

Normalizácia relačných databáz

Normálizácia

1NF (1. normálna forma)

2NF závislosť medzi kľúčmi a atribútmi

3NF závislosť medzi kľúčmi a atribútmi

HBCNF (Heat-Boyce-Codd) závislosť medzi kľúčmi

Normalizácia

Normalizácia je proces eliminovania duplicit, nekonzistentností v dátach a závislostiach v relačných databázach pomocou

- pravidiel medzi atribútmi a funkčnými závislosťami
- dekompozície tabuľky.

Pritom atribúty sa rozdeľujú na kľúčové a neklúčové.

- Kľúče**
- NK - nad kľúč: $NK \rightarrow R$
 - KK - kandidátny kľúč: a) $KK \rightarrow R$ b) $\neg \exists K \subset KK: K \rightarrow R$
 - PK - primárny kľúč (jeden z KK)
- Atribúty**
- KA - kľúčový/é atribút/y: je/sú súčasťou nejakého KK
 - NA - neklúčový atribút: nie je súčasťou žiadneho KK

Tab1

idStudent	Meno	Aresa	idPredmet	NazovPred	Ucitel	Miestnost
1	On	Ke	11	PAZ	M	P3
1	On	Ke	12	DBS	T	P1
2	Ona	Rv	12	DBS	T	P1

- Duplicita - redundantnosť
- Problémy s
 - obnovením
 - vložení
 - vymazaním

	Výhody	Nevýhody
Pôvodná tabuľka	Prehľadnejšia	Duplic. + Problémy
Dekompozícia	Menej Duplic. + Problémov	Viac tab. => View

Dekompozícia:

Tab2

idStudent	Meno	Aresa
1	On	Ke
2	Ona	Rv

Tab2

idStudent	idPredmet
1	11
1	12
2	12

Tab3

idPredmet	NazovPred	Ucitel	Miestnost
11	PAZ	M	P3
12	DBS	T	P1

Tab1 obsahuje $3 \cdot 7 = 21$ údajov a Tab2-Tab3 obsahujú $2 \cdot 3 + 3 \cdot 2 + 2 \cdot 4 = 20$ údajov. To nevyzerá na prvý pohľad veľkou úsporou. Ale skúsme vnoriť do nich údaj, že ona si zapísala aj predmet PAZ. V Tab1 potrebujeme na to v novom riadku až 7 údajov (1 nový + 6 ďalších) kým v prípade Tab2-Tab3 iba 2 údaje (1 nový + jeden ďalší) ako nový riadok v Tab3.

Máme 6 normálnych foriem NF. Edgar Frank Codd definoval prvé 3+1 (1971+1974), ale iba druhá nie je problematická.

Ďalšie normálne formy boli definované Faginom a Date.

- 4,5,DKNF Fagin 1977-1981,

- 6NF Chris Date 2002

Hovorí sa, že tabuľka je v KNF, ak spĺňa požiadavky KNF.

Každá NF má silnejšie požiadavky ako predchádzajúce.

KNF vyžaduje, aby (K-1)NF bola splnená, kde $K \geq 2,6$.

Ale tabuľka, ktorá je v KNF nemusí byť v (K+1)NF.

1NF (jedna formulácia) v podstate zodpovedá definícii relácie.

Druhá a tretia NF sa zaoberajú vzťahom neklúčových atribútov ku kompozitným kľúčom. Požadujú, aby všetky neklúčové atribúty boli funkčne závislé na „(celom) kľúči a ničom inom ako na kľúči“.

Denormalizácia

Keď tabuľky sú už príliš štíhle, je ch príliš veľá a obsahujú malý počet atribútov, často pristupujú k opačnému procesu – k denormalizácii tabuliek.

1NF

Tu v podstate ide o požiadavky, pravidlá, ktoré súvisia s definíciou relácie, tabuľky.

Codd pôvodne požiadaval splnenie nasledujúcich podmienok (prvé dva sú problematické)

1. - jedinečnosť riadkov, tzn. aby relácia mala primárny kľúč (dnes DB systémy netrávajú na tom, aby tabuľky mali PK)
2. - atomické (ďalej už nedeliteľné) hodnoty pre atribúty (dátum, desatinné číslo, reťazec – nie sú skutočne atomické, iba na základe dohody).
3. - tabuľka nesmie obsahovať podobné, opakujúce sa atribúty!

Príklad:

Tab-Student

idStud	Krstne	Priezvisko	CisloTel
1	Jan	X	1
2	Oto	Y	2

Problémy s touto tabuľkou môžu nastať, keď dovolíme, aby študenti mohli mať viac telefónnych čísel. Napríklad: 2, Oto, Y, 6. Ako to vyriešiť?

Ukážeme 4 riešenia, z ktorých iba posledné je akceptovateľné.

1. riešenie: zoznam hodnôt

2	Oto	Y	2, 6
---	-----	---	------

Tu máme nevhodný dátový typ a dané riešenie sťažuje dopytovanie

2. riešenie: nový riadok

2	Oto	Y	2
2	Oto	Y	6

Tu môžu nastať problémy s obnovením a budeme mať viac duplicitných hodnôt.

3. riešenie: opakujúce sa atribúty

idStud	Krstne	Priezvisko	CisloTel1	CisloTel2	CisloTel3
1	Jan	X	1		
2	Oto	Y	2		
2	Oto	Y		6	

Pri tomto riešení na jednej strane môžeme mať stĺpec samých *null* hodnôt a na druhej, čo keď niekto má až štyri čísla.

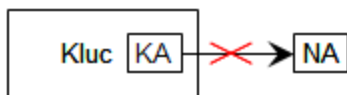
4. riešenie: dekompozícia na dve tabuľky:

idStud	Krstne	Priezvisko
1	Jan	X
2	Oto	Y

idStud	CisloTel
1	1
2	2
2	6

2NF

Definícia: Druhá normálna forma požaduje, aby relácia bola v prvej NF a zároveň žiadny **NA nezávisel od časti žiadneho KK**



- teda, aby každý neklúčový atribút bol **úplne funkčne závislý** na kľúči (na **celom kľúči**, nie len na jeho podmnožine)

⇔ neklúčový atribút upresňuje, charakterizuje celý kľúč

⇔ neexistuje neklúčový atribút, ktorý by upresňoval, charakterizoval časť kľúča

Ak tabuľka je v 1. NF a nemá kompozitný kľúč, potom automaticky je v 2. NF.

Príklad 1.

Tab: Predmety

Kompozitný kľúč

<u>Ucitel</u>	<u>Predmet</u>	Adresa
A	PAZ	Ke
B	DBS1	Rv
B	DBS2	Rv

Adresa charakterizuje učiteľa, ale nie Predmet.

Dôsledok: duplicita, problémy s obnovou

Riešenie:

Tab: Učitelia

<u>Ucitel</u>	Adresa
A	Ke
B	Rv

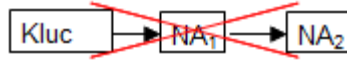
Tab: Predmety

<u>Ucitel</u>	<u>Predmet</u>
A	PAZ
B	DBS1
B	DBS2

3NF

Uvažujme relačnú schému $R(\mathcal{A})$.

Definícia: Tretia normálna forma požaduje, aby relácia bola v 2. NF a zároveň - (Codd 1972) žiadny **NA nezávisel tranzitívne** od žiadneho KK



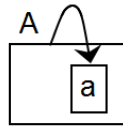
- (Zaniolo 1982) Každá závislosť $A \rightarrow a$, kde $A \subseteq \mathcal{A}$, $a \in \mathcal{A}$, musí byť jedného typu z troch

Zaniolo 1982

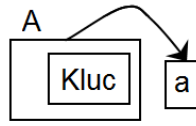
$A \rightarrow a$

A - atribut/y

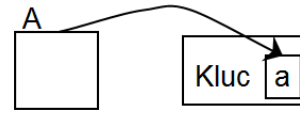
a - atribut



1) Trivialna zavislost



2) A je nadkluc



3) a je cast kluca

Poznámka: je dôležité si uvedomiť, že v 1) $a \in A$ a v 2) $Kluc \subseteq A$.

Pripomíname, že závislosť $A \rightarrow a$ je triviálna, ak $a \subseteq A$.

⇔

Relačná schéma R je v **3NF**, keď pre každú v nej platnú netriviálnu funkčnú závislosť $x \rightarrow y$ platí:

- x je nadkľúč alebo

- y je prvok nejakého kľúča (obsahuje kľúčový atribút) R.

Poznámka:

- dané dve definície sú ekvivalentné

- každý neklúčový atribút má priamo, teda **netranzitívne** závisieť od kľúča.

⇔ neklúčový atribút upresňuje, charakterizuje iba kľúč, ale nie iný neklúčový atribút.

Ak tabuľka je v 2NF a nemá neklúčový atribút, automaticky je v 3NF – **aj v HBCNF**.

Príklad 1.

Kľúč

Turnaj	Rok	Vitaz	DatumNarodenia

Vek charakterizuje Víťaza => dekompozícia:

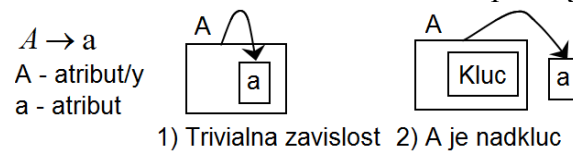
Turnaj	Rok	Vitaz

Pretekar	DatumNarodenia

HBC-NF (Heat-Boyce-Codd)

Uvažujme relačnú schému $R(\mathcal{A})$.

Definícia: HBC normálna forma požaduje, aby relácia bola v 2. NF a zároveň pre každú závislosť $A \rightarrow a$, kde $A \subseteq \mathcal{A}$, $a \in \mathcal{A}$, musí platiť jedna z dvoch možností



Relačná schéma R je v BCNF, keď pre každú v nej platnú netriviálnu funkčnú závislosť $x \rightarrow y$ platí:

x je nadkľúč.

Teda:

- tabuľka je v **HBCNF**, ak vždy, keď $x \rightarrow y$ a y nie je súčasťou x , potom x je nadkľúč.

Poznámky :

- Vo väčšine prípadov, ak tabuľka je v 3NF, je aj v **HBCNF**.

- 3NF tabuľka nespĺňa požiadavky HBCNF ak jej kompozitné kľúče sa prelínajú (majú spoločný atribút).

Príklad

Uvažujme tabuľku

Student0

<u>idStud</u>	Ucitel	UciKabinet	Predmet1	Predmet2	Predmet3
1	A	101	DBS1	DBS2	MPJ
2	B	102	PAZ1	PAZ2	Siete

Pri normalizácii hľadáme čo je zlé.

Na ceste k **dosiahnutiu až 3.NF** zrušme najskôr opakujúce sa atribúty pre 1. NF:

Student1

idStud	Ucitel	UciKabinet	Predmet
1	A	101	DBS1
1	A	101	DBS2
1	A	101	MPJ
2	B	102	PAZ1
2	B	102	PAZ2
2	B	102	Siete

Existuje kľúč K: K → Predmet ?



Nepotvrdzujeme to, čo sa zdá byť správne, napr.

idStud → Ucitel alebo idStud → UciKabinet,

ale hľadáme to, čo je určite zlé:

idStud **ne** → Predmet a ani (idStud, Ucitel, UciKabinet) **ne** → Predmet.

Teda predmet **funkčne NEzávisí od kľúča** idStud.

⇒ Dekompozícia (zrušenie duplicit; Student2 už bude v 2NF)

Student2

<u>idStud</u>	Ucitel	UciKabinet
1	A	101
2	B	102

VyberPredmetov

idStud	Predmet
1	DBS1
1	DBS2
1	MPJ
2	PAZ1
2	PAZ2
2	Siete

Zatiaľ obe tabuľky sú problematické.

Riešme najprv ľavú tabuľku. Aby ona bola v 3. NF, musíme zrušiť atribúty, ktoré závisia nie od kľúča: UciKabinet závisí od Ucitel-a.

⇒ Dekompozícia

Pozrime sa teraz na druhú tabuľku. Môže byť Predmet kľúčom? Čo ak DBS1 zapíše aj ďalší študent? Teda čo bude kľúčom, ktorá najmenšia množina atribútov určuje jednoznačne riadok?

Každá z daných troch tabuliek už bude v 3. NF:

Student3

Ucitel

VyberPredmetov

<u>idStud</u>	<u>Ucitel</u>
1	A
2	B
1	C

<u>Ucitel</u>	<u>UciKabinet</u>
A	101
B	102

<u>idStud</u>	<u>Predmet</u>
1	DBS1
1	DBS2
1	MPJ
2	PAZ1
2	PAZ2
2	Siete
2	DBS1