

# Реляционная модель данных

Наумов Д.А., доц. каф. КТ

Базы данных и базы знаний, 2020

# Содержание лекции

## 1 Реляционная модель данных

## 2 Другие подходы и модели данных

- Модель Сущность-Атрибут-Значение
- Неполно структурированные модели данных
- Документ-ориентированные модели данных

## Немного истории

- Эдгар Кодд, "Реляционная модель данных для больших совместно используемых банков данных"[1];
- Основа - теория отношений (XIX век);

### Эволюция концепции реляционной модели данных

- версия 1, опубликованная в статье в CACM в 1970 г.;
- версия 2, определена в статье по поводу Тьюринговской премии в 1981 г.;
- версия 3, доопределена "дюжиной Кодда"— двенадцатью правилами и оценочной системой в 1985 году [2,3];
- версия 4, определена в опубликованной в 1990 году книге Кодда [4].

# Немного истории

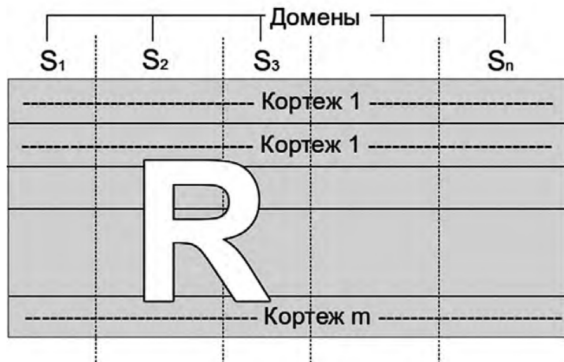
## Выводы

- Существуют два мира с условными названиями "реляционная модель" и "промышленные РСУБД" и эти два мира пересекаются, но не полностью, создавая, в частности, проблемы переносимости приложений между СУБД разных производителей.
- Реляционная модель относится к логическим, то есть определённым в рамках модели структуры, операции над ними и задаваемые ограничения не зависят от способов реализации физической организации данных и управления ими.
- Каждая СУБД имеет свои особенности физической организации, поэтому одна и та же схема данных в рамках реляционной модели может иметь специфические параметры и опции, оптимизирующие работу с данными.

# Определение

## Отношение

Для заданных множеств  $S_1, S_2, \dots, S_n$  (не обязательно различных)  $R$  является отношением на этих  $n$  множествах, если представляет собой множество кортежей степени  $n$ , у каждого из которых первый элемент взят из множества  $S_1$ , второй — из множества  $S_2$  и т.д.



# Определение

## Графический смысл отношения

таблица с зафиксированным типом данных для каждой колонки.

Реляционные СУБД предлагают своим пользователям оперировать не определениями реляционной алгебры, а их адаптированными аналогами (таблицы, строки, столбцы и т. п.).

Таблица (файл) →

Столбец (поле записи) →

НП	ФАМ.	ДОЛЖН.	СТАЖ
104	ИВАНОВ	ДОЦ.	10
108	ПЕТРОВ	СТ. ПРЕП.	5
103	ИЛЬИН	АСС.	2
101	ШАНЬГИН	ПРОФ.	36
102	КОСТИН	ПРОФ.	30

Строка (запись) →

Значение атрибута (значение поля в записи) →

## Свойства отношения

### Схема отношения

перечень атрибутов данного отношения с указанием домена, к которому они относятся.

$$S_R = (A_1, A_2, A_3, \dots, A_n), A_i \subset D_i$$

### Экземпляр отношения

отражает состояние данного объекта в текущий момент времени.

### Свойства отношения

- в таблице нет двух одинаковых строк;
- таблица имеет столбцы, соответствующие атрибутам отношения;
- каждый атрибут в отношении имеет уникальное имя;
- порядок строк в таблице произвольный.

## Первичный ключ отношения

$R1(\underline{\text{НП}}, \text{ФАМ.}, \text{ДОЛЖН.}, \text{СТАЖ})$        $R2(\underline{\text{НП}}, \underline{\text{НК}})$

Первичный ключ отношения (таблицы)

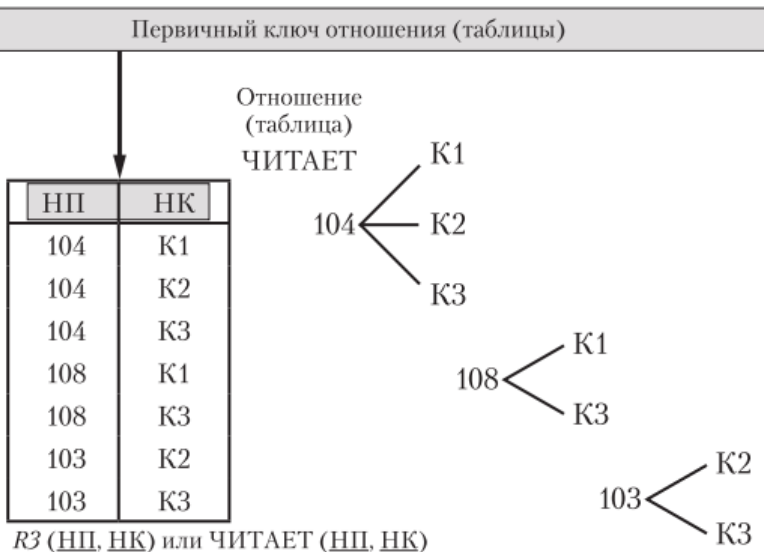
Отношение (таблица)  
ПРЕПОДАВАТЕЛЬ

НП	ФАМ.	ДОЛЖН.	СТАЖ
104	ИВАНОВ	ДОЦ.	10
108	ПЕТРОВ	СТ. ПРЕП.	5
103	ИЛЬИН	АСС.	2

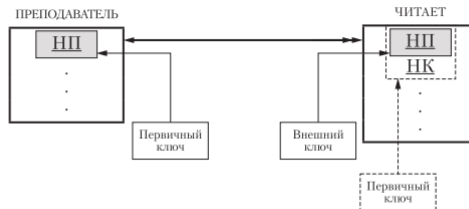
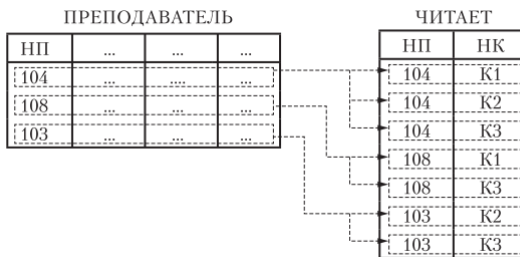
$R1(\underline{\text{НП}}, \text{ФАМ.}, \text{ДОЛЖН.}, \text{СТАЖ})$  или  
ПРЕПОДАВАТЕЛЬ ( $\underline{\text{НП}}, \text{ФАМ.}, \text{ДОЛЖН.}, \text{СТАЖ}$ )



# Первичный ключ отношения



# Внешний ключ отношения



# Реляционная алгебра

R

НП	ЗавК	Нтел
102	Шаньгин	2854
104	Вернер	2882
108	Вернер	2882
125	Шаньгин	2854

r

Отношение R (r – кортеж)

## Проекция отношения R на X

называют новое отношение  $R[X] = r[X] | r \in R$

 $R2 = R[\text{ЗавК}, \text{Нтел}]$ 

ЗавК	Нтел
Шаньгин	2854
Вернер	2882

 $R1 = R[\text{НП}, \text{ЗавК}]$ 

НП	ЗавК
102	Шаньгин
104	Вернер
108	Вернер
125	Шаньгин

Проекции  $R[\text{ЗавК}, \text{Нтел}]$  и  $R[\text{НП}, \text{ЗавК}]$

# Реляционная алгебра

R

НП	ЗавК	Нтел
102	Шаньгин	2854
104	Вернер	2882
108	Вернер	2882
125	Шаньгин	2854

r

Отношение R (r – кортеж)

Выборкой из отношения R по условию  $\Theta$  для множества атрибутов X называют новое отношение  $R[\Theta(X)] = \{r[X] \mid r \in R \wedge \Theta(r[X])\}$

$$R3 = R[\text{ЗавК} = \text{'Шаньгин'}]$$

НП	ЗавК	Нтел
102	Шаньгин	2854
125	Шаньгин	2854

Выборка

# Реляционная алгебра

Соединением отношения  $R$  с отношением  $S$  по условию  $X=Y$  называют новое отношение

$$R[X = Y]S = \langle r, s \rangle \mid r \in R \wedge s \in S \wedge (r[X] = s[Y])$$

 $R2 = R[\text{ЗавК}, \text{Нтел}]$ 

ЗавК	Нтел
Шаньгин	2854
Вернер	2882

 $R1 = R[\text{НП}, \text{ЗавК}]$ 

НП	ЗавК
102	Шаньгин
104	Вернер
108	Вернер
125	Шаньгин

НП	ЗавК	Нтел
102	Шаньгин	2854
104	Вернер	2882
108	Вернер	2882
125	Шаньгин	2854

# Реляционная алгебра

## Объединением отношения R с отношением S

называют новое отношение, объединяющее кортежи, содержащиеся в отношениях R и S.

$$R \cup S = \{x \mid x \in R \vee x \in S\}.$$

*R1*

НП	ЗавК
102	Шаньгин
104	Вернер
108	Вернер
125	Шаньгин

*R2*

ЗавК	Нтел
Шаньгин	2854
Вернер	2882

 $X = \{\text{ЗавК}\}$ 
 $Y = \{\text{ЗавК}\}$ 
 $R1 \mid \text{ЗавК} = \text{ЗавК} \mid R2$ 

<i>R</i>	НП	ЗавК	ЗавК	Нтел	Естественное соединение
	102	Шаньгин	Шаньгин	2854	
	104	Вернер	Вернер	2882	
	108	Вернер	Вернер	2882	
	125	Шаньгин	Шаньгин	2854	

# Реляционная алгебра

## Пересечением отношения R с отношением S

называют новое отношение, состоящее из общих кортежей

$$R \cap S = x | x \in R \wedge x \in S$$

## Разностью отношения R с отношением S

называют новое отношение, состоящее кортежей R, не являющиеся

кортежами S:  $R \setminus S = x | x \in R \wedge x \notin S$

# Реляционная алгебра

Произведением отношения  $R$  с отношением  $S$

называют новое отношение  $R \otimes S = \langle r, s \rangle \mid r \in R \wedge s \in S$

$R =$	<table><tr><td><math>G2</math></td></tr><tr><td><math>A</math></td></tr><tr><td><math>B</math></td></tr></table>	$G2$	$A$	$B$	$S =$	<table><tr><td><math>G1</math></td><td><math>F1</math></td></tr><tr><td><math>d</math></td><td><math>3</math></td></tr><tr><td><math>h</math></td><td><math>7</math></td></tr></table>	$G1$	$F1$	$d$	$3$	$h$	$7$						
	$G2$																	
	$A$																	
$B$																		
$G1$	$F1$																	
$d$	$3$																	
$h$	$7$																	
TO																		
$R \otimes S =$	<table><tr><td><math>G2</math></td><td><math>G1</math></td><td><math>F1</math></td></tr><tr><td><math>A</math></td><td><math>d</math></td><td><math>3</math></td></tr><tr><td><math>A</math></td><td><math>h</math></td><td><math>7</math></td></tr><tr><td><math>B</math></td><td><math>d</math></td><td><math>3</math></td></tr><tr><td><math>B</math></td><td><math>h</math></td><td><math>7</math></td></tr></table>	$G2$	$G1$	$F1$	$A$	$d$	$3$	$A$	$h$	$7$	$B$	$d$	$3$	$B$	$h$	$7$		
	$G2$	$G1$	$F1$															
	$A$	$d$	$3$															
	$A$	$h$	$7$															
$B$	$d$	$3$																
$B$	$h$	$7$																



## Преимущества реляционной модели данных

- теоретическая основа. Формально определяет базовые понятия модели, язык описания и операции над отношениями;
- стандартизация. Стандарты SQL-NN (SQL-89, SQL-92, SQL-99 и т. д.), имеющие несколько уровней полноты реализации, позволяют создавать приложения, переносимые между СУБД разных поставщиков;
- полное разделение доступа к данным от способа их физической организации;
- универсальность. Информационное моделирование сущностей реального мира в виде набора связанных таблиц является достаточно хорошим подходом в большинстве случаев;
- простота манипуляции данными с точки зрения конечного пользователя;
- SQL — развитый стандартизованный декларативный язык 4-го поколения.

## Недостатки реляционной модели данных

- в общем случае, более низкое быстродействие по сравнению с сетевыми и иерархическими СУБД или другими подходами, обеспечивающими доступ к данным непосредственно на уровне их физической организации, например, индексированные файлы;
- неполнота реализации стандартов SQL-NN, а также специфические языковые и процедурные расширения СУБД разных поставщиков, осложняющие переносимость приложений (так называемый vendor lock);
- необходимость учёта некоторых особенностей модели на концептуальном уровне (ключи — идентификаторы сущностей), отсутствующая, например, в сетевой модели.

## САЗ (Сущность-Атрибут-Значение, EAV (Entity-Attribute-Value))

подход, заключающийся в максимальном упрощении структуры хранения данных, за счет "вертикального хранения атрибутов что обеспечивает высокую гибкость изменения логической схемы данных.

Товар №795	Наименование	Мука пшеничная в/с
Товар №795	Вес упаковки	50
Клиент №123	Наименование	"Пирожки" ООО
Доставка №24	Дата создания	2014-06-15
Доставка №24	Статус	В пути
Продажа №867	Дата создания	2014-06-12
...	...	...

	Наименование	Вес упаковки	Дата создания	Статус
Товар №795	Мука пшеничная в/с	50		
Клиент №123	"Пирожки" ООО			
Доставка №24			2014-06-15	В пути
Продажа №867			2014-06-12	

САЗ даёт возможность менять логическую схему данных, что называется, "на лету". Действительно, достаточно добавить в метаданные описание нового атрибута и связать его с типами сущностей, как все приложения могут считывать и записывать значения данного атрибута, не принимая во внимание структурные изменения.

## Преимущества модели САЗ

- гибкость схемы данных и возможность её изменения в динамике;
- высокая степень унификации и повторного использования структур хранения;
- возможность выполнения контекстно-независимых запросов, например, не принимающих в расчёт тип сущностей;
- возможность реализации на основе реляционной СУБД;
- относительно простая базовая концепция.

## Недостатки модели САЗ

- сложность манипуляций данными вне CRUD-запросов;
- необходимость вводить собственный язык манипуляции данными;
- более низкая производительность за счёт большего числа соединений в запросах;
- необходимость оптимизации физического хранения.

## Неполно структурированные модели данных

(НСМД, semi-structured data models) - способы организации данных, в которых схема и данные не разделены.

Случаи применения:

- исходные данные неполно структурированы, а возможности по их анализу и классификации ограничены;
- обмен информацией в гетерогенной среде.

```
<!-- Случай 1. Заказы списком -->
<orders>
  <order>
    <number>123</number>
    <items />
  </order>
</orders>
...
<!-- Случай 2. Заказы в составе продаж -->
<sales>
  <sale>
    <order>
      <number>123</number>
    </order>
  </sale>
</sales>
```

## JSON

JavaScript Object Notation, формат хранения НСМД, менее удобный для восприятия человеком, но более прост для обработки анализаторами и более экономичный.

```
{
  "номер_заказа": "123-45",
  "дата": "2014-02-15",
  "состояние": "выполнен",
  "адрес_доставки": {
    "улица": "Ивановский пр. 5",
    "город": "Крыжополь",
    "индекс": "333222"
  },
  "телефоны": [
    { "тип": "домашний", "номер": "921-1112233" },
    { "тип": "мобильный", "номер": "921-1114455" }
  ]
}
```

Концепция самодокументированных данных: информация передаётся между системами в НСМД, при этом структура считается самодостаточной для распознавания и интерпретации смысловых и контекстных связей в документе без предварительного согласования

## В документ-ориентированной модели данных (ДОМ)

основным элементом являются документы, организованные в линейные списки — коллекции, которые могут содержать другие коллекции.

- В отсутствии строгого определения, понятие "документ" соответствует иерархическим XML, JSON-документам и подобным структурам.
- ДОМ представляет собой синтез иерархической и неполно-структурированной моделей данных.

СУБД, реализующие документ-ориентированную модель, получили название NoSQL.

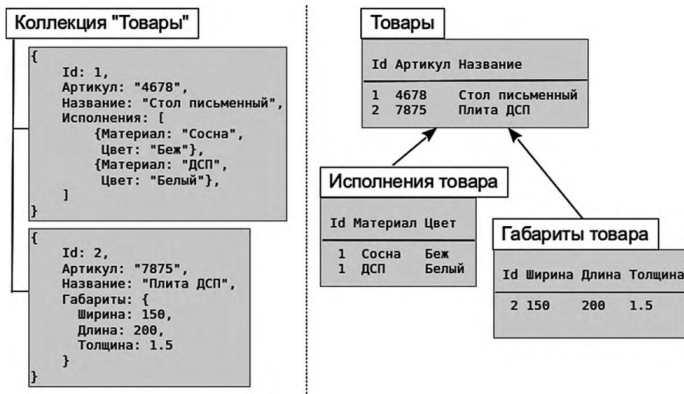
## Реализации NoSQL СУБД

Cassandra, CouchDB, MongoDB.

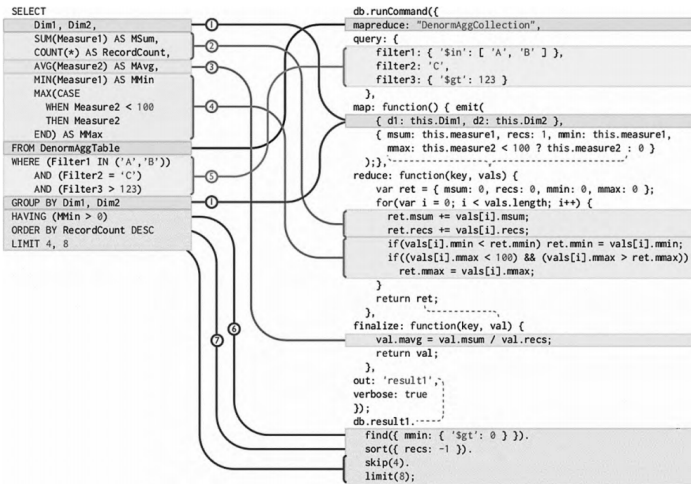


## Отличия от РМД и сетевой МД

В отличие от реляционной модели, где каждая строка таблицы имеет строго оговорённое число полей, определяемых заданными колонками и их типами, или от канонической иерархической модели, где каждый сегмент содержит список записей одного типа, документы списка в ДОМ могут иметь структуру с общей и различающимися частями.



Слабым звеном NoSQL является отсутствие стандартизованного декларативного языка запросов.



- Существует класс задач сбора, хранения и первичной обработки неполно структурированной информации, где продукты данного типа могут быть рассмотрены в качестве альтернативы реляционным СУБД, не имеющих соответствующих расширений для работы с XML или JSON.
- Некомпетентность в SQL и реляционных СУБД не может служить серьёзным доводом в выборе NoSQL СУБД.

## Рекомендуемая литература:

- ① Эдгар Кодд, Реляционная модель данных для больших совместно используемых банков данных. Перевод Communications of the ACM, Volume 13, Number 6, June, 1970. — Журнал "СУБД №1/1995;
- ② Codd, E.F. Is Your DBMS Really Relational?; Computerworld, October 14th, 1985.
- ③ Codd, E.F. Does Your DBMS Run By The Rules?; Computerworld, October 21st, 1985.
- ④ Codd, E.F. The Relational Model For Database Management Version 2. Reading, Mass.: Addison-Wesley, 1990.