»Гл. ас. д-р Георги Чолаков »Бази от данни	
Правила за цялостност >	
правила за цялостност	

### Въведение

- » Във всеки момент всяка БД съдържа определена конфигурация от стойности на данните - тази конфигурация отразява (е модел на) една реална ситуация;
- » Определени конфигурации нямат реален смисъл те не представляват някакво възможно състояние на реалния свят;
- Дефиницията на БД трябва да бъде разширена тя трябва да включва правила за цялостност;
- » Тяхната цел е да информират БД за някои ограничения в реалния свят напр., теглото не може да бъде отрицателно число;
- » Така могат да се избегнат някои конфигурации, които нямат смисъл в реалния свят.

# Въведение

БД са субект на многобройни правила за цялостност – обикновено те са зависими от приложната област.

Пример за таблица с книги от библиотека:

- Не е реално цената да е отрицателно число т.е. стойността й трябва да е по-голяма от нула;
- Годината на публикуване не може да е по-голяма от текущата година;
- > ISBN номерът не се формира по произволен начин генерира се по определени международни правила и е уникален за всяка книга.



Въведение			
	авила за цялостност релационният модел		
включва два типа общ	овалидни правила за цялостност, валидни		
за всяка релационна Б	д:		-
> Концепцията за п	тървични ключове		
> Концепцията за в	ъншни ключове		
			•
		>	
Ключове			
ключове			
Ключовете в релационно » Уникална идентифика	ия модел се използват за:		
» Създаване на взаимос	отношения между таблици и референциален		
интегритет.			
Ще разгледаме: » Суперключ;			
» Кандидат-ключ;			
» Първичен ключ; » Алтернативен ключ;			
» Външен ключ.		>	•
Зависимости			
<ul> <li>Ролята на ключа се б (изчисляване);</li> </ul>	базира на концепцията за определяне		
» Определянето е про	цес, при който знаейки стойността на един		
атрибут определяме » Пример:	е (намираме) стойността на друг атрибут;		
	д - РАЗХОД = ПЕЧАЛБА		-
	пример за това как знаейки стойностите на а определим (изчислим) еднозначно		
стойността на трети ат	рибут.	>	

-					
За	DIA		~	CT	14

- » Определянето, обаче, в БД не се базира на формула, а на взаимоотношения между атрибутите;
- » Пример:

(ФАКУЛТЕТЕН НОМЕР, ДИСЦИПЛИНА) -> ОЦЕНКА

Т.е. знаейки кой е студентът и коя е дисциплината ние можем еднозначно да разберем каква оценка е поставена на този студент за конкретната дисциплина – в тази формула ясно се вижда, че няма изчисления, а се касае за взаимоотношение между студент и дисциплина, което завършва с поставяне на оценка.

)

### Зависимости

fac_no	first_name	last_name	email	specialty
1801681001	Иван	Петров	ipetrov@mail.com	STD
1801321011	Силвия	Георгиева	silvito.g@hotmail.com	SE
1801261024	Стефи	Огнянова	fire stefi@abv.bg	INF

В таблицата със студентите можем да открием някои взаимоотношения, напр. знаейки факултетния номер еднозначно можем да определим:

- » Име;
- » Фамилия;
- » E-mail;
- » Специалност.

Защото на всеки факултетен номер съответства точно по една стойност на всеки от изброените атрибути.

### Зависимости

Терминът за подобни взаимоотношения, базирани на определяне, се нарича функционална зависимост.

### Нека:

- > R релация;
- $\,>\,$  X, Y произволни подмножества на множеството на атрибутите на R.

### Тогава

> X → Y: "Х функционално определя Y" или "Y функционално зависи от X", ако и само ако всяка X-стойност в R е свързана точно седна У-стойност в R за всички възможни стойности на R.

\_\_\_\_

_					
<b>3</b> a	DI	101	 -	CT	14

### Примери:

- » (студент, дисциплина) ightarrow оценка
- » фак. номер  $\rightarrow$  (име, фамилия, email)
- » рег. номер ightarrow (марка автомобил, модел, собственик)

Атрибутът от лявата страна (който определя) се нарича детерминант, а вдясно (чиято стойност е определяна) - зависим.

Ако детерминантът се състои от повече от един атрибут се нарича съставен.

Видове	зависи	имости
--------	--------	--------

» Пълна функционална зависимост - атрибут е напълно функционално зависим от множество атрибути ако и само ако той е функционално зависим от цялото това множество от атрибути и не е зависим от никое негово подмножество от атрибути.

Пример: нека имаме релацията EXAMS, съдържаща данни за проведените изпити, с идентификатор (FNO, SUBJ).

FNO	SUBJ	HALL	GRADE
1701261010	DB	547	6
1701261016	SE	446	5

### (FNO, SUBJ) $\rightarrow$ (GRADE)

Фак. номер и дисциплината определят оценката. Това означава, че студенть може изкара добра оценка по дадена дисциплина, но не е сигурно, че ще изкара добри оценки по всички дисциплини, т.е. оценката замси и от цизата комбинация студент и дисциплината, и не зависи от никой от тях

### Видове зависимости

- » Частична функционална зависимост такава е налице, когато само подмножество от съставен детерминант е достатъчно, за да определи функционално даден атрибут.
- » Пример: при провеждане на изпити те се насрочват в конкретни зали, дори за някои от тях има изискване за конкретна зала, и може да се каже, че залата зависи от конкретната дисциплина:
- » (FNO, SUBJ) ightarrow (HALL), т.е. фак. номер не играе роля за определянето, защото:
- $\text{» (SUBJ)} \to \text{(HALL)}$

FNO	SUBJ	HALL	GRADE
1701261010	DB	547	6
1701261016	SE	446	5
1701261016	DB	547	4

RUJORE	SARUC	MMOCT

» Транзитивна функционална зависимост - транзитивна ФЗ съществува, когато има "преходна" функционална зависимост. Т.е., ако A ightarrow B, а B ightarrow C, то A ightarrow C.

FNO	NAME	CITY	ZIP
1701261010	Никола Михалев	Смолян	4700
1701261016	Евгения Михайлова	Пловдив	4000
1701261017	Лимитър Георгиев	Swfon	8600

Тук FNO ightarrow CITY, CITY ightarrow ZIP, следователно FNO ightarrow ZIP транзитивно.

1	
->	
-	

# Суперключ

 » Подмножество от атрибути на релация, което уникално идентифицира всяка n-торка от тялото й, т.е. всеки ред от таблицата.

Суперключът определя функционално всеки атрибут на релацията.

Суперключовете представят ограничение, което предпазва две п-торки да имат съвпадащи стойности за тези атрибути, т.е. стойностите на суперключа за всеки ред трябва да са уникални.

### Кандидат-ключ

» Кандидат-ключ за дадена релация е минималното множество от атрибути, чиито стойности уникално идентифицират n-торка в тази релация – т.е. суперключ без атрибутите, които нямат отношение към уникалната идентификация.

Таблица може да има множество кандидат-ключове, от които се избира един за идентификатор, наречен първичен ключ.

Ключ-кандидат, който се състои от повече от един атрибут, се нарича <u>съставен</u>, а с един атрибут – <u>прост.</u>

Пъ	nei	ичен	VП	M
HD	NDA	ичен	KJI	ЮЧ

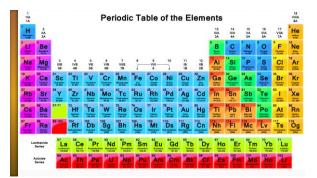
Нека R е релация с атрибути  $\{A_1,A_2,...,A_n\}$ . Множеството на атрибутите  $K=(A_i,A_j,...,A_k)$  от R е ключ-кандидат (candidate key) на R, ако то удовлетворява следните две, независещи от времето условия:

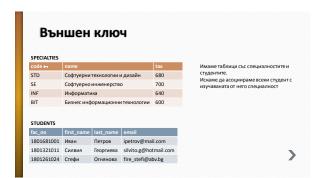
- > <u>уникалност (uniqueness)</u> в един и същ момент не съществуват два различни записа на R с еднаква стойност за  $A_i$ ,  $A_i$ ,....,  $A_k$ ;
- > <u>достатъчност (minimality)</u> никой от  $A_{i_{\ell}}$   $A_{j_{\ell}}$ ,...,  $A_{k}$  не може да бъде премахнат, без това да наруши уникалността.

Интегритет на обект (entity integrity) е състояние, в което всеки ред от таблицата (всяка инстанция на обекта) може да бъде уникално идентифицирана. Затова първичният ключ трябва да гарантира уникалност, като в същото време всички съставящи го атрибути трябва да имат стойност (различна от NULL).

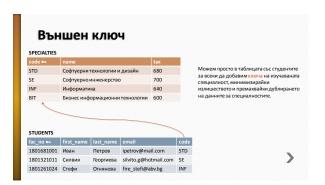
# Алтернативен ключ

- » Ако една релация има повече от един ключ-кандидат релационният модел изисква един от тях да бъде избран за първичен;
- » Останалите кандидат-ключове се наричат алтернативни в таблицата на тях може да им бъде наложено (или не) ограничение за уникалност на стойностите им, докато това автоматично се прилага за първичния ключ.









### Външен ключ

Външен ключ: нека  $R_1$  е базова релация. Външен ключ в  $R_1$  е едно подмножество на множеството от атрибутите  $\{A_1^R,A_2^R,\dots,A_n^R\}$  така, че:

- $\,>\,$  съществува една базова релация  ${
  m R_2}$  с ключ-кандидат;
- > във всеки един момент стойностите на външния ключ в  ${\bf R}_1$  са измежду стойностите на кандидат-ключа в някой запис на  ${\bf R}_2$ .

Важно е да се отбележи, че не е задължително  $\mathsf{R}_1$  и  $\mathsf{R}_2$  да са различни редации

>

соde         name         tax           TD         Софтуерни технологии и дизайн         680           SE         Софтуерно инженерство         700           INF         Информатика         640           BIT         Бизнес информационни технологии         600           Стойностите му са измежду стойностите на референцирания първичен ключ.           STUDENTS           fac_no № [first_name last_name email         code           1801361001         Иван         Петров           1801321011         Силвия         Георгиева silvito g@hotmail.com         STD           1801321012         Силвия         Георгиева         Introductive name         STD	Вън	шен	ключ	4			
SE Софтуерно инженерство 700  INF Информатика 640  BIT Бизнес информационни технологии 600  STUDENTS  fac_no № first_name last_name email code 1801321011 Иван Петров ipetrov@mail.com STD  1801321011 Силвия Георгиева silvito.g@hotmail.com SE	code ⊷	name			tax		
SE         Софтуерно инженерство         700         У Подмномество от атрибутите су тойностите на референцирания първичен ключ.           ВІТ         Бизнес информационни технологии         600         У Стойностите на референцирания първичен ключ.           STUDENTS           Гас_по тойностите на референцирания първичен ключ.           1801681001         Иман         Петров         ipetrov@mail.com         STD           1801321011         Силвия         Георгиева         silvito.g@hotmail.com         SE	STD	Софтуерни технологии и дизайн			680		Проверка дали STUDENTS.code отговаря
INF Информатика 640 Стойностите му са измежду стойностите на референцирания транев референцирания транев референцирания транев измежду стойностите на транев и на транев измежду стойнос	SE	Софтуерно инженерство			700		
STUDENTS    fie_no →   first_name   tast_name   email   code     1801321011 Силвия   feoprisea   silvito_g@hotmail.com   SE	INF	Информатика			640		
STUDENTS  fac_no in first_name last_name email code  1801.681001 / Mean	BIT	Бизнес информационни технологии			600		
1801681001 Иван Петров ipetrov@mail.com STD 1801321011 Силвия Георгиева silvito.g@hotmail.com SE	STUDENTS				_		
1801321011 Силвия Георгиева silvito.g@hotmail.com SE	fac_no ⊷	first_name	last_name	email		code	
	1801681001	Иван	Петров	ipetrov@ma	il.com	STD	
1801261024 Стефи Огнянова fire stefi@abv.bg INF	1801321011	Силвия	Георгиева	silvito.g@hc	tmail.com	SE	· ·
	1801261024	Стефи	Огнянова	fire_stefi@a	bv.bg	INF	

### Външен ключ

В дефиницията беше подчертано, че двете релации  ${\sf R}_1$  и  ${\sf R}_2$  не е необходимо да се различават.

Като пример за това можем да споменем релацията EMPLOYEES, в която се съхраняват данните за служителите на определена компания, включително с йерархията им.

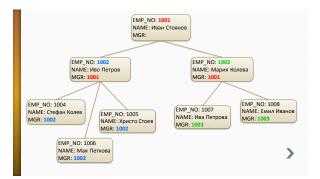
Това е типична дървовидна структура, която може да бъде съхранена в таблица.

### Външен ключ

В дефиницията беше подчертано, че двете релации  ${\sf R}_1$  и  ${\sf R}_2$  не е необходимо да се различават.

Като пример за това можем да споменем релацията EMPLOYEES, в която се съхраняват данните за служителите на определена компания, включително с йерархията им.

Това е типична дървовидна структура, която може да бъде съхранена в таблица.



MP_NO ⊷ 1001	NАМЕ Иван Стоянов	SALARY 800	IVIGIC	Тук EMP_NO е първичният ключ, а MGR є
1001	Иво Петров	630	1001	външният ключ, който референцира
1003	Мария Колева	600	1001	първичния ключ на същата релация FMPLOYFFS.
1004	Стефан Колев	430	1002	EIVIF LOTELS.
1005	Христо Стоев	450	1002	Логиката е следната – служителят с номе
1006	Мая Петкова	390	1002	1001 е мениджър на служителите с
1007	Ива Петрова	540	1003	номера 1002 и 1003.
1008	Емил Иванов	300	1003	

### Свойства на външните ключове

- » ВК представлява подмножество от атрибути на релация;
- По дефиниция всяка стойност на даден ВК във всеки един момент трябва да е измежду стойностите на съответния референциран първичен ключ;
- » Обратното не се изисква т.е. КК може да има стойност, която не се появява във ВК;
- Ако даден ред е референциран от ред(ове) от друга релация, то изтриването му се предотвратява от ВК (освен ако не е указано пруго):
- Един ВК ще бъде съставен, ако кореспондиращият КК е също съставен; ако КК е прост, ВК също ще е прост;
- » Всеки атрибут на един ВК трябва да бъде дефиниран върху същия домейн, върху който е дефиниран кореспондиращият атрибут от референцирания КК.

### Пример за съставен външен ключ RECIPES recipe\_name date\_published type author • Иван Петров Кюфтета от тиквички 12.10.2019 Вегетарианска Атрибутите, съставящи първичния ключ, мигрират в таблицата, където ще бъдат външе Исяке EIERTS.author и INGREEIERTS.author и INGREEIERTS.author трябва да бъдат от един и същ тип; RECIPES recipe, name и INGREEIERTS recipe, name трябва да бъдат от един и същ тип; RECIPES recipe, name трябва да бъдат от един и същ тип; при и същ тип; при и същ тип; при и същ тип; при и същ табищата, която да играе роля на прост ключ, за да се избелнат усложненията от съставния ключ. INGREDIENTS product\_name quantity author recipe\_name 500 гр. Иван Петров Кюфтета от тиквички 300 rp. Мария Иванова Боб с наденица Наденица 200 rp. Мария Иванова Боб с наденица Домати 250 гр. Мария Иванова Боб с наденица

- » <u>Референциална цялостност</u>: ситуация, в която БД не съдържа невалидни стойности на ВК.
- Референциално ограничение: ограничението, че стойностите на един ВК трябва да съответстват на стойностите на съответния пк
- » Релация:
  - > референцираща релацията, която съдържа ВК;
  - > референцирана (целева) релацията на ПК.

## Референциални диаграми

- Нека разгледаме таблиците за държави и региони отново. Можем да представим референциалните ограничения чрез референциалната диаграма на фигурата;
- » Всяка стрелка показва наличие на външен ключ в релацията, от която тръгва, референциращ съответен първичен ключ в релацията, към която сочи. Добра идея е да се указват атрибутите, съставящи външния ключ.

REGIONS	REGION ID = REGION ID	COUNTRIES			
REGION_ID smallint <pl>smallint <pl>smallint NAME varchar(25) not null not null</pl></pl>	-	COUNTRY_ID NAME REGION_ID	char(2) varchar(40) smallint	≤pk≥ <fk></fk>	not null not null null



# Референциален път

- » Нека релациите от  $R_n$  до  $R_1$  са такива, че има референциално ограничение от  $R_n$  към  $R_{n-1}$ , от  $R_{n-1}$  към  $R_{n-2}$  и т.н. до  $R_1$ .
- » Последователността от референциални ограничения от  $R_n$  до  $R_1$  представя референциален път от  $R_n$  до  $R_1.$



 $\boxed{\mathsf{R}_{\mathsf{n}} \to \mathsf{R}_{\mathsf{n-1}} \to ... \to \mathsf{R}_{\mathsf{2}} \to \mathsf{R}_{\mathsf{1}}}$ 

## Референциален цикъл

- » Нека релациите от  $R_n$  до  $R_1$  са такива, че има референциално ограничение от  $R_n$  към  $R_{n,2}$  от  $R_{n,1}$  към  $R_{n,2}$  и т.н. до  $R_1$ , като  $R_1$  също има външен ключ към  $R_n$ . Така представени, релациите формират референциален цикъл.
- » Самореференциращите се релации са частен случай на референциален цикъл.
- » По-точно, референциален цикъл съществува, ако има референциален път от  $R_{_{\rm n}}$  до  $R_{_{\rm n}}.$



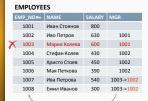
## Изтриване на редове при референциален цикъл

### **EMPLOYEES**

EMP_NO⊷	NAME	SALARY	MGR
1001	Иван Стоянов	800	
1002	Иво Петров	630	1001
1003	Мария Колева	600	1001
1004	Стефан Колев	430	1002
1005	Христо Стоев	450	1002
1006	Мая Петкова	390	1002
1007	Ива Петрова	540	1003
1008	Емил Иванов	300	1003
1			1

- » Ако е налице референциален цикъл изтриването на редове от таблиците в него може да се окаже проблемно;
- » Например, как да изтрием реда с EMP\_NO = 1003, след като той е референциран от други редове?

# Изтриване на редове при референциален цикъл



» Можем да прехвърлим служителите с MGR = 1003 под ръководството на 1002, след което да изтрием ред 1003.

### Правила за външните ключове

- Правило за референциалната цялостност: БД не трябва да съдържа стойности на ВК, които нямат съответна стойност на референцирания ПК в целевата релация;
- Всяко състояние на БД, което не удовлетворява правилото за референциалната цялостност, по дефиниция е некоректно;
- » Но правилото не казва как да се предпазим от такива некоректни състояния.

### Възможности:

- Системата да отхвърли всяка операция, която би довела до некоректно състояние, ако бъде изпълнена;
- Системата извършва операцията, а след това извършва допълнителни компенсиращи операции.

» За всеки ВК съществуват два основни въпроса, на които трябва да се отговори:

1. Какво ще се случи при опит да се изтрие целевата на ВК референция? Например при опит да се изтрие регион, за който съществува поне една държава.

### Съществуват два подхода:

- » RESTRICTED операцията изтриване е ограничена само за случая, когато няма свързани данни; ако такива са налице, изтриване не се извършва;
- » CASCADES операцията се разширява каскадно и изтрива също така и свързаните записи.

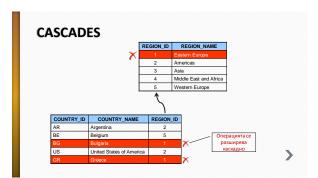
RESTRICTED

Onepaujusta ce
Otxis-juna

Ti Eastern Europe
1 Eastern Europe
2 Americas
3 Asia
4 Middle East and Africa
5 Western Europe

COUNTRY\_ID COUNTRY\_NAME REGION\_ID
AR Argentina 2
BE Belgium 5
BG Bugaria 1
US United States of America 2
BS General 5

-		

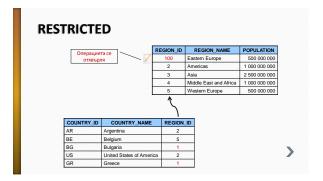


2. Какво ще стане при опит да се промени стойността на един ПК, който е целева референция на ВК? Например - опит да се промени идентификатор на регион, за който съществува поне една държава.

Отново се използват същите два подхода:

» RESTRICTED - операцията промяна е ограничена само за случая, когато няма държава за региона; ако такава е налице, промяна не се извършва;

» CASCADES - операцията се разширява каскадно и променя също така и записа за държавата.



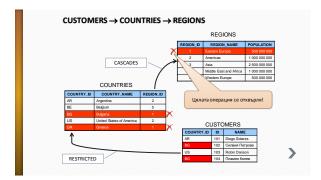
		REGION_ID	REGION_NAME	POPULATION
	7	100	Eastern Europe	500 000 000
		2	Americas	1 000 000 000
		3	Asia	2 500 000 000
		4	Middle East and Africa	1 000 000 000
		5	Western Europe	500 000 000
COUNTRY_ID	COUNTRY_NAME	REGION	ID	
	COUNTRY_NAME Argentina	REGION 2	ID.	
COUNTRY_ID AR BE				0
2	Argentina	2		Операцията се
₹	Argentina Belgium	2 5 100		Операцията се ширява каскади

### Забележка 1

- Опциите за правилата DELETE и UPDATE за ВК не изчерпват възможностите - те по-точно представят най-общите, които се изискват от практиката.
- » По принцип съществуват повече възможности. Например, опит за изтриване на един регион възможностите могат да бъдат:
  - > инициализира се диалог с краен потребител за потвърждение;
  - > информацията може да се архивира (какво ще се прави);
  - > държавата може да се причисли към друг регион.

### Забележка 2

- » Нека  $R_2$  и  $R_1$  са референцираща и целева  $R_2 \to R_1$  и нека delete-правилото за това референциално ограничение е **CASCADES**;
- » Следователно изтриването на n-торка в  ${f R}_1$  ще предизвика изтриване на n-торка в  ${f R}_2$ ;
- » Да допуснем, че  $R_2$  е целева за  $R_3$ :  $R_3 \to R_2 \to R_1$ . Тогава изтриването на n-торка зависи и от delete-правилото между  $R_3$  и  $R_2$ . Ако то забранява изтриване (RESTRICTED), тогава не се изтрива нищо и БД остава непроменена. И т.н. до произволно ниво.



# Задача от интервю » Да се провери налице ли е референциална цялостност в представената база от данни с SQL команда. REGIONS REGIONS REGIONS REGION D smallint sole not null not null NAME varchar(25) sole not null REGION\_ID smallint varchar(40) not null REGION\_ID smallint des null REGION\_ID smallint des null



# Задача от интервю - решение » Какво е референциален интегритет? Състояние на БД, в която няма стойности на външни ключове, които не се срещат сред стойностите на референциарните от тях първични. » Значи трябва да проверим има ли държави със стойност в атрибута ССОИТRES гедоп і d, която да не се среща сред стойностите на REGIONS.region i d. » Ако има такива – няма референциален интегритет! Ако няма – налице е референциален интегритет. REGIONA R

Задача от интерв	ю - решение
SELECT *	
FROM COUNTRIES	
WHERE region id NOT IN	N (SELECT region id
- <del>-</del>	FROM REGIONS)
Results (N) Messages	⊞ Results DW Messages
Results Messages  COUNTRY ID NAME REGION ID	Results
	COUNTRY_ID NAME REGION_ID  1 AR Apxentina 20
	COUNTRY_ID NAME REGION_ID

# Нулеви стойности » Нулева стойност: липса на стойност (данни) за даден атрибут. » В реалния свят има много такива ситуации - напр. неизвестен адрес; » В релационния модел е приет един специален маркер за означаване на липсваща информация — NULL; » Един атрибут може да съдържа или може да му бъде забранено да съдържа нулеви стойности.

## Външни ключове и нулеви стойности

- » Да се върнем на примера "Държава-Регион";
- » Възможно е някоя държава да не е причислена към конкретен географски регион ⇒ допуска се ВК да има нулеви стойности.

<u>Забележка:</u> за всеки ВК конструкторът на БД трябва да реши дали да се допускат нулеви стойности или не в зависимост от конкретната логика.

>

- » Да се върнем на въпроса какво ще стане, ако искаме да изтрием п-торка от целевата релация, например, да изтрием регион, за който съществува държава:
  - Ако за ВК са разрешени NULL съществува нова трета възможност NULLIFIES - т.е. ВК получава стойност NULL и тогава се изтрива основната п-торка;
  - > Същото важи и за промяната.

# REGION ID REGION NAME POPULATION 1 Extent Europe 500 000 000 2 Americas 1 000 000 000 3 Assia 2 500 000 000 4 Middle East and Africa 1 000 000 000 5 Western Europe 500 000 000 5 Western Europe 500 000 000 COUNTRY ID COUNTRY NAME REGION ID AR Argentina 2 BE Belgium 5 BG Bulgaria NULL US United States of America 2 GR Gracce NULL

NULL и логическите оператори	
» SQL има три логически стойности: TRUE, FALSE и UNKNOWN;	
<ul> <li>Стойността UNKNOWN произлиза в резултат от използването на NULL в сравнения с други предикати, но UNKNOWN е логическа стойност и не е еквивалентна на NULL, което е маркер за липса на стойност;</li> </ul>	
» Ето защо в SQL пишем	
x IS [NOT] NULL	
вместо	
x = NULL	>

AND	TRUE	UNKNOWN	FALSE	
TRUE	TRUE	UNKNOWN	FALSE	
UNKNOWN	UNKNOWN	UNKNOWN	FALSE	
FALSE	FALSE	FALSE	FALSE	
OR	TRUE	UNKNOWN	FALSE	
TRUE	TRUE	TRUE	TRUE	
UNKNOWN	TRUE	UNKNOWN	UNKNOWN	
FALSE	TRUE	UNKNOWN	FALSE	
x	NOT	x * unknor	IN = UNKNOWN	
TRUE	FALSE	X / UNKNO		
UNKNOWN	UNKNOWN	X + UNKNOT		,
FALSE	TRUE	x % UNKNO		