# CTECH.DB.02

# ОТНОШЕНИЯ. КЛЮЧИ. СВЯЗИ.

Базовые понятия реляционных БД: отношения, ключи и их виды, связи и их виды. Индексы. Целостность и консистентность данных, триггеры.

**Author: Svyatoslav Kulikov** 

Training And Education Manager

svyatoslav\_kulikov@epam.com

#### Содержание

- 1. Отношения.
- 2. Ключи.
- 3. Индексы.
- 4. Связи.
- 5. Ссылочная целостность данных.
- 6. Триггеры.

# **ОТНОШЕНИЯ**

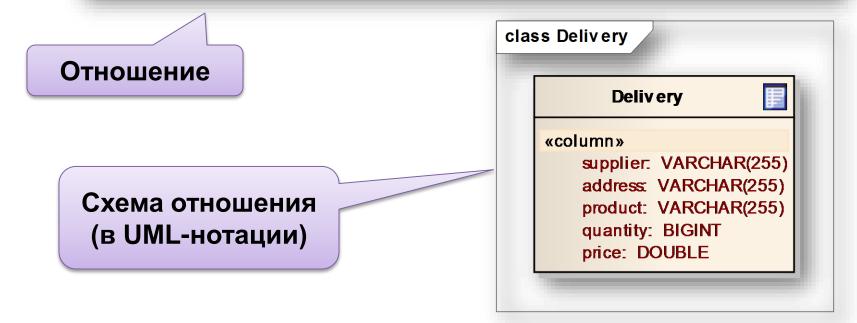
### Определение (сначала научно)

**Отношение, сущность** (relation, entity) R степени n – подмножество декартового произведения множеств  $D_1, D_2, ..., D_n$  (n≥1). Исходные множества  $D_1, D_2, ..., D_n$  называются **доменами (в СУБД** – **«тип данных»)**.

	D <sub>1</sub>	$D_2$	D <sub></sub>		
	supplier	address	product	quantity	price
Наб. зн. 1	"Рога и копыта" Ltd	Далёкое-далёко, 999	Веники	200	1500
Наб. зн. 2	"Рога и копыта" Ltd	Далёкое-далёко, 999	Ложки	100	500
Наб. зн	"Рога и копыта" Ltd	Далёкое-далёко, 999	Вилки	100	500
	"Vasya and partners" Inc.	Близкое-близко, 111	Кирпичи	100000	3000000
	"Vasya and partners" Inc.	Близкое-близко, 111	Корм для канареек	20	12345

### Определение (теперь «на пальцах»)

supplier	address	product	quantity	price
"Рога и копыта" Ltd	Далёкое-далёко, 999	Веники	200	1500
"Рога и копыта" Ltd	Далёкое-далёко, 999	Ложки	100	500
"Рога и копыта" Ltd	Далёкое-далёко, 999	Вилки	100	500
"Vasya and partners" Inc.	Близкое-близко, 111	Кирпичи	100000	3000000
"Vasya and partners" Inc.	Близкое-близко, 111	Корм для канареек	20	12345



# Определение (теперь «на пальцах»)

Заголовок (схема) отношения

Значение отношения Атрибуты отношения («столбцы таблицы»)

Мощность этношения (число «строк таблицы»)

	supplier	address	product	quantity	price
	"Рога и копыта" Ltd	Далёкое-далёко, 999	Веники	200	1500
l	"Рога и копыта" Ltd	Далёкое-далёко, 999	Ложки	100	500
	"Рога и копыта" Ltd	Далёкое-далёко, 999	Вилки	100	500
	"Vasya and partners" Inc.	Близкое-близко, 111	Кирпичи	100000	3000000
	"Vasya and partners" Inc.	Близкое-близко, 111	корм для канареек	20	12345

Тело отношения (кортежи НЕ упорядочены, порядок задаётся операторами сортировки)

Кортеж отношения («строка таблицы», «запись таблицы»)

#### «В жизни всё не так, как на самом деле»

Теория реляцис может считатьс

В реальности в таблицу БД можно поместить 2+ одинаковые строки. Этого не делают, но всё же...

ца

- В таблице нет двух одинаковых строк.
- У таблицы есть столбцы, соответствующие атрибутам отношения.

OK

- Каждый атрибут-столбец имеет уникальное имя.
- Порядок строк в таб те произвольный.

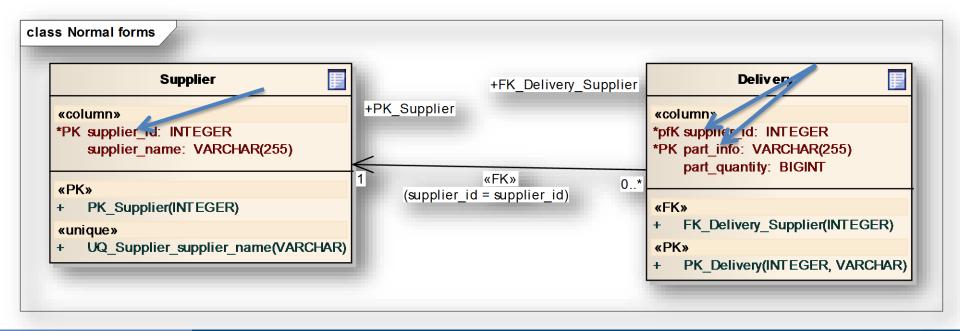
Например, JOIN'ы в MySQL запросто возвращают таблицы с 2+ одинаковыми именами.

OK

# **КЛЮЧИ**

#### Ключ

Ключ (key) – атрибут (или совокупность атрибутов) отношения, обладающий некоторыми специфическими свойствами, зависящими от вида ключа.



# Первичный ключ

Первичный ключ (primary key, PK) – минимальное множество атрибутов, являющееся подмножеством заголовка данного отношения, составное значение которых уникально определяет кортеж

отношения.

class Normal forms

«column»

«PK»

Log

event date: DATE

PK Log(BIGINT)

event info: VARCHAR(50)

\*PK event id: BIGINT

## Первичный ключ – ВАЖНО!

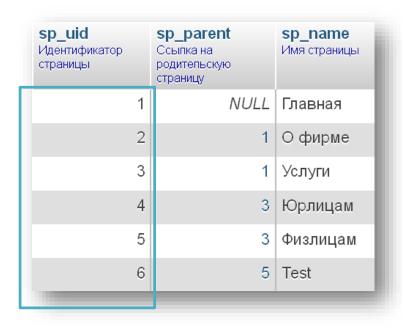
# Значение первичного ключа не может повторяться в двух и более строках.

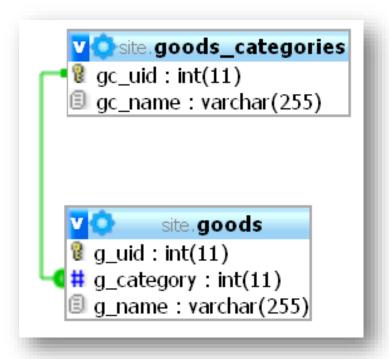


### Первичный ключ – ВАЖНО!

# Первичный ключ служит для:

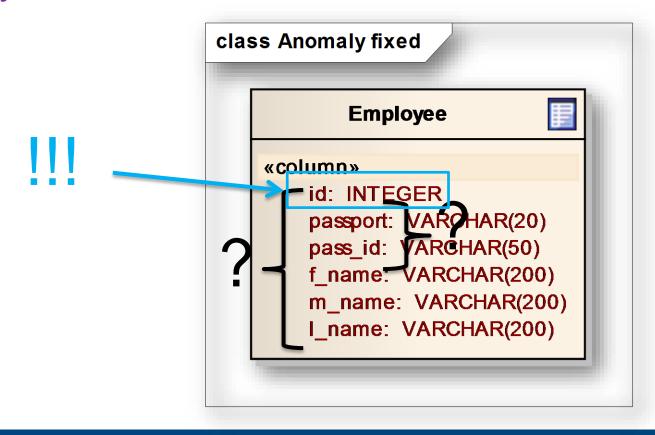
- однозначной идентификации строки таблицы;
- организации связей между таблицами.





## Первичный ключ – ВАЖНО!

Из всех вариантов нужно выбирать в качестве первичного ключа самое «короткое» поле или самую «короткую» комбинацию полей.



# Простой первичный ключ

Простой первичный ключ (simple primary key) – первичный ключ, состоящий из единственного поля таблицы (атрибута отношения).

passport	fml_name
MA334455	Сидоров С.С.
MM991122	Сидоров С.С.
MM998877	Петров П.П.
MP112233	Иванов И.И.
MP334455	Иванов И.И.

# Составной первичный ключ

Составной первичный ключ (compound PK, composite PK, concatenated PK) – первичный ключ, состоящий из нескольких полей таблицы.

passportcountryfml\_nameMA998877BYСидоров С.С.MM112233BYИванов И.И.ММ112233RUПетров П.П.

## Составной первичный ключ – последовательность полей

Последовательность полей в составных РК имеет значение: СУБД может проводить поиск очень быстро целиком по всему составному РК или по первому полю, входящему в его состав.

Если поиск производится чаще по:

- номеру паспорта => {passport, country}
- стране => {country, passport}

<b>V</b>	<b>V</b>	
passport	country	fml_name
MA998877	BY	Сидоров С.С.
MM112233	BY	Иванов И.И.
MM112233	RU	Петров П.П.

# Естественный первичный ключ

**Естественный первичный ключ** (natural PK) – поле таблицы, хранящее полезные данные.

passport	fml_name
MA334455	Сидоров С.С.
MM991122	Сидоров С.С.
MM998877	Петров П.П.
MP112233	Иванов И.И.
MP334455	Иванов И.И.

# Естественный первичный ключ

У естественных первичных ключей есть недостатки:

- Большой размер.
- Необходимость каскадных изменений.
- Невозможность вставки данных при отсутствии части информации.





fml\_name

passport

passport	fml_name	passport	money
MA334455	Сидоров С.С.	MA334455	999
MM991122	Сидоров С.С.	MA334455	5754
		MA334455	6433634
		MM991122	454534
м, паспорт	га нет.		

fml\_name

passport

Какой-то мужик с бодуна, готов работать двоником, паспорта нет.

# Синтетический (суррогатный) первичный ключ

Синтетический (суррогатный) первичный ключ (synthetic, surrogate PK) – искусственно добавленное в таблицу поле, единственная задача которого – быть первичным ключом.

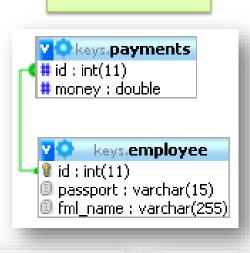
id 🖖		passport	fml_name
	1	MP334455	Иванов И.И.
	2	MP112233	Иванов И.И.
	3	MM998877	Петров П.П.
2	4	MM991122	Сидоров С.С.
į	5	MA334455	Сидоров С.С.
(	3		Какой-то мужик с бодуна,

# Синтетический (суррогатный) первичный ключ

У синтетических ключей есть преимущества:

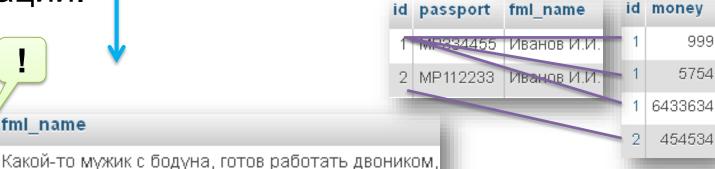
- Малый размер.
- Нет необходимости каскадных изменений.
- Можно вставить данные при отсутствии части информации.

fml name



4 байта

id



passport

id

6

#### Возможные и альтернативные ключи

Возможный ключ (possible key) – поле или попой с ушикальными COBOKYPHOCTL На альтернативных значен ичные ключи. ключах, как правило, Альте nate key) – поле создают уникальные индексы. Но об этом чуть или со Ікальными позже. качестве значен первичного ключа. Возможные ключи Если это license plate vin info – PK, ... 1122 EH-7 LICECBL CX11000237 Машинко 2233 AA-9 ZZCPCBLCX11999111 Тоже машинко ... то это – альтернативный ключ.

## Интеллектуальный первичный ключ

# Интеллектуальный первичный ключ

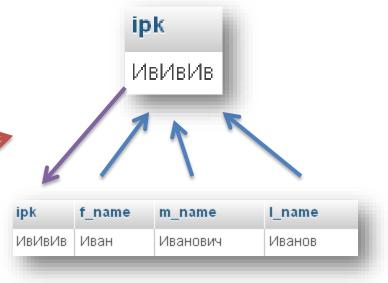
(intelligent primary key) – поле, значение которого формируется на основе значений других полей.

ipk	f_name	m_name	I_name
ИвИвИв	Иван	Иванович	Иванов
СиСиСи	Сидор	Сидорович	Сидоров
ПёПеПе	Пётр	Петрович	Петров
АлАлАл	Александр	Александрович	Александров

# Интеллектуальный первичный ключ

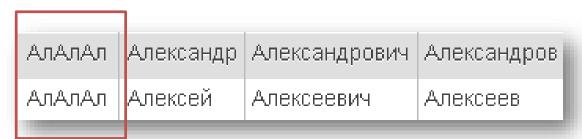
С интеллектуальными ключами есть **проблемы**:

 Их нужно генерировать и обновлять.



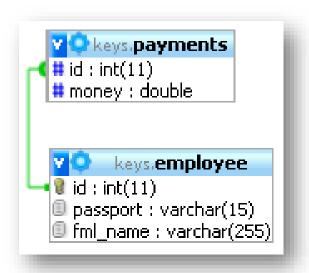
• Высока вероятность коллизий (совпадений значений). /

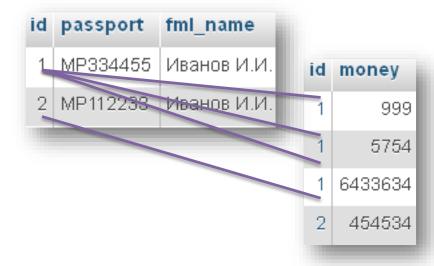
Поэтому они почти не используются.



#### Внешний ключ

Внешний ключ (foreign key) – поле таблицы, предназначенное для хранения значения первичного ключа другой таблицы с целью организации связи между этими таблицами.

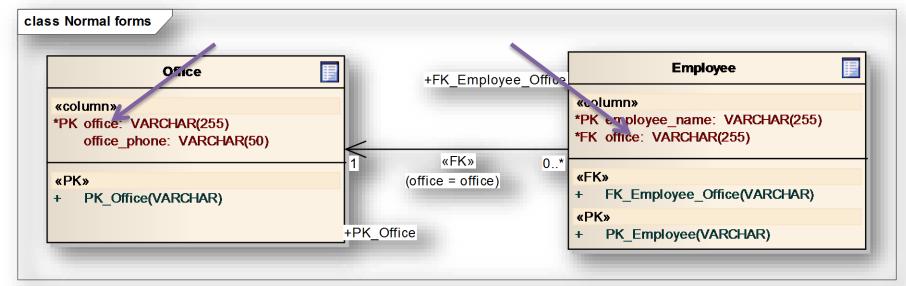




## Внешний ключ - как происходит миграция

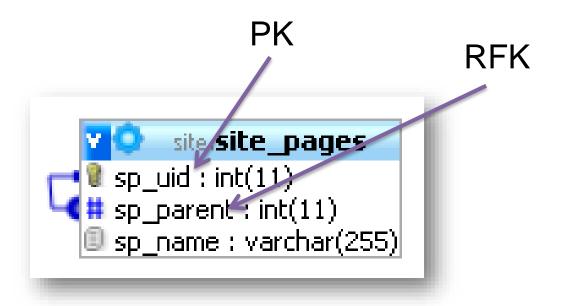
Миграция первичного ключа из родительской таблицы в дочернюю либо производится:

- автоматически (ErWin);
- вручную (Sparx EA) создать в дочерней таблице поле такого же типа (имя можно брать другое!!!), что и мигрирующий РК (включая регистрочувствительность, «знаковость» и т.п.), и указать его в качестве внешнего ключа для проведённой между таблицами связи.



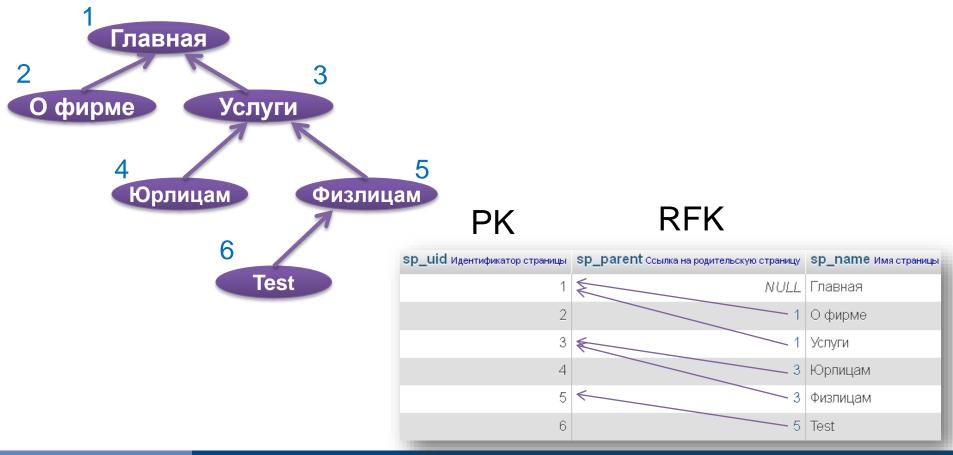
## Рекурсивный внешний ключ

Рекурсивный внешний ключ (recursive foreign key) – внешний ключ, полученный из поля этой же таблицы.



# Рекурсивный внешний ключ

Классика применения RFK – хранение древовидных структур (например, структуры сайта).

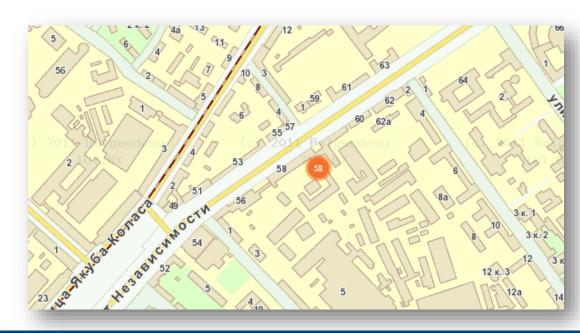


# **ИНДЕКСЫ**

#### Понятие индекса

**Индекс** (index) – объект БД, создаваемый с целью повышения производительности поиска данных.

Простая аналогия «из жизни» — карта города. Можно посмотреть по карте, где находится искомый объект, а не бродить в поисках по всему городу.



#### Преимущества индексов

#### **SHOW TABLE STATUS FROM 'dbname'**

**B-Tree** 

(сбалансированное

Размер индексов позволяет разместить их в оперативной памяти.

Rows	Data	Index
1+ Million	>100 Mb	<20 Mb

Структура индексов оптимизирована для выполнения операций поиска.

Индексы значительно ускоряют операции поиска данных в БД.

Иногда – на 2-3 порядка. Но так бывает не всегда, нужно исследовать ситуацию.

#### Недостатки индексов

#### **SHOW TABLE STATUS FROM 'dbname'**

**Занимают** оперативную память.

Rows	Data	Index
1+ Million	>100 Mb	<20 Mb

В большинстве случаев значительно замедляют операции вставки, обновления, удаления, т.к. требуется обновлять сам индекс.

Иногда — в разы и на порядки. С этим можно «бороться», отключая индексы на момент вставки большого числа записей. В MySQL — так:

ALTER TABLE `table\_name`
DISABLE KEYS;
A потом:

ALTER TABLE `table\_name`
ENABLE KEYS;

# Признаки того, что индекс нужен:

- операции чтения из таблицы выполняются гораздо чаще, чем операции модификации;
- поле или совокупность полей часто фигурируют в запросах в секции WHERE;
- исследование показало, что наличие индекса повышает производительность.

Column	Туре	Collation
n_uid ?	int(11)	
n_rubric	int(11)	
n_dt 🥎	int(11)	
n_title	text	utf8_general_ci
n_annotation	text	utf8_general_ci
n_author	varchar(255)	utf8_general_ci
n_text	text	utf8_general_ci
n_source	text	utf8_general_ci

Проведём **исследование**. Для таблицы со следующей структурой сначала не будем создавать индексов, кроме первичного ключа:

Column	Туре	Collation
n_uid	int(11)	
n_rubric	int(11)	
n_dt	int(11)	
n_title	text	utf8_general_ci
n_annotation	text	utf8_general_ci
n_author	varchar(255)	utf8_general_ci
n_text	text	utf8_general_ci
n_source	text	utf8_general_ci

«Разброс» порядка 0.1 секунды

# Выполним по 1000 раз:

# Среднее время (с):

INSERT no 1000 d	трок
------------------	------

**0.027**289

SELECT \* from `news` where `n rubric`='....'

**4.035**899

SELECT \* from `news` where `n\_rubric`='...' AND `n dt`>='...' AND `n dt`<='...'

**4.065**648

SELECT \* from `news` where `n\_dt`>='...' AND `n\_dt`<='...'

**4.508**579

SELECT \* from `news` where `n\_title`='...' AND `n\_dt`>='...' AND `n dt`<='...'

**4.207**702

SELECT \* from `news` where `n\_title`='...' AND `n\_author`='...' AND `n\_dt`>='...' AND `n\_dt`<='...'

**4.187**432

SELECT \* from `news` where `n\_title`='...' AND `n\_author`='...'

**4.210**264

SELECT \* from `news` where `n\_title`='...'

**4.173**025

SELECT \* from `news` where `n\_author`='...'

**4.161**251

# Создадим следующие индексы:

- n\_rubric
- n\_rubric, n\_dt
- n\_title, n\_dt
- n\_dt, n\_author

Что здесь лишнее и почему?

n\_rubric, т.к. есть {n\_rubric, n\_dt}, и первое поле в индексе может использоваться как «самостоятельно проиндексированное»

n_author	varchar(255)	utf8_general_ci
n_text	text	utf8_general_ci
n_source	text	utf8_general_ci

# «Разброс» порядка 20 секунд!!!

# Снова выполним по 1000 раз: Среднее время (с):

INSERT I

**SELECT** \*

Индекс «сработал», просто в выборку попадает очень много записей, извлечение которых отнимает время.

0.027

0.896445

4.036

1.200757

4.066

**0.207**999

SELECT \*

SELECT \* from `news` where `n\_dt`>='...' AND `n\_dt`<='...'

4.509

**4.318**613

SELECT \* from `news` where `n\_title`='...' AND `n\_dt`>='...'
AND `n dt`<='...'

4.208

**0.003**918

SELECT \* from `news` where `n\_title`='...' AND `n\_author`='...' AND `n\_dt`>='...' AND `n\_dt`<='...'

4.187

0.000468

SELECT \* from 'nowe' where 'n title'-! ! AND

`n\_author`

При поиске по этому полю применимых индексов нет.

4.210

**0.000**909

SELECT \* from news where n\_une - ...

4.173

**0.000**279

SELECT \* from `news` where `n\_author`='...'

4.161

**6.567**766

#### Какие индексы создавать?

Итак, в общем случае индексы лучше создавать, чем не создавать ©.

Поля и их комбинации, на которых лучше создать индексы, определяются исходя из наиболее часто выполняемых запросов на чтение.

Проверить, какие индексы использует СУБД, можно (в MySQL) командой EXPLAIN запрос;

**EXPLAIN SELECT** \* from `news` where `n\_author`='...'

id	select_type	table	type	possible_keys					
1	SIMPLE	news	ALL	NULL	NULL	NULL	NULL	1000000	Using where

### Поле с уникальным индексом

Поле с НЕуникальным индексом

Уникальные (unique) — запрещают вставку одинаковых значений в поле таблицы. Как правило, создаются на «альтернативном ПК».

**Неуникальные** (non-unique) – просто индексы ©, созданные для ускорения поиска.

wui	wnui
1	2
3	2
4	2
7	3
9	3

Здесь может быть кластерный индекс.

Здесь может быть НЕкластерный индекс.

Кластерные (claster index) — строятся на поле, по значению которого упорядочена таблица. В таблице может быть только один кластерный индекс (и это, как правило — ПК).

**Некластерные** (non-claster index) – строятся на произвольном неупорядоченном поле таблицы.

sn_uid	sn_dt	sn_title
1	118758071	Title 64379878
2	121669144	Title 64046656
3	1569875561	Title 344103318
4	1631586361	Title 647355496
5	388175780	Title 8939791
6	160759710	Title 288420257

Одно поле, простой индекс.

Два поля, составной индекс.

**Простые** (simple index) – строятся на одном поле.

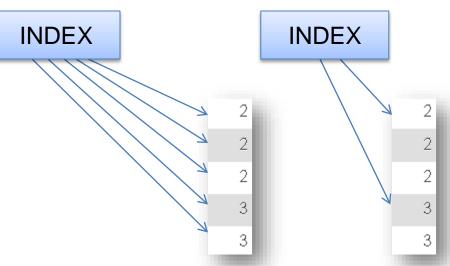
**Составные, сложные** (complex index) – строятся на нескольких полях или даже на выражениях.

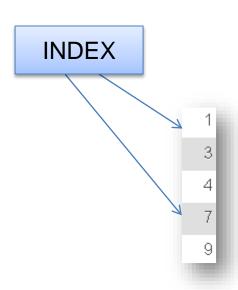
sn_uid	sn_dt	sn_title
1	118758071	Title 64379878
2	121669144	Title 64046656
3	1569875561	Title 344103318
4	1631586361	Title 647355496
5	388175780	Title 8939791
6	160759710	Title 288420257

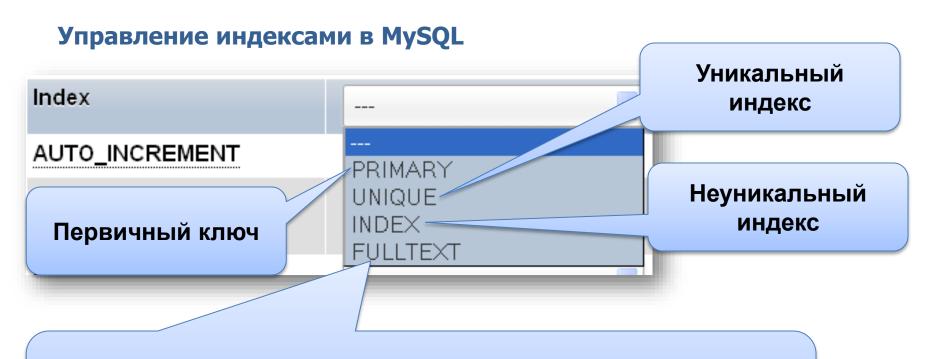
Сложный индекс по UPPER(`sn\_title`) (HE поддерживатеся MySQL)

Плотные (dense index) – указывают на конкретную запись в таблице или блок записей с одинаковыми значениями индексированного поля.

**Неплотные, редкие** (spare index) – указывают на блок (отсортированных) записей.







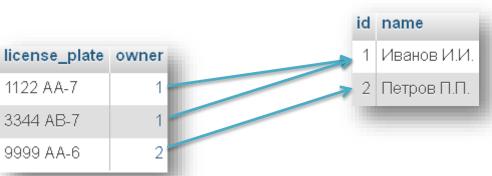
Полнотекстовый индекс для таблиц MylSAM и полей типа CHAR, VARCHAR, TEXT. Для поиска используется синтексис MATCH() ... AGAINST (далее – см. мануал по MySQL).

### СВЯЗИ

#### Понятие связи

Связь (relationship) – ассоциация, установленная между двумя и более сущностями (relations, entities).

Простая аналогия «из жизни» — указание в описании автомобиля информации о его владельце.



🗸 💽 relationship. owner

name: varchar(255)

🛭 license\_plate : varchar(50)

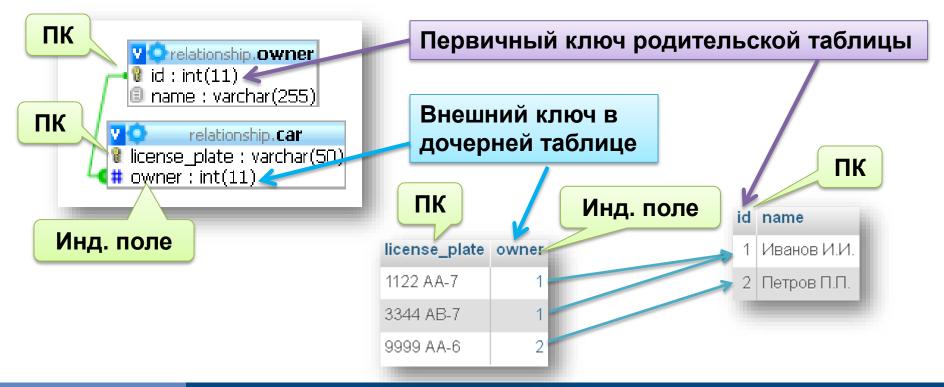
relationship.car

🐧 id : int(11).

# owner : int(11)

### Организация связи

Связь организуется за счёт миграции первичного ключа родительской таблицы в дочернюю таблицу. Получившееся в результате поле называется внешним ключом (foreign key).



### Виды связей: «1 ко многим» («многие к 1-му»)

Связь «1 ко многим» («многие к 1-му») определяет ситуацию, когда одной записи родительской таблицы соответствует несколько записей дочерней таблицы.

Разница между «1 ко многим» и «многие к 1-му» – исключительно в том, «кто является главным».





«Стрелка» рисуется в направлении родительской сущности.

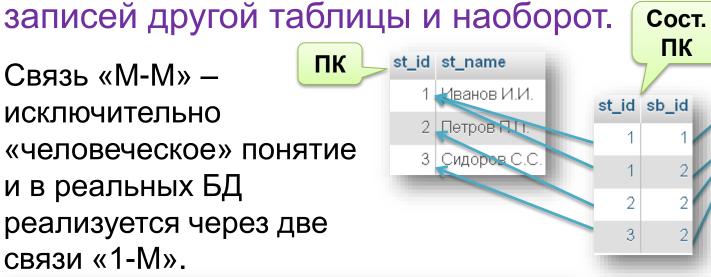


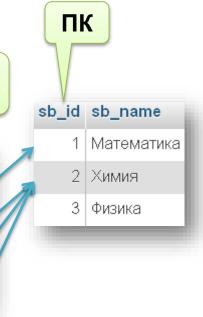
ПК

### Виды связей: «многие ко многим»

Связь «многие ко многим» определяет ситуацию, когда любой записи одной таблицы может соответствовать много

ПК Связь «М-М» исключительно «человеческое» понятие и в реальных БД реализуется через две связи «1-М».



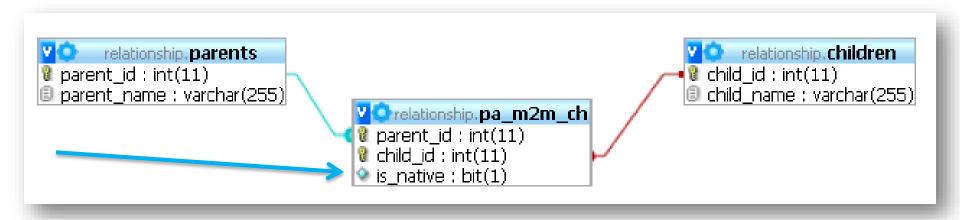


```
relationship.st m2m sb
                              8 st id : int(11)
                             🔞 sb_id : int(11)
  relationship.subjects
                                                                 relationship.students
🛭 sb id : int(11)
                                                              🛭 st_id : int(11)
  sb name : varchar(255)
                                                                st_name : varchar(255)
```

### Виды связей: «многие ко многим»

Связь «М-М» может также отражать некие свои свойства, не характерные для связываемых сущностей.

Например, в связи «родители-дети» можно отразить признак родства, который является именно свойством связи, но не родителя или ребёнка.



### Виды связей: «1 к 1-му»

Связь «1 к 1-му» определяет ситуацию, когда любой записи одной таблицы может соответствовать ровно одна запись другой таблицы и наоборот. Пк пк

tr\_id tr\_dt

2012-03-22 14:30:52

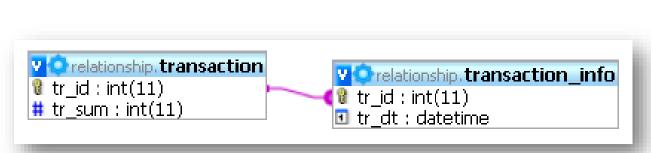
2 2012-03-28 09:40:20

tr\_id tr\_sum

500

2000

Наличие неаргументированной связи «1-1» в простых БД – признак ошибки в формировании структуры.



И ещё

20-30

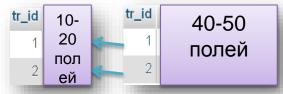
полей

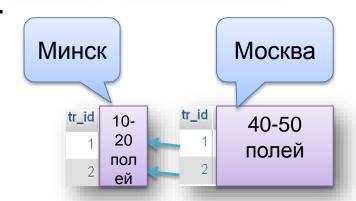
### Виды связей: «1 к 1-му»

# Связь «1-1» оправдана в том случае, если:

- Нужно разнести очень часто и очень редко обрабатываемые данные по разным таблицам для ускорения работы.
- Получилась «мега-таблица», количество полей в которой «упирается» в ограничения СУБД.
- Нужно разнести данные по удалённым друг от друга местам, где они активно обрабатываются.





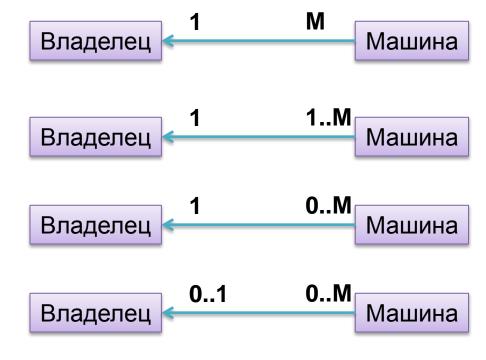


### Мощность связи

Мощность (кардинальность) связи (relationship cardinality) — указание возможного числа записей в таблице с каждой стороны связи.

При серьёзном проектировании указывают обе (нижнюю и верхнюю) границы с каждой стороны связи.

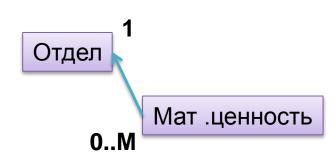
В чём разница таких записей?



### Идентифицирующая и неидентифицирующая связи

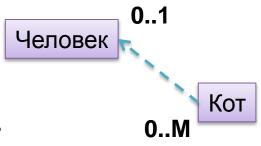
### Идентифицирующая связь

(identifying relationship) определяет ситуацию, когда запись в дочерней таблице обязана быть связана с записью в родительской таблице.



### НЕидентифицирующая связь

(non-identifying relationship) определяет ситуацию, когда запись в дочерней таблице может быть НЕ связана с записью в родительской таблице.



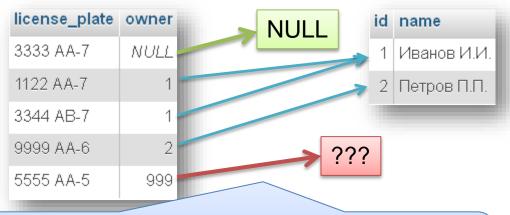
## **ССЫЛОЧНАЯ ЦЕЛОСТНОСТЬ ДАННЫХ**

### Понятие ссылочной целостности

Ссылочная целостность (referential integrity) — необходимое качество реляционной БД, заключающееся в отсутствии в любом её отношении внешних ключей, ссылающихся на несуществующие кортежи.

Простым языком: «если ключ на что-то ссылается, это что-то должно

существовать».



При наличии ЯВНО прописанных между таблицами связей СУБД не допустит такой ситуации.

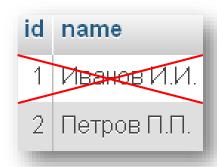
### Каскадные операции

Каскадные операции (cascade operations) – специальные ограничения БД, описывающие её поведение в случае удаления записи из родительской таблицы или изменения её первичного ключа.

Каскадное удаление	Каскадное обновление	Установка пустых ключей	Установка значения по умолчанию	Запрет каскадной операции
Записи в дочерней таблице удаляются	Значения внешних ключей в дочерней таблице обновляются	Во внешние ключи в дочерней таблице выставляется значение NULL	Во внешние ключи в дочерней таблице выставляется значение по умолчанию	Запись из родительской таблицы нельзя удалить, пока ей соответствуют записи в дочерней таблице

### Каскадное удаление

При каскадном удалении записи дочерней таблицы, соответствующие удаляемой записи родительской таблицы, тоже удаляются.

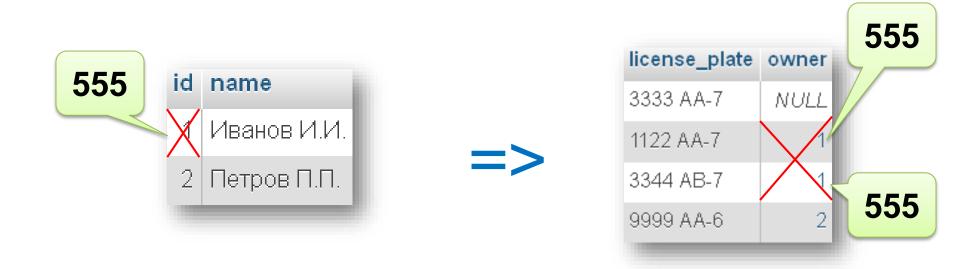




license_plate	owner
3333 AA-7	NULL
1122 AA-7	1
3344 AB-7	1
9999 AA-6	2

### Каскадное обновление

При обновлении первичного ключа родительской таблицы внешние ключи соответствующих записей в дочерней таблице принимают это же новое значение.

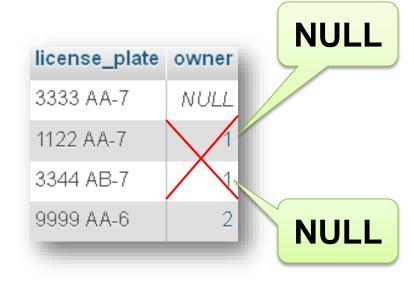


### Установка пустых ключей

При удалении записи из родительской таблицы во внешние ключи соответствующих записей дочерней таблицы устанавливается значение NULL.

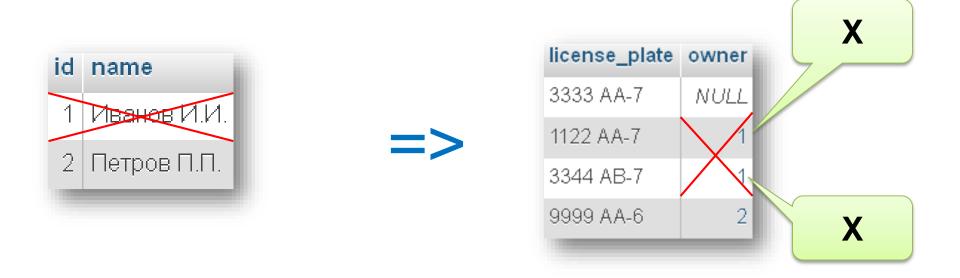






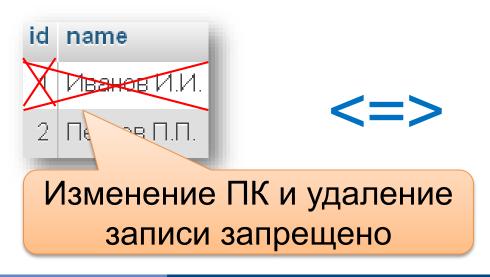
### Установка значения по умолчанию

При удалении записи из родительской таблицы во внешние ключи соответствующих записей дочерней таблицы устанавливается значение по умолчанию.



### Запрет каскадной операции

Пока существует хотя бы одна запись в дочерней таблице, соответствующая некоей записи в родительской таблице, эту запись из родительской таблицы нельзя удалить и/или нельзя изменить её первичный ключ.

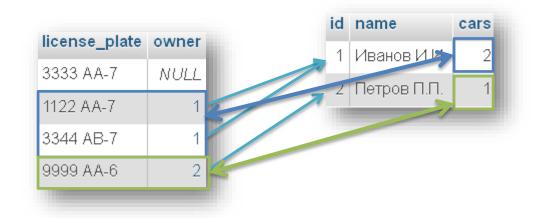


license_plate	owner
3333 AA-7	NULL
1122 AA-7	1
3344 AB-7	1
9999 AA-6	2

### Консистентность данных

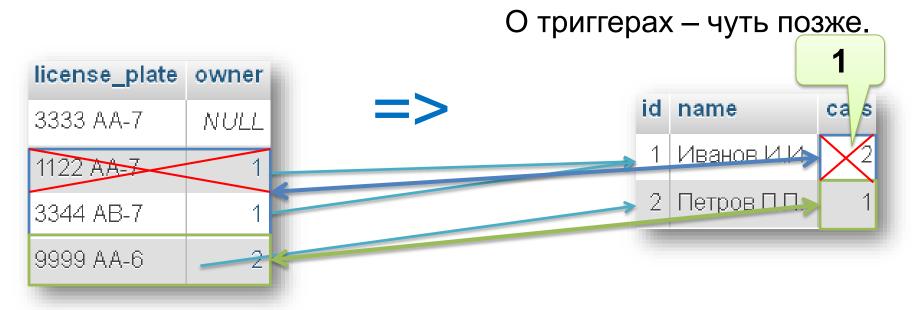
Консистентность данных (data consistency, data validity) – согласованность данных друг с другом: ссылочная целостность и внутренняя непротиворечивость.

Простым языком: фрагменты данных в БД не должны противоречить друг другу.



### Консистентность данных

Консистентность данных, как правило, обеспечивается созданием триггеров, контролирующих операции с таблицами и корректирующих соответствующие данные или блокирующих операцию.



### **ТРИГГЕРЫ**

### Понятие триггера

**Триггер** (trigger) – элементарная программа, написанная на некотором расширении языка SQL, и используемая для обеспечения консистентности данных.

Более подробно о триггерах – в разделах, посвящённых языку SQL.

	Вставка	Обновление	Удаление
Перед	BEFORE INSERT	BEFORE UPDATE	BEFORE DELETE
После	AFTER INSERT	AFTER UPDATE	AFTER DELETE

### Задачи триггеров

С использованием триггеров, как правило, решаются задачи, которые нельзя решить только с использованием ссылок, например:

- Обновление агрегированных данных («сколько у владельца машин»).
- Сложный запрет удаления («в системе должен быть хотя бы один администратор»).
- Контроль числовых значений («рост от 50 до 300 см»).
- ...

### СПАСИБО ЗА ВНИМАНИЕ!

### ВОПРОСЫ?

Отношения. Ключи. Связи.

**Author: Svyatoslav Kulikov** 

**Training And Education Manager** 

svyatoslav\_kulikov@epam.com