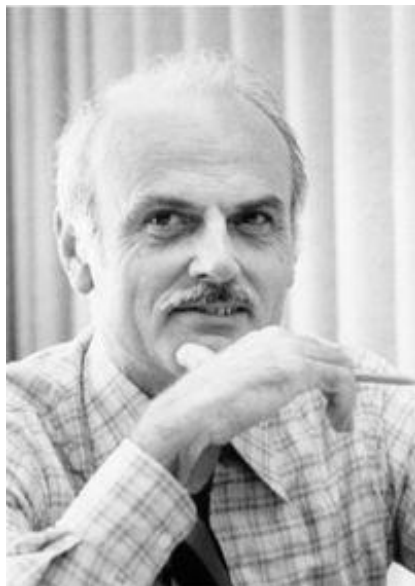


2. Реляционная модель данных. Нормальные формы.

Лучинин Захар
zakhar.luchinin@gmail.com

Реляционные базы данных



Эдгар Франк Кодд
23.08.1923 — 18.04.2003

- Особенности реляционной базы данных:
 - Данные хранятся в таблицах, состоящих из столбцов и строк;
 - На пересечении каждого столбца и строчки стоит в точности одно значение;
 - У каждого столбца есть своё имя, которое служит его названием, и все значения в одном столбце имеют один тип.
 - Столбцы располагаются в определённом порядке, который определяется при создании таблицы, в отличие от строк, которые располагаются в произвольном порядке. В таблице может не быть ни одной строчки, но обязательно должен быть хотя бы один столбец;
 - Запросы к базе данных возвращают результат в виде таблиц, которые тоже могут выступать как объект запросов.

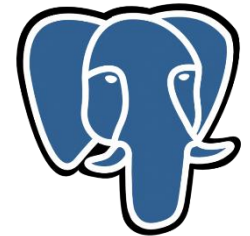
Реляционная модель данных

Реляционная модель данных - разработанная Э.Коддом в 1970 г. логическая модель данных, описывает:

- структуры данных в виде (изменяющихся во времени) наборов отношений; в теории множеств таблице соответствует термин отношение (relation).
- теоретико-множественные операции над данными: объединение, пересечение, разность и декартово произведение;
- специальные реляционные операции: селекция, проекция, соединение и деление;
- специальные правила, обеспечивающие целостность данных.

Примеры Реляционных СУБД

- MySQL
- PostgreSQL
- Microsoft Access
- Microsoft SQL Server
- Oracle



PostgreSQL



ORACLE®
D A T A B A S E

Понятия реляционной модели данных

Отношение	целое	строка		целое		Типы данных
	номер	имя	должность	деньги		Домены
	Табельный номер	Имя	Должность	Оклад	Премия	Атрибуты
	2934 2935 2936	Иванов Петров Сидоров	инженер вед. инженер бухгалтер	112 144 92	40 50 35	Кортежи

↑
Ключ

Ключи и ключевые атрибуты

- **Ключ отношения** – атрибут, или набор атрибутов, однозначно идентифицирующий объект в БД
- **Первичный ключ** – ключ, используемый для идентификации объекта

Персоны

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80

Ключи и ключевые атрибуты

- **Ключ отношения** – атрибут, или набор атрибутов, однозначно идентифицирующий объект в БД
- **Первичный ключ** – ключ, используемый системой для идентификации объекта
- **Суррогатный ключ** – ключ, значения которого генерируются самой системой

Персоны

ID	Имя	Возраст	Вес
1	Harry	34	80
2	Donald	29	70
3	Helena	54	54
4	Peter	34	80

Ключи и ключевые атрибуты

- **Ключ-кандидат** – ключ, по каким либо причинам неиспользуемый как первичный
- **Составной ключ** – ключ, использующий несколько атрибутов

Персоны

ID	Паспорт	Имя	Возраст	Вес
1	8809 643496	Harry	34	80
2	8808 423386	Donald	29	70
3	8810 833946	Helena	54	54
4	8808 812937	Peter	34	80

Хорошая или плохая схема?

Предметная область: необходимо хранить сотрудников университета, конференции и посещаемые конференции.

- Других таблиц нет.
- Исследователей с одинаковым именем не бывает.

Университет	Исследователь	Конференция	Город
ПГТУ	Иванов	Highload'16	Москва
МГУ	Петров	Highload'16	Москва
ИТМО	Сидоров	РИТ++'17	Санкт-Петербург
ПГТУ	Васенев	РИТ++'17	Санкт-Петербург

Удаление информации

Университет	Исследователь	Конференция	Город
ПГТУ	Иванов	Highload'16	Москва
МГУ	Петров	Highload'16	Москва
ИТМО	Сидоров	РИТ++'17	Санкт-Петербург
ПГТУ	Васенев	РИТ++'17	Санкт-Петербург
СПбГУ	Андреев	Blockchain Conf'18	Казань

Аномалии-удаления — при удалении какого либо кортежа из таблицы может пропасть информация, которая не связана на прямую с удаляемой записью.

- Где будет Blockchain Conf'18?
- В каком университете работает Андреев?

Добавление информации

Университет	Исследователь	Конференция	Город
ПГТУ	Иванов	Highload'16	Москва
МГУ	Петров	Highload'16	Москва
ИТМО	Сидоров	РИТ++'17	Санкт-Петербург
СПбГУ	Андреев	?	?
ПГТУ	Васенев	РИТ++'17	Санкт-Петербург
?	?	Blockchain Conf'18	Казань

Аномалии-добавления возникают, когда информацию в таблицу нельзя поместить, пока она не полная, либо вставка записи требует дополнительного просмотра таблицы.

Обновление информации

Университет	Исследователь	Конференция	Город
ПГТУ	Иванов	Highload'16	Москва
МГУ	Петров	Highload'16	Казань
ИТМО	Сидоров	РИТ++'17	Санкт-Петербург
ПГТУ	Васенев	РИТ++'17	Санкт-Петербург

Аномалии-модификации проявляются в том, что изменение данных может повлечь к потере достоверности данных.

Устранение избыточности

Устранение избыточности производится за счёт декомпозиции отношений таким образом, чтобы в каждом отношении хранились только первичные факты (то есть факты, не выводимые из других хранимых фактов).

Университет	Исследователь	Конференция	Город
ПГТУ	Иванов	Highload'16	Казань
МГУ	Петров	Highload'16	Казань
ИТМО	Сидоров	РИТ++'17	Санкт-Петербург
ПГТУ	Васенев	РИТ++'17	Санкт-Петербург

Устранение избыточности

Университет	Исследователь
ПГТУ	Иванов
МГУ	Петров
ИТМО	Сидоров
ПГТУ	Васенев

Конференция	Город
Highload'16	Казань
РИТ++'17	Санкт-Петербург

Исследователь	Конференция
Иванов	Highload'16
Петров	Highload'16
Сидоров	РИТ++'17
Васенев	РИТ++'17

Нормальная форма

Нормальная форма — требование, предъявляемое к структуре таблиц в теории реляционных баз данных для устранения избыточных зависимостей между атрибутами.

Нормализация

Процесс приведения отношений к нормальным формам с целью :

- устранение избыточности
- устранение аномалий
- упрощение процедуры применения необходимых ограничений целостности

Цель нормализации

Процесс проектирования правильной базы данных – это не процесс приведения ее к самой высокой нормальной форме, это компромисс между отсутствием аномалий и приемлемой производительностью.

Первая нормальная форма(1NF)

Таблица находится в первой нормальной форме (1NF) только в том случае, когда каждая ее запись содержит только **одно значение для каждого из атрибутов**

Первая нормальная форма

- Исходная ненормализованная таблица

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

- Таблица, приведённая к 1НФ

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

Вторая нормальная форма(2NF)

Таблица находится во второй нормальной форме, если она находится в первой нормальной форме, и при этом любой её атрибут, не входящий в состав первичного ключа, функционально полно зависит от первичного ключа

Вторая нормальная форма

Пусть в следующем отношении первичный ключ образует пара атрибутов {Филиал компании, Должность}

Исходная ненормализованная таблица

<u>Филиал компании</u>	<u>Должность</u>	Зарплата	Наличие компьютера
Филиал в Томске	Уборщик	20000	Нет
Филиал в Москве	Программист	40000	Есть
Филиал в Томске	Программист	25000	Есть

Функциональные зависимости:

- *Филиал компании, Должность* → *Зарплата*
- *Должность* → *Наличие компьютера*

Если изменение содержимого одного столбца должно приводить к изменению другого, говорят что второй столбец **функционально зависим** от первого.

Таблицы, приведённые к 2НФ

R1

<u>Филиал компании</u>	<u>Должность</u>	Зарплата
Филиал в Томске	Уборщик	20000
Филиал в Томске	Программист	25000
Филиал в Москве	Программист	40000

R2

<u>Должность</u>	Наличие компьютера
Уборщик	Нет
Программист	Есть

Третья нормальная форма (3NF)

Таблица находится в третьей нормальной форме, если она находится во второй нормальной форме, и при этом любой её **неключевой атрибут зависит только от первичного ключа.**

Третья нормальная форма

- Каждый сотрудник относится исключительно к одному отделу
- Каждый отдел имеет единственный телефон
- Атрибут Сотрудник является первичным ключом
- Личных телефонов у сотрудников нет, и телефон сотрудника зависит исключительно от отдела

Исходная ненормализованная таблица R1

<u>Сотрудник</u>	Отдел	Телефон
Гришин	Бухгалтерия	11-22-33
Васильев	Бухгалтерия	11-22-33
Петров	Снабжение	44-55-66

Функциональные зависимости:

- *Сотрудник* → *Отдел*
- *Отдел* → *Телефон*
- *Сотрудник* → *Телефон*

Транзитивная зависимость

- *Сотрудник* → *Телефон*

Транзитивная зависимость

Таблицы, приведённые к 3НФ

R1

<u>Отдел</u>	Телефон
Бухгалтерия	11-22-33
Снабжение	44-55-66

R2

<u>Сотрудник</u>	Отдел
Гришин	Бухгалтерия
Васильев	Бухгалтерия
Петров	Снабжение

Нормальная форма Бойса — Кодда(BCNF)

Часть составного первичного ключа не должна зависеть от неключевого столбца.

Нормальная форма Бойса — Кодда

Исходная ненормализованная таблица

<u>Проект</u>	<u>Направление</u>	Куратор
1	Разработка	Иванов И.И.
1	Бухгалтерия	Сергеев С.С.
2	Разработка	Иванов И.И.
2	Бухгалтерия	Петров П.П.
2	Реализация	John Smith
3	Разработка	Андреев А.А.

- **Сотрудник может быть куратором только того направления, на котором он специализируется**
- Зная куратора, мы можем четко определить, какое направление он курирует. Иными словами, часть составного ключа, т.е. «Направление», зависит от неключевого атрибута, т.е. «Куратора».

Таблицы, приведённые к BCNF

Таблица кураторов R2

<u>ID куратора</u>	ФИО	Направление
1	Иванов И.И.	Разработка
2	Сергеев С.С.	Бухгалтерия
3	Петров П.П.	Бухгалтерия
4	John Smith	Реализация
5	Андреев А.А.	Разработка

Таблица связи кураторов и проектов R3

<u>Проект</u>	<u>Идентификатор куратора</u>
1	1
1	2
2	1
2	3
2	4
3	5

Четвертая нормальная форма

Переменная отношения R находится в четвёртой нормальной форме, если она находится в НФБК и все нетривиальные **многозначные зависимости** фактически являются функциональными зависимостями от её потенциальных ключей.

Четвертая нормальная форма

Исходная ненормализованная таблица R1

<u>Курс</u>	<u>Преподаватель</u>	<u>Аудитория</u>
SQL	Иванов И.И.	101
SQL	Иванов И.И.	203
SQL	Сергеев С.С.	305
SQL	Сергеев С.С.	407
Python	John Smith	502
Python	John Smith	305

Многозначная зависимость

Курс ->-> Преподаватель

Курс ->-> Аудитория

Т.е. для каждого курса в этой таблице может быть несколько преподавателей, а также несколько аудиторий.

Таблицы, приведённые к 4НФ

Связь курсов и преподавателей R2

<u>Курс</u>	<u>Преподаватель</u>
SQL	Иванов И.И.
SQL	Сергеев С.С.
Python	John Smith

Связь курсов и аудиторий R3

<u>Курс</u>	<u>Аудитория</u>
SQL	101
SQL	203
SQL	305
SQL	407
Python	502
Python	305

Пятая нормальная форма

Требование пятой нормальной формы (5NF) заключается в том, чтобы в таблице каждая нетривиальная **зависимость соединения** определялась потенциальным ключом этой таблицы.

Декомпозиция без потерь – процесс разбиения одной таблицы на несколько, при условии, что в случае соединения таблиц, которые были получены в результате декомпозиции, будет формироваться ровно та же самая информация, что и в исходной таблице до декомпозиции.

Пятая нормальная форма

Связь сотрудников с проектами и направлениями работы в проектах.

Исходная ненормализованная таблица R1

<u>Сотрудник</u>	<u>Проект</u>	<u>Направление</u>
Иванов И.И.	Интернет магазин	Разработка
Сергеев С.С.	Интернет магазин	Бухгалтерия
Сергеев С.С.	Новый офис	Реализация
John Smith	Личный кабинет	Бухгалтерия
Иванов И.И.	Личный кабинет	Разработка
Иванов И.И.	Информационная система	Разработка

Пятая нормальная форма

Необходимо разобраться в предметной области и определить ограничения

Поработав с предметной областью, мы выясняем, что:

- Иванов И.И. может работать только в направлении «Разработка»
- Сергеев С.С. может работать в любом направлении, за исключением «Разработка»
- Иванов И.И. может участвовать в большом количестве проектов
- John Smith может участвовать только в одном проекте

Таблица находится в четвертой нормальной форме, так как у нас отсутствует многозначная зависимость, ведь у нас нет таких атрибутов, которые зависели бы от другого атрибута.

Пятая нормальная форма

Связь сотрудников и проектов R2

Сотрудник	Проект
Иванов И.И.	Интернет магазин
....	...

Связь сотрудников и направлений R3

Сотрудник	Направление
Иванов И.И.	Разработка
...	...

Связь проектов и направлений R4

Проект	Направление
Интернет магазин	Разработка
...	...

Доменно-ключевая нормальная форма

Каждое наложенное ограничение на таблицу являлось логическим следствием ограничений доменов и ограничений ключей, которые накладываются на данную таблицу.

Ограничение домена – это ограничение, предписывающее использование для определенного атрибута значений только из некоторого заданного домена (набора значений).

Ограничение ключа – это ограничение, утверждающее, что некоторый атрибут или комбинация атрибутов представляет собой потенциальный ключ.

Доменно-ключевая нормальная форма

Исходная ненормализованная таблица R1

<u>Wealthy Person</u>	Wealthy Person Type	Net Worth in Dollars
Steve	Millionaire	124,543,621
Roderick	Billionaire	6,553,228,893
Katrina	Billionaire	8,829,462,998
Gary	Millionaire	495,565,211

Таблицы, приведенные к ДКНФ

<u>Wealthy Person</u>	Net Worth in Dollars
Steve	124,543,621
Roderick	6,553,228,893
Katrina	8,829,462,998
Gary	495,565,211

<u>Status</u>	Minimum	Maximum
Millionaire	1,000,000	999,999,999
Billionaire	1,000,000,000	999,999,999,999

Шестая нормальная форма

Таблица должна удовлетворять всем нетривиальным зависимостям соединения.

Работники

<u>Таб. №</u>	<u>Время</u>	<u>Должность</u>	<u>Домашний адрес</u>
6575	[01-01-2000:10-02-2003]	слесарь	ул. Ленина, 10
6575	[11-02-2003:15-06-2006]	слесарь	ул. Советская, 22
6575	[16-06-2006:05-03-2009]	бригадир	ул. Советская, 22

Должности работников

<u>Таб. №</u>	<u>Время</u>	<u>Должность</u>
6575	[01-01-2000:15-06-2006]	слесарь
6575	[16-06-2006:05-03-2009]	бригадир

Домашние адреса работников

<u>Таб. №</u>	<u>Время</u>	<u>Домашний адрес</u>
6575	[01-01-2000:10-02-2003]	ул. Ленина, 10
6575	[11-02-2003:15-06-2006]	ул. Советская, 22

Результаты нормализации

- Уменьшение избыточности данных (занимает меньше места)
- Увеличение количества таблиц
- Потеря в скорости выполнения запросов
- Легче обозначить связи между сущностями и наложить ограничения бизнес-логики.

РБД. Преимущества

- Теоретическим базисом реляционного подхода к организации баз данных служит простой и мощный математический аппарат теории множеств и математической логики
- Изложение информации в простой и понятной для пользователя форме (таблица)
- Контроль целостности данных
- Гибкость: выборка, изменение схемы данных
- Декларативные языки определения и манипулирования данными

РБД. Недостатки

- Относительно медленный доступ к данным.
- Большое количество таблиц в реальных БД.
- Некоторые предметные области плохо представляются в форме отношений.
- Несоответствие классов ПО к отношениям БД

ER диаграмма

ENTITY-RELATIONSHIP диаграмма – это диаграмма, которая отображает отношения набора сущностей, хранящиеся в базе данных.

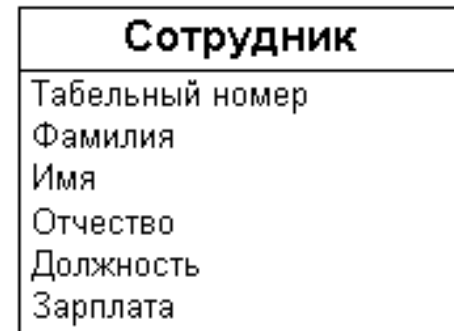
В основе ER-диаграмм лежит принцип «рисунок нагляднее текста»

Составные части ER диаграммы

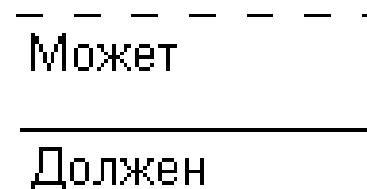
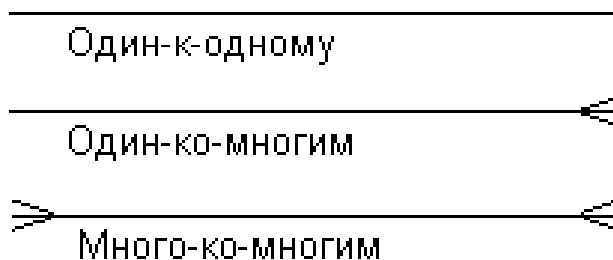
- **Сущность** - это класс однотипных объектов, информация о которых должна быть учтена в модели.



- **Атрибут сущности** - это именованная характеристика, являющаяся некоторым свойством сущности.



- **Ключ сущности** - это избыточный набор атрибутов, значения которых в совокупности являются уникальными для каждого экземпляра сущности. Избыточность заключается в том, что удаление любого атрибута из ключа нарушается его уникальность.
- **Связь** - это некоторая ассоциация между двумя сущностями. Одна сущность может быть связана с другой сущностью или сама с собою.



ER диаграмма логического уровня проектирования

