(\underline{A1}, \underline{A2}, A3, A5)$

$R2(\underline{A2}, A4)$

- соответствуют 2НФ

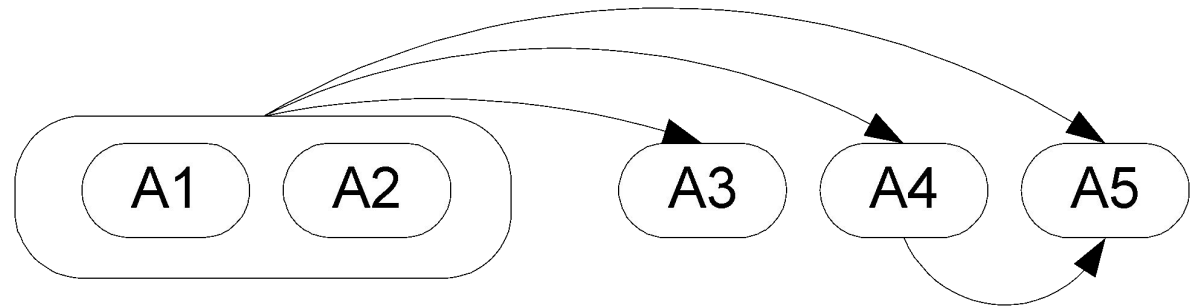


3-ая нормальная форма (3НФ)

- Схема отношения R находится в третьей нормальной форме, если не существует ключа X для R , множества атрибутов $Y \leq R$ и непервичного атрибута A из R , таких, что:
 - $X \rightarrow Y$ справедливо в R ,
 - $Y \rightarrow A$ справедливо в R , но $Y \rightarrow X$ не имеет места.
- Схема отношения R находится в третьей нормальной форме, если она находится в 2НФ и каждый непервичный атрибут нетранзитивно зависит от первичного ключа.

Пример 3НФ

Исходное отношение $R(\underline{A1}, A2, A3, A4, A5)$ - нарушает 3НФ



Результат
декомпозиции
исходного

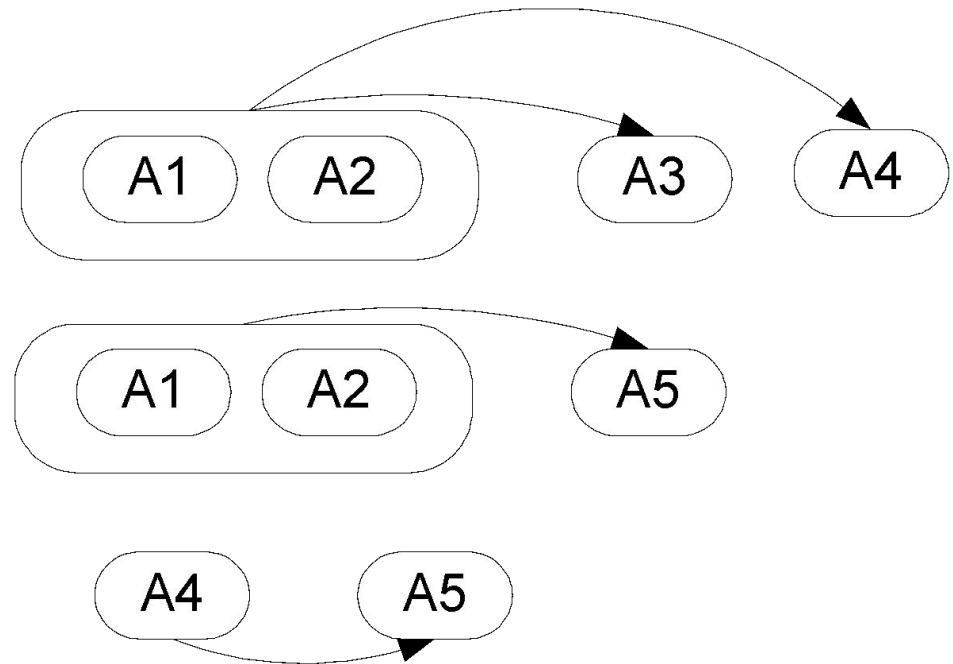
отношения:

$R1(\underline{A1}, A2, A3, A4)$

$R2(\underline{A1}, \underline{A2}, A5)$

$R3(\underline{A4}, A5)$

- Соответствуют 3НФ



Свойство 3НФ

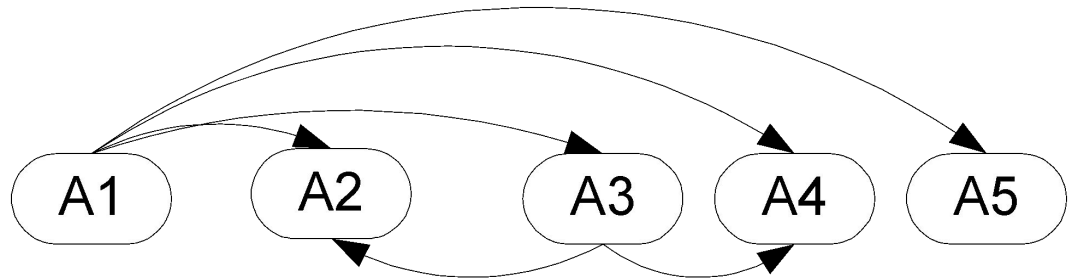
- Когда отношение имеет только один ключ и другие зависимости, в том числе многозначные, в нем отсутствуют, то 3НФ освобождает от избыточности и аномалий выполнения операций включения, удаления и модификации.
- Если в отношении, находящемся в 3НФ, отсутствуют многозначные зависимости, но имеются другие зависимости, кроме зависимости от ключа, то 3НФ будет иметь аномалии операций.
- В этом случае рассматривают усиленную третью нормальную форму (нормальная форма Бойса — Кодда).

3-ая усиленная нормальная форма (НФ Бойса-Кодда)

- Схема отношения R с зависимостями F находится в усиленной третьей нормальной форме, если всякий раз, когда в R имеет место зависимость $X \rightarrow A$ и A не принадлежит X , то X является возможным ключом отношения R .
- Атрибут (или комбинацию атрибутов), от которого какой-либо другой атрибут зависит функционально (полно), называют детерминантой.
- В этом случае нормализованная схема отношения R находится в усиленной третьей нормальной форме, если каждая детерминанта является возможным ключом.
- Если в отношении отсутствуют многозначные зависимости, то снимаются все аномалии операций включения, удаления и модификации.

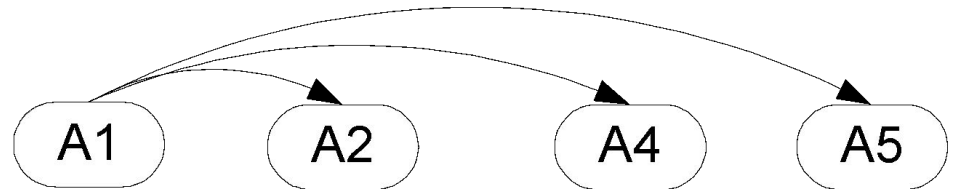
Пример НФБК

Исходное отношение $R(\underline{A1}, A2, A3, A4, A5)$ - нарушает НФБК



Детерминанты

A1 и A3



Результат

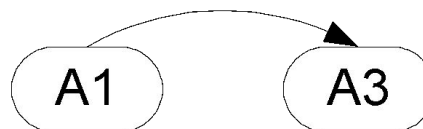
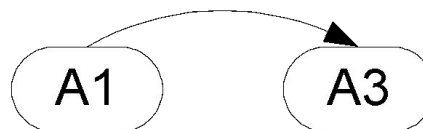
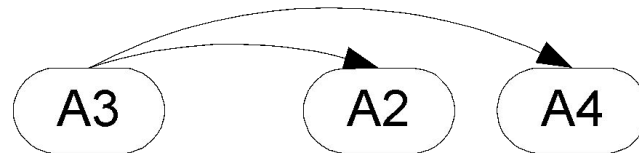
декомпозиции исходного
отношения:

$R1(\underline{A1}, A2, A4, A5)$

$R2(\underline{A3}, A2, A4)$

$R3(\underline{A1}, A3)$ –

соответствуют НФБК

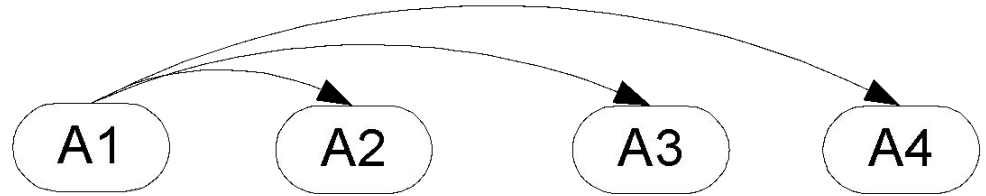


4 -ая нормальная форма (4 НФ)

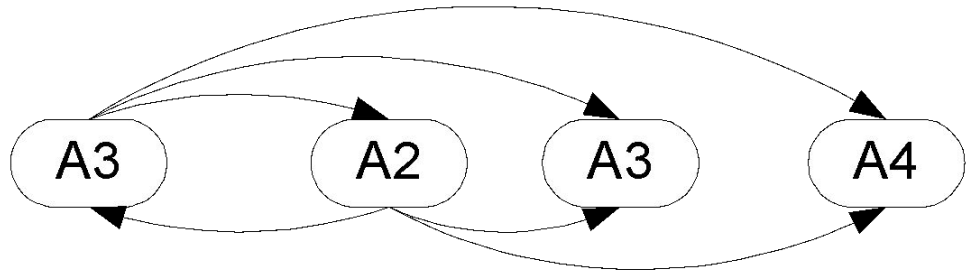
- Если в отношении присутствуют многозначные зависимости, то схема отношения должна находиться в четвертой нормальной форме. В противном случае снова возникнут соответствующие аномалии.
- Схема отношения R будет находиться в четвертой нормальной форме, если всякий раз, когда существует многозначная зависимость $X \twoheadrightarrow Y$ (где Y — не пустое множество и не является подмножеством X и XY состоит не из всех атрибутов R), также существует зависимость $X \rightarrow A$ для любого атрибута A в R .

Примеры 4 НФ

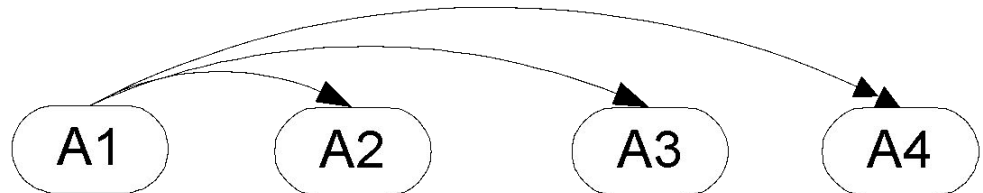
- 3НФ, 3УНФ, 4 НФ



- 3УНФ, 4 НФ



- 4 НФ



Аномалии

- Сложность выполнения операций
- Отсутствуют при чтении
- Встречаются при добавлении, удалении, изменении
- Дубликаты, повтор действий, удаление данных

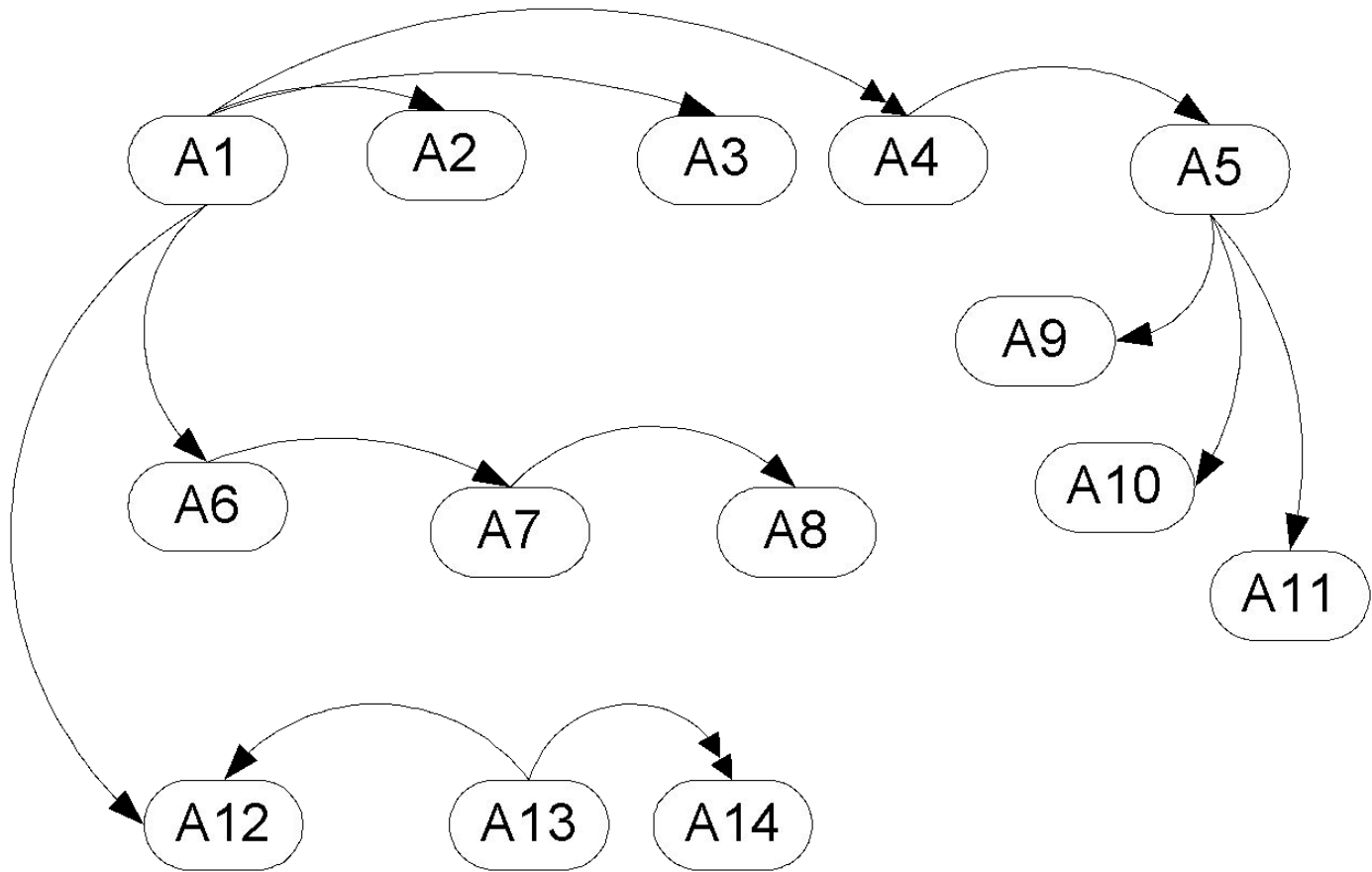
- Персона (ФИО, телефон, почта, организация, раб.телефон, адр.орг, должность)
 - При удалении персоны удалятся сведения об организации
 - При обновлении данных организации надо изменять всех персон
 - Дублируются сведения об организации
 - Нельзя добавить сведения об организации без персоны

- Персона (ФИО, телефон, почта, организация, должность)
- Организация(название, раб.телефон, адрес)
 - Аномалии отсутствуют

Нормализация

- Нормализация отношений выполняется декомпозицией их схем.
- Декомпозицией схемы отношения $R = \{A_1, A_2, \dots, A_n\}$ называется замена схемы совокупностью схем $\rho = \{R_1, R_2, \dots, R_n\}$ подмножеств, таких, что
$$R_1 \cup R_2 \cup \dots \cup R_k = R = \{A_1, A_2, \dots, A_n\}.$$
- При этом не требуется, чтобы схемы R_t были непересекающимися.
- Однако не все докомпозиции могут обладать свойством соединения без потерь (если естественное соединение не восстанавливает первоначальное отношение, то нет способа, выполняющего его однозначное восстановление) и свойством сохранения зависимостей.

Задание



Решение

- R1 (A1, A2, A3, A 4, A6, A12)
- R2(A 4 , A5)
- R3 (A5, A9, A1 0 , A11)
- R 4 (A6, A 7)
- R5 (A7 , A8)
- R6 (A13, A12, A14)

Атрибуты ПО (Студент)

- ФИО, ном-зачетки, г.рождения, телефон, инд.группы, уч.план, кафедра, тел. кафедры

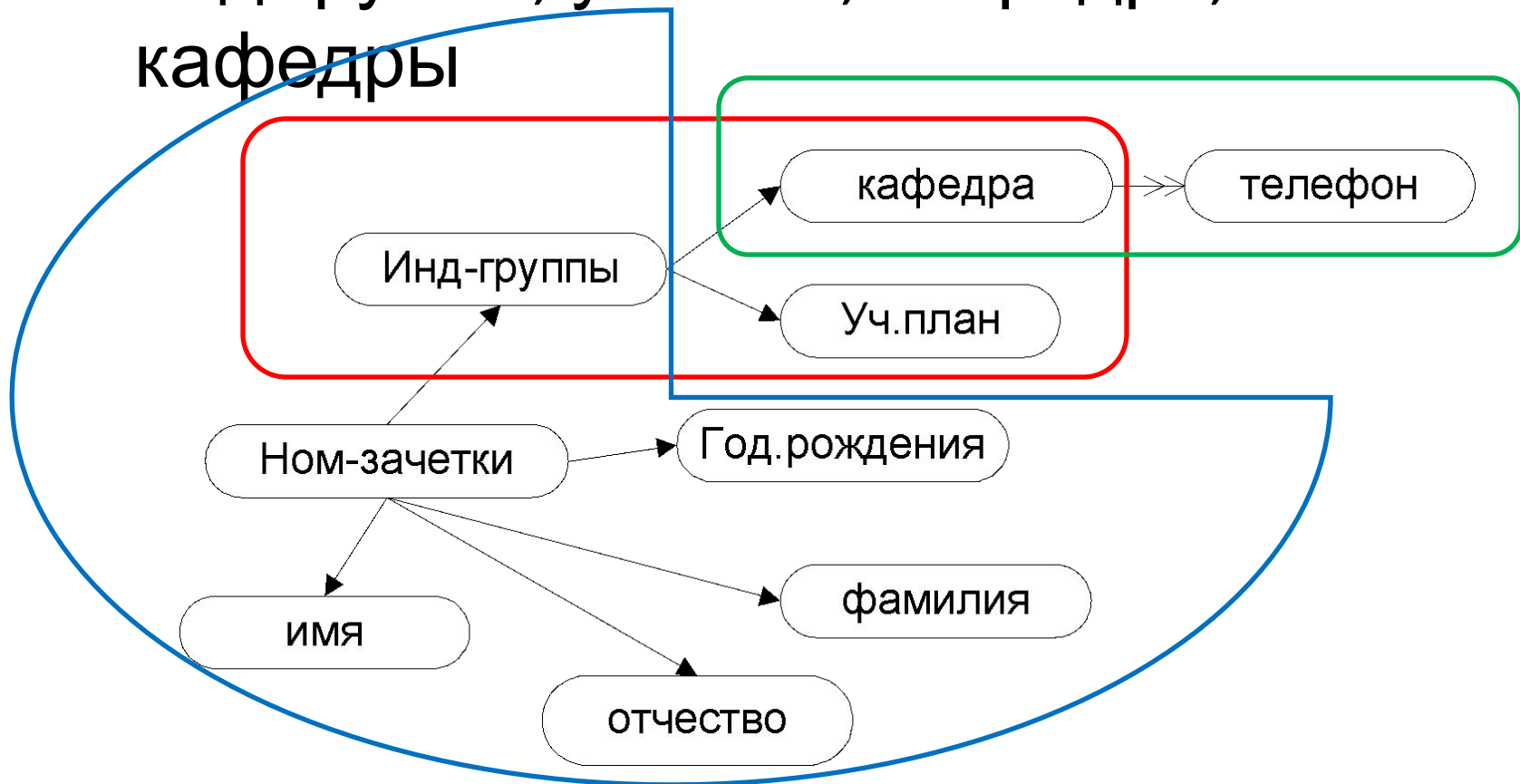


Схема БД Студента (4 НФ)

- Студент (Ном-зачетки, Фамилия, Имя, Отчество, год.рождения, Инд-кафедры)
- Группа(инд-группы, Уч.план, Кафедра)
- Кафедра(кафедра, телефон)

Таблицы БД студента

Ном-зачетки	Фамилия	Имя	Отчество	Год.рождения	Инд-группы
12345	Иванов	Иван	Петрович	2000	ИУ5-41

Инд-группы	Уч. план	кафедра
ИУ5-41	Б123	ИУ5
РТ4-61	Б444	РТ4

Кафедра	телефон
ИУ5	60-20
ИУ5	54-67
РТ4	90-18