

Реляционная модель данных

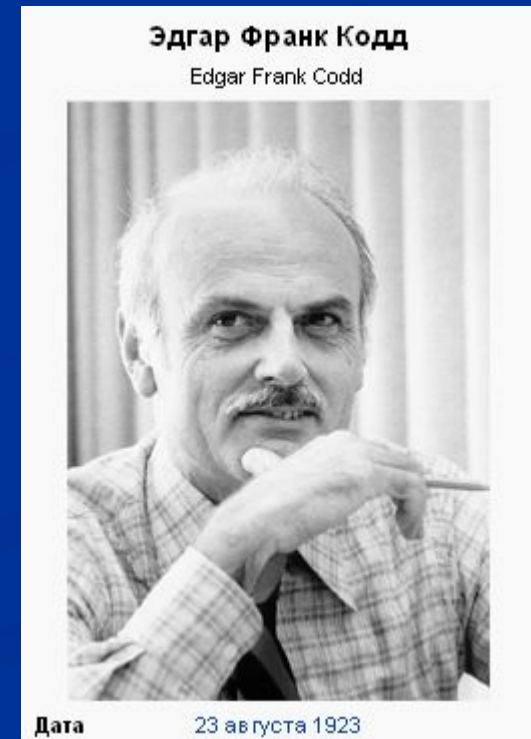
Технологии баз данных

Из истории РМД

19 августа 1969 года вышла статья Эдгара Кодда посвященная реляционному представлению данных. Т.о. данный подход существует уже более 40 лет. Реляционная модель данных жива и здравствует, и пока что не видно реальных претендентов на ее замену. Возможно, это объясняется отсутствием в настоящее время в сообществе баз данных людей масштаба доктора Кодда, но, скорее всего, просто отсутствуют потребности в других абстрактных средствах представления данных. Все существующие модели происходят из нее и на ее базе.

С другой стороны, современное представление "истинной" реляционной модели данных обеспечивается в работах последних десятилетий Криса Дейта и Хью Дарвена. Не нарушая предписания, сформулированные в первых статьях Эдгара Кодда Дейт и Дарвен предлагают концепцию построения новых СУБД.

Сам термин «реляционная модель» у Кодда появилась только в 1979 году, а понятие модели— только в 1981



Основные понятия РМД

- **Реляционная модель данных** (РМД, relation – отношение, математический термин, обозначающий таблицу) – это теория данных, основанная на некоторых положениях математики (теория множеств и предикативной логики).
- Любое предприятие имеет большое количество данных, связанных с его деятельностью (о продукции, о счетах, о планировании, о студентах и др.) Эти данные представляют основные **объекты**, они связаны между собой **отношениями**. Поэтому в БД отражаются и объекты и отношения между ними.

Основные принципы РМД

Основные принципы РМД:

- представление данных в виде таблиц;
- использование стандартных операторов обработки данных, позволяющих генерировать новые таблицы на основе существующих.
- РМД характеризуется *структурой данных*, *целостностью* и *обработкой данных*.

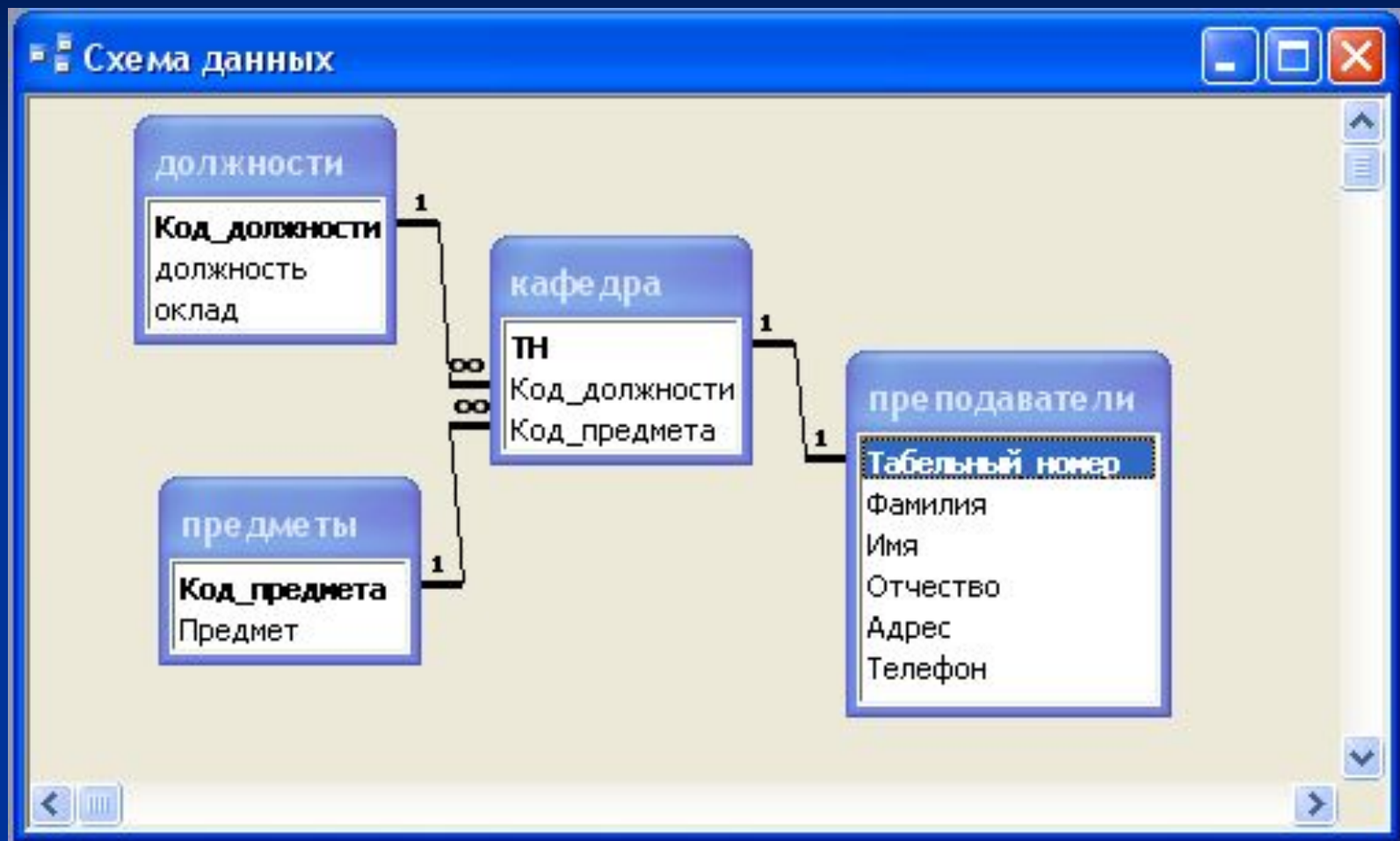
Основные понятия РМД

Структура данных связана с несколькими понятиями: *отношение* — это данные представленные в виде таблицы; таблица состоит из строк и столбцов. Строка таблицы называется *кортежем*, количество кортежей — *кардинальным числом*. Столбец — *атрибутом*, количество атрибутов — *степенью*. Каждая строка содержит *запись* БД. Каждая запись одной таблицы содержит равное количество полей. В таблице не должно быть одинаковых записей (кортежей). Каждая запись характеризуется уникальным идентификатором, который называется *первичный ключ*.

ВИДЫ ОТНОШЕНИЙ В РМД

- *Целостность данных* обеспечивается наличием ключей и связями, устанавливаемыми между таблицами.
- Связи между таблицами устанавливаются на базе следующих возможных типов отношений:
 1. ОДИН К ОДНОМУ (О-О);
 2. ОДИН КО МНОГИМ (О-М);
 3. МНОГО К ОДНОМУ (М-О);
 4. МНОГО КО МНОГИМ (М-М).

Виды отношений в РМД



Операции над данными

Обработка данных подразумевает совокупность возможных операций над БД.

Следует рассмотреть следующие виды операций:

- над строками;
- над таблицами (отношениями);
- специальные.

Операции над данными в РМД (на уровне строк)

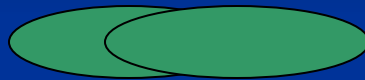
На уровне строк БД возможно их *удаление, обновление, добавление.*

Операции над данными в РМД (на уровне таблиц)

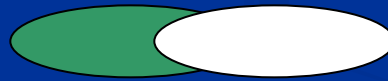
- На уровне таблиц – *объединение* таблиц, *пересечение*, *разность*, *декартово произведение* (над таблицами разной структуры – результат – все возможные сочетания – ненормализованная таблица с полной информацией об объекте).
Остальные операции производятся над таблицами одинаковой структуры.

Операции над данными в РМД (на уровне таблиц)

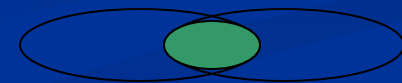
- **Объединение** – новая таблица (отношение) на базе двух существующих, включающая все кортежи, принадлежащие и первому отношению и второму.



- **Вычитание** – возвращает отношение, содержащее все кортежи, которые принадлежат первому и не принадлежат второму



- **Пересечение** – возвращает отношение, содержащее все кортежи, принадлежащие и первому и второму отношениям.



■ *Произведение декартово* – возвращает отношение, содержащее всевозможные кортежи, которые являются сочетанием двух кортежей, принадлежащих двум отношениям.

■ *Произведение декартово* – возвращает отношение, содержащее всевозможные кортежи, которые являются сочетанием двух кортежей, принадлежащих двум отношениям.

A
B
C

X
Y

результат:

A	X
A	Y
B	X
B	Y
C	X
C	Y

Операции над данными в РМД (специальные)

- Специальные операции – *выборка* данных, *проекция* и др.
- *Выборка* – возвращает отношение, содержащее все кортежи, удовлетворяющие определенным условиям.
- *Проекция* – возвращает отношение, содержащее все кортежи существующего отношения после исключения из него некоторых атрибутов.

Нормализация таблиц

При проектировании РБД необходимо обеспечить эффективную работу с данными, поэтому следует организовывать такую структуру, которая позволит:

- организовать быстрый доступ к данным;
- исключить ненужное повторение данных, которое может являться причиной ошибок при вводе и нерациональное использование дискового пространства;
- обеспечить целостность данных так, чтобы при изменении одних данных происходило автоматическое изменение связанных с ними данных.

Процесс уменьшения избыточности данных называется *нормализацией таблицы*.

Нормализация таблиц

- При существенном дублировании данных таблица называется **ненормализованной**. Из нее можно сформировать несколько нормализованных таблиц и установить между ними связи.
- Связи устанавливаются между **первичным ключом** и **совпадающими полями** другой таблицы. Первичный ключ однозначно определяет запись в таблице, таким ключом может быть код детали, порядковый номер в списке и т.д. **Внешний ключ** — это поле, содержащее такой же тип информации в таблице, которая является подчиненной.

Нормализация таблиц

Следует различать простое (неизбыточное) и избыточное дублирование данных. Наличие первого допускается в РБД. Второе может приводить к проблемам при обработке данных. Пример избыточного дублирования:

ненормализованная
таблица

сотрудник	телефон	комната	корпус
Иванов	2-22-22	105	2
Петров	4-44-44	107	1
Сидоров	4-44-44	107	1
Егоров	4-44-44	107	1

нормализованные
таблицы

сотрудник	комната
Иванов	105
Петров	107
Сидоров	107
Егоров	107

телефон	комната	корпус
2-22-22	105	2
4-44-44	107	1

Нормализация таблиц

Существует следующая последовательность нормальных форм:

- первая нормальная форма (1НФ);
- вторая нормальная форма (2НФ);
- третья нормальная форма (3НФ);
- нормальная форма Бойса-Кодда (БКНФ);
- четвертая нормальная форма (4НФ);
- пятая нормальная форма (5НФ)

и еще 2 формы сформулированы в последнее время.

Каждая форма – свод правил и ограничений, которые необходимо соблюдать при формировании таблиц.

Нормализация таблиц

Пример приведения таблицы к первой нормальной форме

Исходная, ненормализованная, таблица:

Сотрудник	Номер телефона
Иванов И. И.	283-56-82
	390-57-34
Петров П. Ю.	708-62-34

Таблица, приведённая к 1NF:

Сотрудник	Номер телефона
Иванов И. И.	283-56-82
Иванов И. И.	390-57-34
Петров П. Ю.	708-62-34

Нормализация таблиц

Пример приведения таблицы к третьей нормальной форме

Исходная таблица:

Фамилия	Отдел	Телефон
Гришин	1	11-22-33
Васильев	1	11-22-33
Петров	2	44-55-66

В результате приведения к ЗНФ получаются две таблицы:

Фамилия	Отдел
Гришин	1
Васильев	1
Петров	2

Отдел	Телефон
1	11-22-33
2	44-55-66

Достоинства и недостатки РМД

Достоинства РМД:

- простота представления данных (таблицы просты для восприятия информации);
- минимальная избыточность данных при нормализации таблиц;
- универсальность процедур обработки данных.

Недостатки РМД:

- фрагментация данных, из-за нормализации таблиц, а в большинстве задач требуется общая картина данных, но эта проблема может быть разрешена через различные средства СУБД – выборку данных в запросах и формирование отчетов на основе данных из различных связанных между собой таблиц.

СПАСИБО!

ВОПРОСЫ?