Тема 9

ТЕОРИЯ НОРМАЛИЗАЦИИ

Теория нормализации

Нормализация данных – это декомпозиция (разбиение путем проецирования) отношения, находящегося в предыдущей нормальной форме, на два или более отношений, удовлетворяющих требованиям следующей нормальной формы.

Нормализация данных – это разложение отношений на большее количество более простых таблиц.

Цели нормализации

1. Основной целью теории нормальных форм первоначально была экономия места на диске.
2. Данные должны быть устроены так, чтобы при их редактировании или удалении, необходимо было исправлять данные только в одном месте БД.
3. Группировка данных по содержанию (каждая таблица - определенная тематика).
4. Принцип модульности (несколько унифицированных независимых блоков).

Первая нормальная форма

Отношение ***R*** находится в ***первой нормальной форме*** (***1NF***) тогда и только тогда, когда значения всех его атрибутов атомарны.

Замечания:

1. Отношение ***R*** должно быть регулярно.
2. Отношение ***R*** не должно содержать вычисляемых полей.

Примеры

1НФ

Преподаватели-предметы	
Преподаватель	**Предмет**
Иванов	Физика
Иванов	Химия
Петров	Русский язык
Петров	Литература

Не удовлетворяет 1НФ

Преподаватели-предметы	
Преподаватель	**Предмет**
Иванов	Физика Химия
Петров	Русский язык Литература

1НФ

СТУДЕНТЫ	
Студент	**Дата рождения**
Иванов	06.06.1996
Иванов	06.08.1995
Петров	06.09.1995
Петров	09.06.1997

Не удовлетворяет 1НФ

СТУДЕНТЫ	
Студент	**Возраст**
Иванов	13
Иванов	14
Петров	14
Петров	12

Декомпозиция без потерь

Основой нормализации является процесс разбиения - или ***декомпозиции***. Причем нас будет интересовать не просто процесс декомпозиции, а процесс ***декомпозиции без потерь***.

Определение. Процесс декомпозиции будем называть ***декомпозицией без потерь***, если из полученных отношений можно полностью восстановить исходное отношение без потери информации.

По своей сути декомпозиция представляет собой ***проекцию***. А обратное преобразование - ***операция соединения***.

Корректные и некорректные декомпозиции отношений

СЛУЖАЩИЕ_ПРОЕКТЫ

СЛУ_НОМЕР	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРОЕКТ_РУК
2934	Иванов	22000.00	1	Иванов
2941	Иваненко	22000.00	2	Иваненко

Декомпозиция (1). Отношения **СЛУЖ** и **СЛУ_ПРО**

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000.00
2941	Иваненко	22000.00

СЛУ_НОМ	ПРО_НОМ	ПРОЕКТ_РУК
2934	1	Иванов
2941	2	Иваненко

Декомпозиция (2). Отношения **СЛУЖ** и **ЗАРП_ПРО**

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000.00
2941	Иваненко	22000.00

СЛУ_ЗАРП	ПРО_НОМ	ПРОЕКТ_РУК
22000.00	1	Иванов
22000.00	2	Иваненко

Корректные и некорректные декомпозиции отношений

STUDENTS

Student_ID	Surname	Rating	Course	Leader
17	Иванов	25	1	Козлов
23	Смехова	25	2	Фурсенко

Декомпозиция (1) (без потерь)

STUD

Student_ID	Surname	Rating
17	Иванов	25
23	Смехова	25

STUD—LEADER

Student_ID	Course	Leader
17	1	Козлов
23	2	Фурсенко

Результат соединения (1)

Student_ID	Surname	Rating	Course	Leader
17	Иванов	25	1	Козлов
23	Смехова	25	2	Фурсенко

Корректные и некорректные декомпозиции отношений

STUDENTS

Student_ID	Surname	Rating	Course	Leader
17	Иванов	25	1	Козлов
23	Смехова	25	2	Фурсенко

Декомпозиция (2) (с потерями)

STUD

Student_ID	Surname	Rating
17	Иванов	25
23	Смехова	25

RATING—LEADER

Rating	Course	Leader
25	1	Козлов
25	2	Фурсенко

Результат соединения (2)

Student_ID	Surname	Rating	Course	Leader
17	Иванов	25	1	Козлов
23	Смехова	25	2	Фурсенко
17	Иванов	25	2	Козлов
23	Смехова	25	1	Фурсенко

Теорема Хита

Пусть дано отношение ***r (A, B, C),***
где ***A***, ***B*** и ***C*** – непересекающиеся подмножества множества (в противном случае получим избыточность FD) атрибутов ***R***.
Причём множество атрибутов ***R*** равно объединению подмножеств ***A***, ***B***, ***C***.
Если ***R*** удовлетворяет функциональной зависимости ***B*** → ***C***, то ***R*** равно естественному соединению его проекций **{A, B}** и **{B, C}**.

r = (r PROJECT {A, B}) NATURAL JOIN (r PROJECT {B, C}).

ГРУППЫ атрибутов: A, B и C. B → C

A		B	C			
Surname	**Rating**	**Student_ID**	**Sex**	**Birthday**	**Course**	**Leader**
Иванов	25	17	М	21.12.1992	1	Козлов
Козлов	22	18	М	12.09.1992	1	Козлов
Семибаба	23	19	М	11.08.1991	1	Козлов
Мудрый	17	20	М	05.03.1992	1	Козлов
Фурсенко	44	21	Ж	01.01.1990	2	Фурсенко
Иванова	47	22	Ж	21.12.1992	2	Фурсенко
Смехова	43	23	Ж	07.11.1991	2	Фурсенко

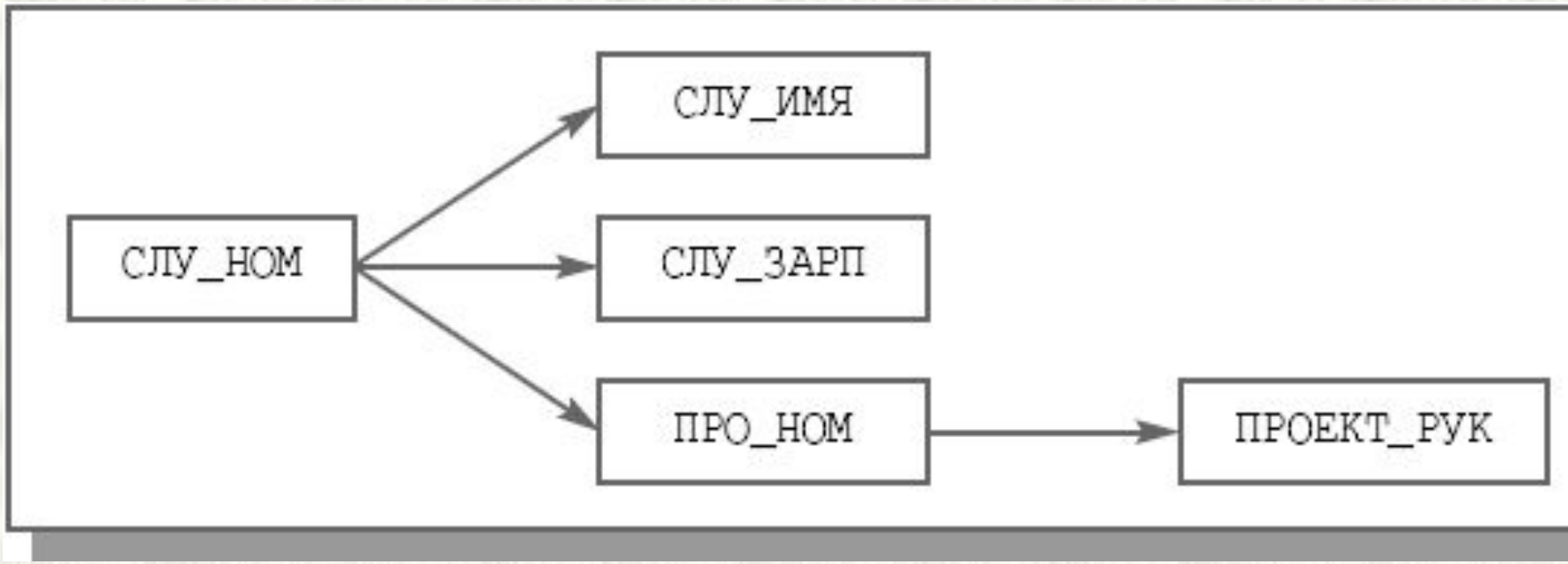
A		B
Surname	**Rating**	**Student_ID**
Иванов	25	17
Козлов	22	18
Семибаба	23	19
Мудрый	17	20
Фурсенко	44	21
Иванова	47	22
Смехова	43	23

B	C			
Student_ID	**Sex**	**Birthday**	**Course**	**Leader**
17	М	21.12.1992	1	Козлов
18	М	12.09.1992	1	Козлов
19	М	11.08.1991	1	Козлов
20	М	05.03.1992	1	Козлов
21	Ж	01.01.1990	2	Фурсенко
22	Ж	21.12.1992	2	Фурсенко
23	Ж	07.11.1991	2	Фурсенко

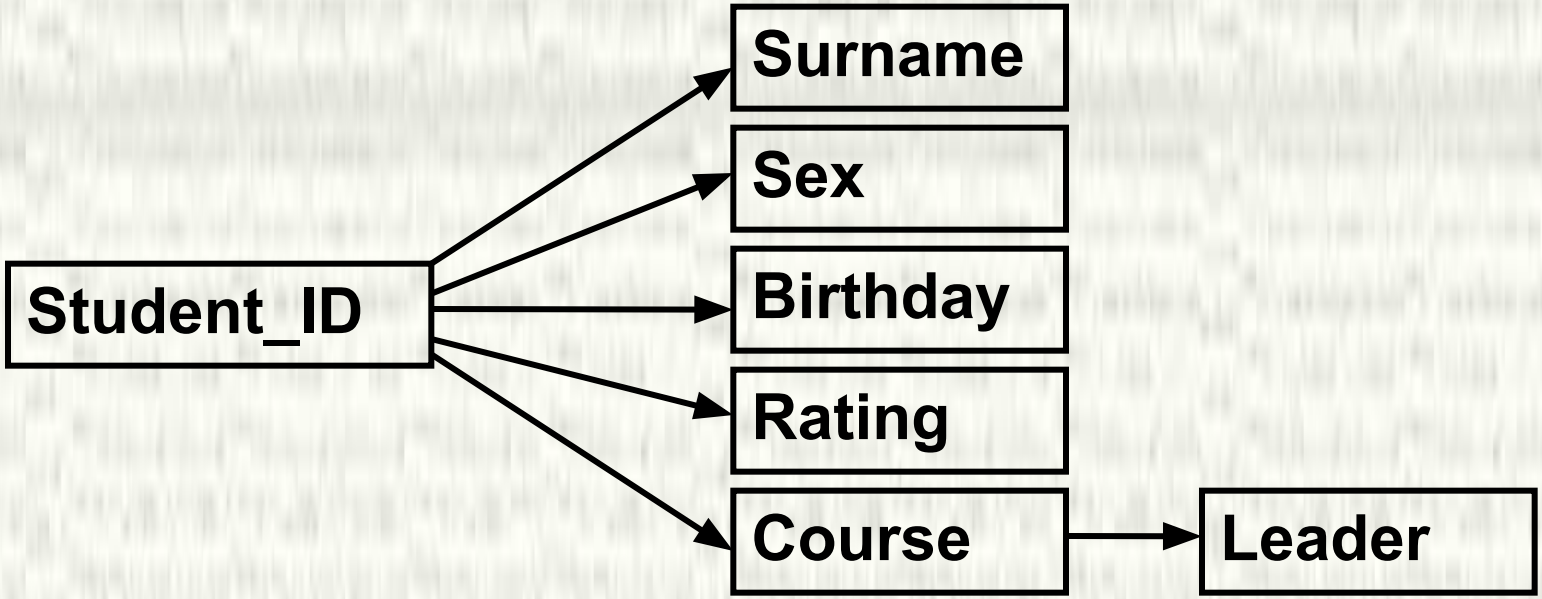
- ***Атрибут B минимально зависит от атрибута A***, если выполняется минимальная слева *FD* **A→B**.

Диаграммы функциональных зависимостей

- *Минимальные множества FD* можно наглядно представлять с помощью *диаграмм FD*.



Диаграммы функциональных зависимостей



Замечание. Первичный ключ всегда является детерминантом, **НО** детерминант не обязательно является первичным ключом!

Данные по студентам ф-та математики, физики и информатики

Student_ID	Surname	Rating	Sex	Birthday	Speciality	Value	Group	Leader
17	Иванов	25	М	21.12.1992	ПИ	56000	1	Козлов
18	Козлов	22	М	12.09.1992		56000	1	Козлов
19	Семибаба	23	М	11.08.1991		56000	1	Козлов
20	Мудрый	17	М	05.03.1992		56000	1	Козлов
21	Фирсов	44	Ж	01.01.1990	ДВ	42000	2	Фирсов
22	Иванова	47	Ж	21.12.1992		42000	2	Фирсов
23	Смехова	43	Ж	07.11.1991		42000	2	Фирсов
20	Мудрый	49	М	05.03.1992		42000	2	Фирсов
22	Смирнова	51	Ж	21.12.1992	ИНФ	56000	3	Петров

Отношение Students_Value_Leader

Student_ID	Surname	Rating	Sex	Birthday	Speciality	Value	Group	Leader
17	Иванов	25	М	21.12.1992	ПИ	56000	1	Козлов
18	Козлов	22	М	12.09.1992	ПИ	56000	1	Козлов
19	Семибаба	23	М	11.08.1991	ПИ	56000	1	Козлов
20	Мудрый	17	М	05.03.1992	ПИ	56000	1	Козлов
21	Фирсов	44	М	01.01.1990	ДВ	42000	2	Фирсов
22	Иванова	47	Ж	21.12.1992	ДВ	42000	2	Фирсов
23	Смехова	43	Ж	07.11.1991	ДВ	42000	2	Фирсов
20	Мудрый	49	М	05.03.1992	ДВ	42000	2	Фирсов
24	Смирнова	51	Ж	21.12.1992	ИНФ	56000	3	Петров

Организация данных: первая нормальная форма 1NF

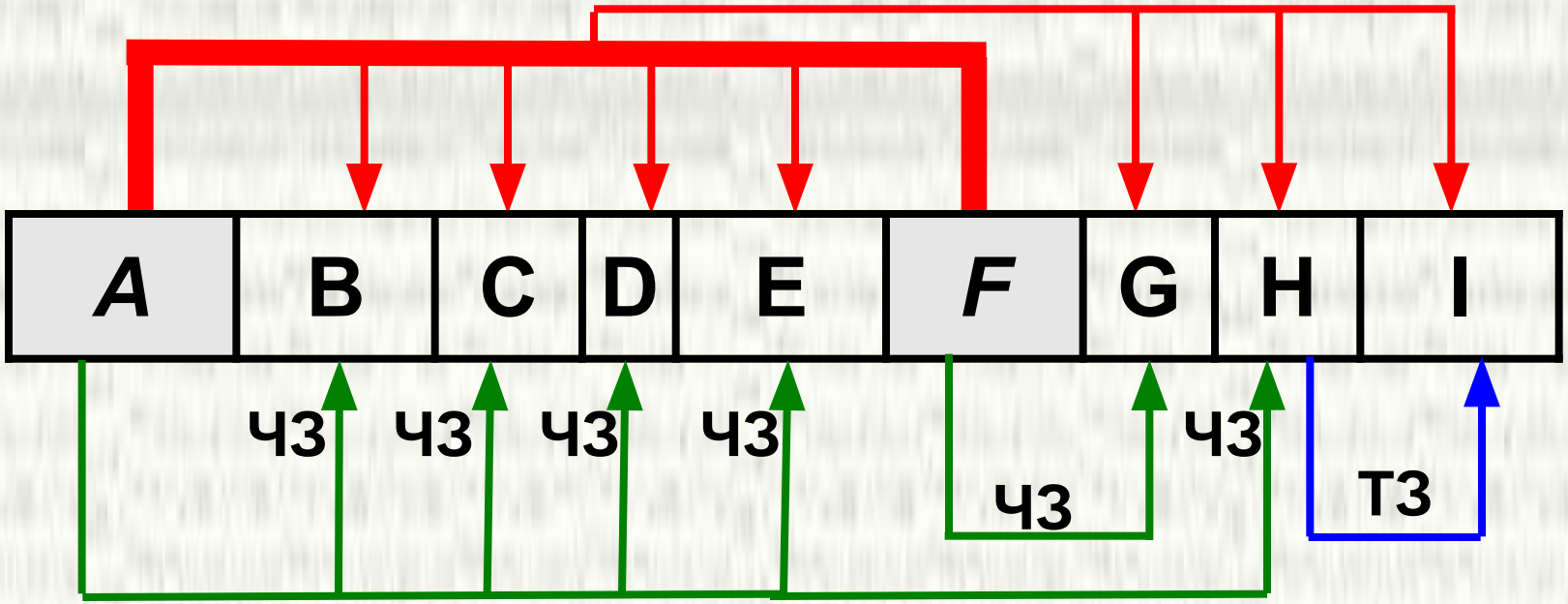
Аномалии модификации

Избыточность данных приводит к следующим аномалиям:

- **аномалия удаления**
- **аномалия вставки**
- **аномалия модификации**

Необходимость нормализации прямо следует из необходимости устранения аномалий

Диаграмма зависимостей: первая нормальная форма (1NF)



A — атрибут первичного ключа

первичный ключ

зависимости, основанные на первичном ключе

частичные зависимости

транзитивные зависимости

Комментарии к диаграмме зависимостей

- **Частичная зависимость (partial dependency)** – зависимость, определяемая только частью составного первичного ключа.
- **Транзитивная зависимость (transitive dependency)** – зависимость одного непервичного атрибута от другого непервичного атрибута.

Замечание: транзитивные зависимости могут стать причиной аномалии данных.

Приведение к 1NF

- **Таблица приведена к 1NF, если в ней:**
  - определены все ключевые атрибуты;
  - отсутствуют повторяющиеся группы (на пересечении каждого столбца и каждой строки содержится только одно (атомарное) значение);
  - все атрибуты зависят от первичного ключа.

Замечания:

1. Отношение должно быть регулярно.
2. Отношение не должно содержать вычисляемых полей.
3. Все реляционные таблицы удовлетворяют требованиям, предъявляемым к 1NF.

Приведение к 2NF

1. Students {**Student_ID;** Surname; Sex; Birthday}

Student_ID	Surname	Sex	Birthday
17	Иванов	М	21.12.1992
18	Козлов	М	12.09.1992
19	Семибаба	М	11.08.1991
20	Мудрый	М	05.03.1992
21	Фирсов	М	01.01.1990
22	Иванова	Ж	21.12.1992
23	Смехова	Ж	07.11.1991
24	Смирнова	Ж	21.12.1992

3. Assignment {**Student_ID**; **Speciality**; Rating; Group; Leader}

Student_ID	Speciality	Rating	Group	Leader
17	ПИ	25	1	Козлов
18	ПИ	22	1	Козлов
19	ПИ	23	1	Козлов
20	ПИ	17	1	Козлов
21	ДВ	44	2	Фирсов
22	ДВ	47	2	Фирсов
23	ДВ	43	2	Фирсов
20	ДВ	49	2	Фирсов
24	ИНФ	51	3	Петров

2. Specialities {**Speciality;** Value}

Speciality	Value
ПИ	56000
ДВ	42000
ИНФ	56000

Результат приведения к 2NF

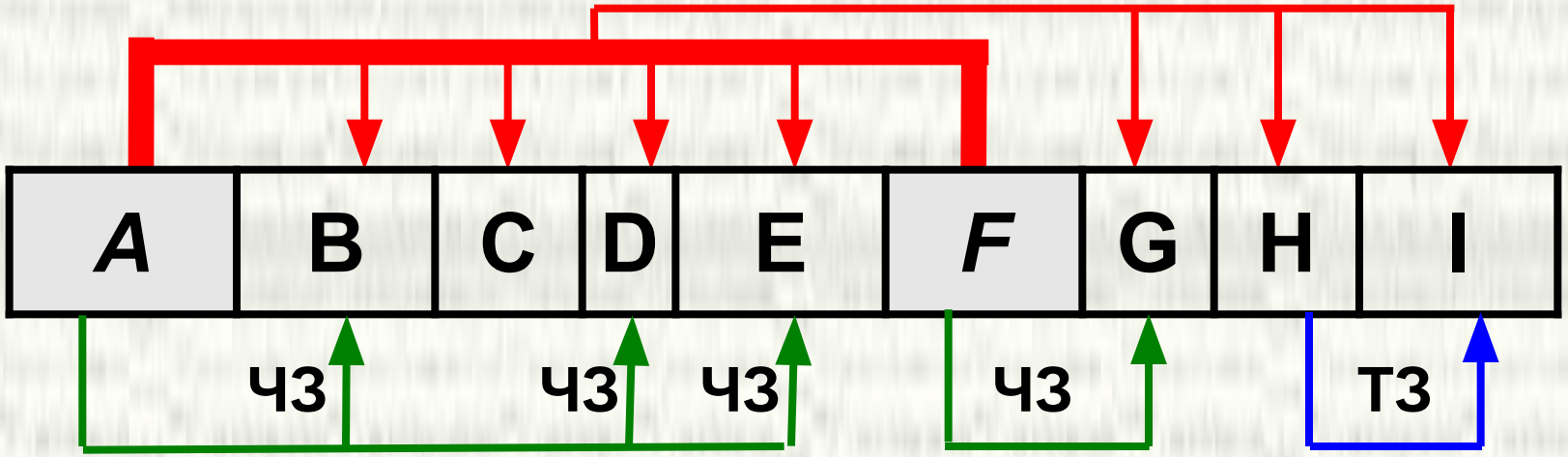


Таблица Students

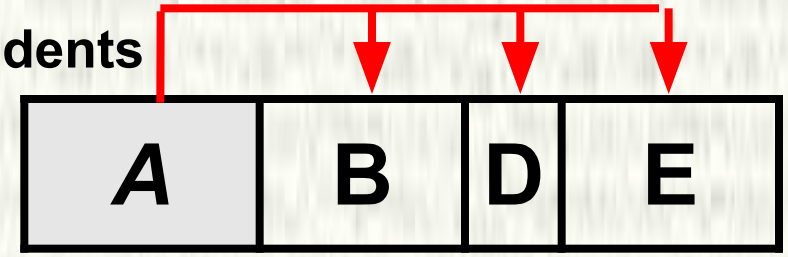


Таблица Specialities

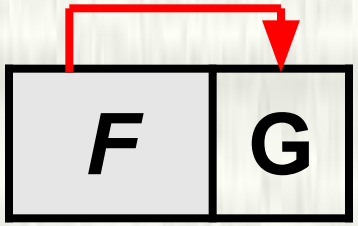
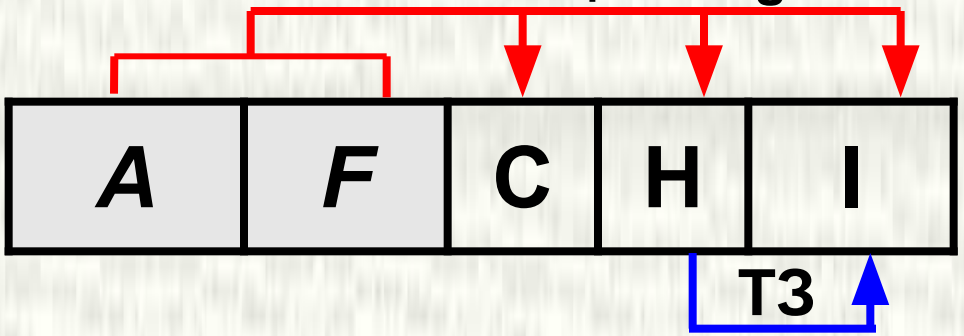


Таблица Assignment



Приведение к 2NF

- **Таблица приведена к 2NF, если:**
  - она приведена к 1NF;
  - в ней отсутствуют частичные зависимости.

Замечания:

1. В таблицах, приведённых к 2NF, может иметь место транзитивная зависимость.
2. Таблицы, в которых первичный ключ содержит всего лишь 1 атрибут, автоматически будут иметь 2NF, если они имеют 1NF.

Приведение к 3NF

1. Students {**Student_ID;** Surname; Sex; Birthday;}

Student_ID	Surname	Sex	Birthday
17	Иванов	М	21.12.1992
18	Козлов	М	12.09.1992
19	Семибаба	М	11.08.1991
20	Мудрый	М	05.03.1992
21	Фирсов	М	01.01.1990
22	Иванова	Ж	21.12.1992
23	Смехова	Ж	07.11.1991
24	Смирнова	Ж	21.12.1992

2.Specialities {**Speciality;**Value}

Speciality	Value
ПИ	56000
ДВ	42000
ИНФ	56000

3. Assignment {**Student_ID**; **Speciality;** Rating; Group}

Student_ID	Speciality	Rating	Group
17	ПИ	25	1
18	ПИ	22	1
19	ПИ	23	1
20	ПИ	17	1
21	ДВ	44	2
22	ДВ	47	2
23	ДВ	43	2
20	ДВ	49	2
24	ИНФ	51	3

4. Leader {**Group**; Leader}

Group	Leader
1	Козлов
2	Фирсов
3	Петров

Результат приведения к 3NF

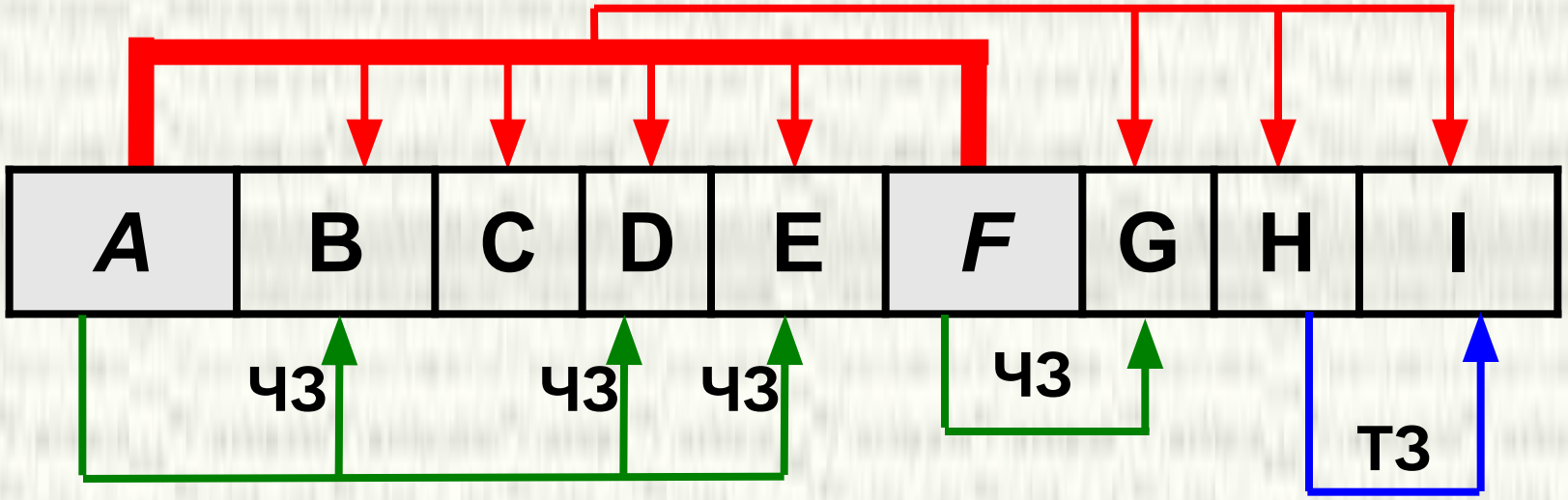


Таблица Students

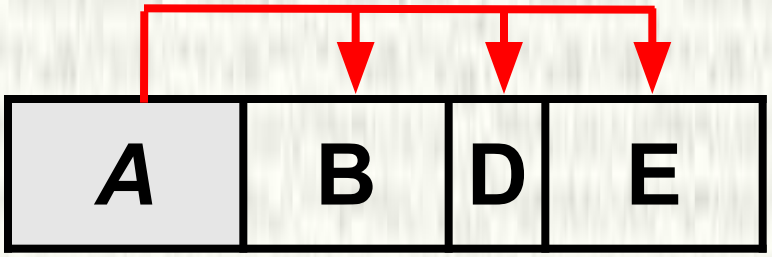


Таблица Specialities

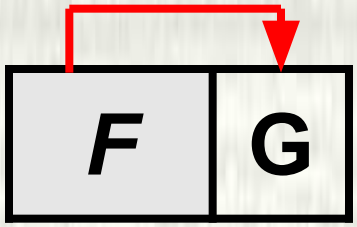


Таблица Assignment

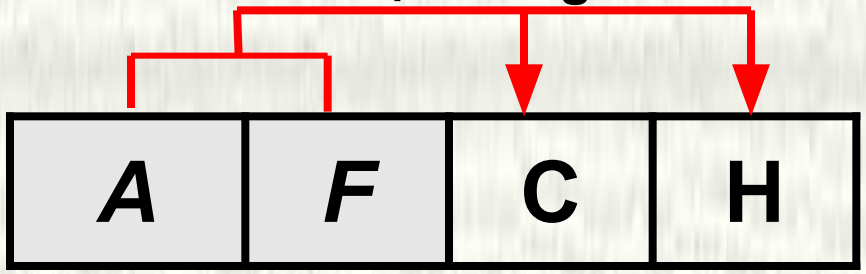
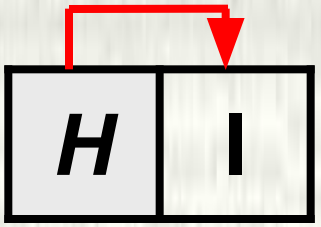


Таблица Leader



Приведение к 3NF

- **Таблица приведена к 3NF, если:**
  - она приведена к 2NF;
  - в ней отсутствуют транзитивные зависимости.

Нормальная форма Бойса-Кодда (BKNF)

- **Таблица приведена к BKNF, если:**
  - она приведена к 3NF;
  - каждый детерминант таблицы является потенциальным ключом.

Замечание.

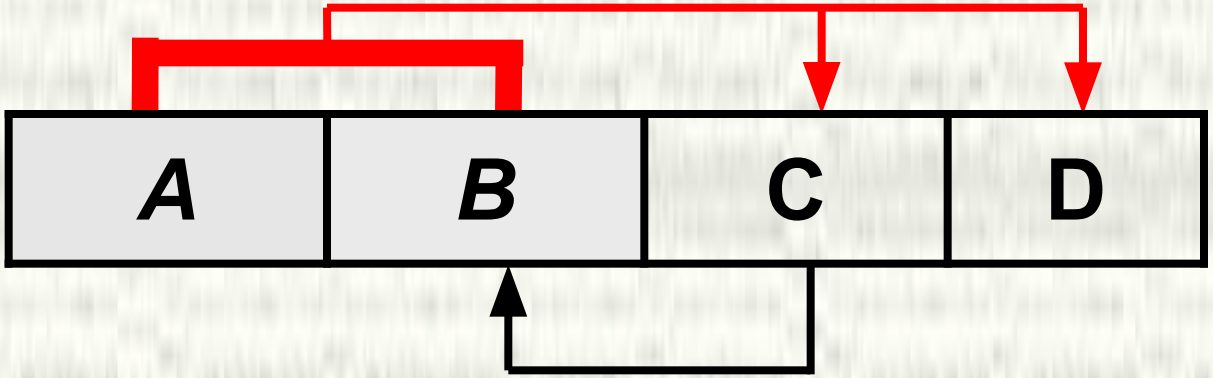
Если таблица содержит только один потенциальный ключ, то формы **3NF** и **BKNF** эквивалентны.

Необходимость приведения 3NF к BKNF

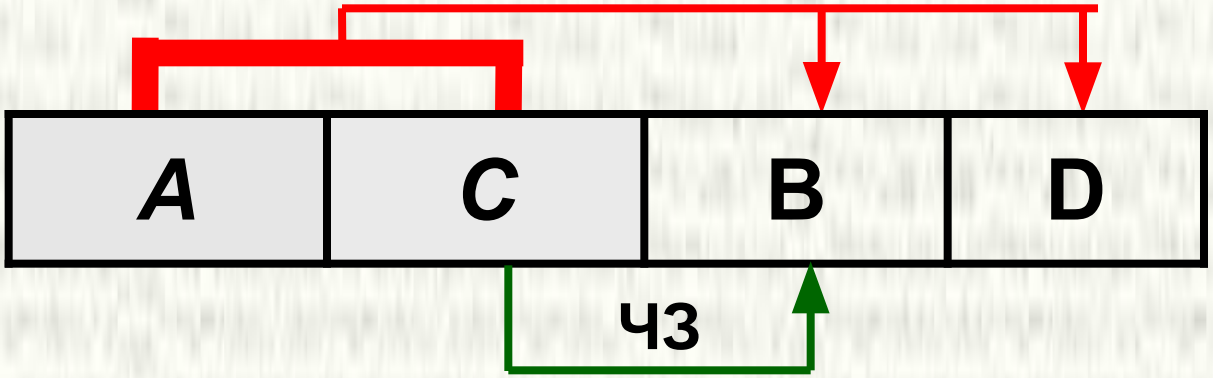
- Если **неключевой атрибут является детерминантом ключевого атрибута**, это удовлетворяет требованиям отношения в **3NF** форме, но не отвечает правилам **BKNF**.
- **A + B → C, D**

  C → B (неключевой атрибут определяет часть первичного ключа, зависимость **нетранзитивна**)

Декомпозиция структуры таблицы для выполнения требований BKNF

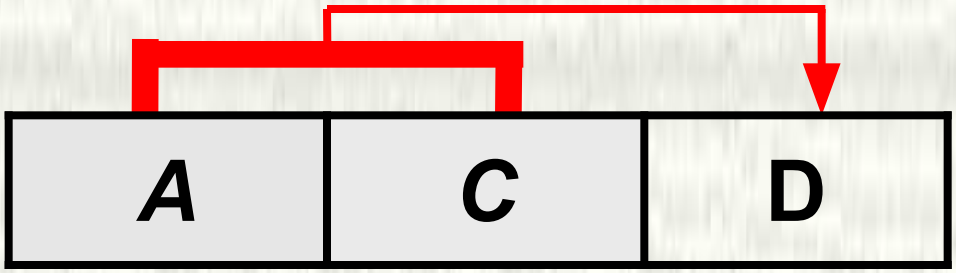


3NF но не BKNF

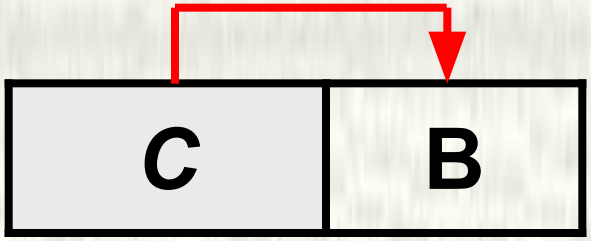


1NF

3NF и BKNF



3NF и BKNF



Пример

Рассмотрим отношение R:

Группа	Кол-во	Специальность
4311	15	5201
4361	15	3515
2311	21	5201

Пример

Декомпозиция без потерь:

Группа	Кол-во
4311	15
4361	15
2311	21

Группа	Специальность
4311	5201
4361	3515
2311	5201

Пример

Декомпозиция с потерями

Группа	Кол-во
4311	15
4361	15
2311	21

Кол-во	Специальность
15	5201
15	3515
21	5201