

Databázové systémy

Funkčné závislosti pri normalizácii databáz

Normalizácia dát

- **Normalizácia** je proces eliminovania duplicít, nekonzistentností v dátach a závislostiach v relačných databázach pomocou
 - **pravidiel medzi atribútmi a funkčnými závislosťami**
 - **dekompozície** tabuľky (náhrada tabuľky s veľkým počtom stĺpcov s viacerými tabuľkami s menším počtom stĺpcov)s cieľom zabezpečiť integritu dát a úsporne šetriť miesto na ukladanie údajov.

- **Ciele:**
 - Eliminácia redundancie dát (duplicitného ukladania)
 - Minimalizácia rizika vzniku anomálií (Insert, Update, Delete)
 - Zabezpečenie logickej konzistentnosti dát
- **Pravidlo:** Uložiť každú jednu informáciu do databázy práve raz.

Normalizácia dát

- Normalizácia prebieha v po sebe idúcich krokoch, ktoré nazývame **Normálne formy (NF)**. Aby bola databáza považovaná za zdravú, musí v komerčnej praxi spĺňať prvé tri:
- **1. Normálna forma (1NF):** Všetky atribúty musia byť atomárne (nedeliteľné). V bunke nesmie byť zoznam hodnôt (napr. v stĺpci meno nemôže byť napísané 'Adam a Eva'). Každý riadok musí byť unikátny (musí existovať primárny kľúč).
- **2. Normálna forma (2NF):** Splnená 1NF + všetky neklúčové stĺpce musia byť plne **závislé** od celého primárneho kľúča (rieši sa najmä pri zložených kľúčoch).
- **3. Normálna forma (3NF):** Splnená 2NF + medzi neklúčovými stĺpcami nesmie existovať tranzitívna **závislosť**. To znamená, že žiadny neklúčový stĺpec nesmie závisieť od iného neklúčového stĺpca (napríklad, ak by ste mali v tabuľke stĺpce PSC a Mesto, porušujete 3NF, pretože názov mesta viete odvodiť z PSC a tieto dáta patria do samostatnej číselníkovej tabuľky).
- **Heat-Boyce-Coddova normálna forma (HBCNF):** Sprísňuje tretiu normálnu formu tým, že vyžaduje, aby každý stĺpec, ktorý určuje hodnotu iného stĺpca, bol striktne kandidátskym kľúčom.

Relačná databáza

- **Relačná databáza** je typ databázy, v ktorej sú údaje uložené v **navzájom prepojených tabuľkách** (stĺpcoch a riadkoch).
- **Vzťahy** medzi tabuľkami sa vytvárajú pomocou **klúčov** (napríklad primárny a cudzí kľúč).

Relačná schéma

- **Relačná schéma:** Predstavuje logický zápis štruktúry tabuľky.
- Zapisuje sa názvom relácie (tabuľky) a zoznamom jej atribútov (stĺpcov):
 - NAZOV_TABULKY (Atribút1, Atribút2, Atribút3, ...)
- **Názov tabuľky** (napr. *Študent*, *Faktúra*).
- **Atribúty** sú názvy stĺpcov uvedených v zátvorke.
- **Podčiarknutie** označuje atribút, ktorý slúži ako **primárny kľúč** (jednojednoznačný identifikátor riadku).
- Napr.
 - KNIHA(ISBN,nazov,autor,rok,vydavatel,cena)

Čo všetko relačná schéma definuje?

- **Názov relácie:** Jedinečné meno tabuľky v rámci databázy.
- **Zoznam atribútov:** Názvy všetkých stĺpcov, ktoré bude tabuľka obsahovať.
- **Dátové typy (domény):** Určuje, aký typ dát môže v stĺpci byť (napr. text, celé číslo, dátum).
- **Integritné obmedzenia:** Pravidlá pre dáta (napr. stĺpec nesmie byť prázdny – NOT NULL).
- **Kľúče:** Určenie primárneho kľúča (identifikátor) a cudzích kľúčov (prepojenia na iné tabuľky).

Relačná schéma

- Uvažujme **relačnú schému** $R(\mathcal{A})$, kde je množina atribútov
 - $\text{NazovTab}(A_1, A_2, \dots, A_n) \equiv \text{NazovTab}(A_1:D_1, A_2:D_2, \dots, A_n:D_n)$
- kde
- $A_1, A_2, \dots, A_n$ je množina všetkých atribútov
- $D_1, D_2, \dots, D_n$ sú domény atribútov
- **Relačná schéma R** = názov relácie, množina atribútov a ich domény
- **Relácia r** ako inštancia R (tabuľka) je množina n-tíc.

Funkčná závislosť

- **Definícia:** Nech X a Y sú množiny atribútov $\mathcal{A}$.
Hovoríme že, **Y funkčne závisí od X** vzhľadom na $\mathcal{A}$
(**X funkčne určuje Y** , resp. **X je determinant**) $X \rightarrow Y$,
ak pre ľubovoľné dve n -tice, v ktorých X atribúty sa rovnajú, platí, že
rovnajú sa aj Y atribúty, tzn.
 $t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$.

- Poznámky:
 - každej hodnote X zodpovedá práve jedna hodnota Y
 - výraz $X \rightarrow Y$ je **funkčná závislosť** (FZ)
 - funkčná závislosť $X \rightarrow Y$ je **triviálna**, ak Y je podmnožinou X

Funkčná závislosť - príklad

- Nech daná tabuľka reprezentuje súčasný stav relácie $R(A,B,C)$. Ktoré z nasledujúcich FZí platia v R na základe jej súčasného stavu?

- a) $A \rightarrow B$,
- b) $B \rightarrow A$
- c) $A \rightarrow C$
- d) $C \rightarrow A$
- e) $B \rightarrow C$
- f) $C \rightarrow B$

A	B	C
1	A	a
1	B	b
2	A	d
3	D	c
4	A	a
5	D	c

Vlastnosti funkčnej závislosti

- $A \rightarrow B_1 B_2 \dots B_k$ **spojovanie (union)**
- $A \rightarrow B_i, i=1,2,\dots,k$ **rozštiepenie (decomposition)**
- $A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_k$
& $B_1 B_2 \dots B_k \rightarrow C_1 C_2 \dots C_m \Rightarrow A_1 A_2 \dots A_n \rightarrow C_1 C_2 \dots C_m$

tranzitívnosť

Kandidátny kľúč

- **Def. 1a** Nech X je množina atribútov hlavičky R . X je **kandidátny kľúč** (pre R), ak:
 - a) X funkčne určuje všetky ostatné (neklúčové) atribúty R .
 - b) žiadna podmnožina X neurčuje funkčne ostatné atribúty.
- **Def. 1b Kandidátny kľúč** (pre r) je ľubovoľná množina stĺpcov X , ktoré majú v každom riadku jedinečnú kombináciu hodnôt a odstránením ľubovoľného stĺpca z X by sa vytvorili duplicitné kombinácie hodnôt.

Kandidátny kľúč - vlastnosti

- Relácia **r** ako inštancia **R** je množina (teda bez duplicitných n-tíc)
- Ak tabuľka má **kľúč**, potom v nej všetky riadky sú rozdielne
- (Kandidátny) kľúč pre reláciu R je **minimálna** množina atribútov X, pre ktorú platí $X \rightarrow \mathcal{A}$
- dva rozdielne riadky **R** sa nemôžu rovnať na **X**

Primárny, kompozitný kľúč, nadkľúč, podkľúč

- Relácia môže mať viac kandidátnych kľúčov – z nich sa vyberie jeden – **primárny kľúč**.
- **Kompozitný kľúč** (alebo **zložený kľúč**) je kandidátny kľúč, ktorý sa skladá z dvoch alebo viacerých atribútov.
- Hovoríme, že množina atribútov X je **podkľúč (PK)** / **nadkľúč (NK)**, ak X je podmnožinou / nadmnožinou **nejakého, aspoň jedného** kandidátneho kľúča, $X \subseteq KK_i$ / $KK_i \subseteq X$.

Určovanie kľúča

- **Atribúty:** Autor, NazovDiela, Rok, Vydavatel, PocetStran, Zaner.
- **Analýza determinancie:**
 - Dvojica **NazovDiela, Rok** určuje vydavateľa, strany aj žáner, ale *neurčuje* autora (dielo môže napísať viac autorov).
 - Samotná dvojica **NazovDiela, Autor** tiež nestačí (autor môže vydať rôzne diela v rôznych rokoch).
- **Výsledný kľúč:** Trojica **NazovDiela, Rok, Autor** spĺňa podmienku minimálnosti aj jednoznačného určenia všetkých ostatných atribútov.

Armstrongove pravidlá/axiómy

- Pomocou nasledujúcich pravidiel môžeme vytvárať nové funkčné závislosti alebo dokázať, že ide o funkčnú závislosť.

A1) $\mathbf{x \subseteq y \Rightarrow y \rightarrow x}$ reflexívnosť (triv.závislosť)

A2) $\mathbf{x \rightarrow y \Rightarrow \forall z: xz \rightarrow yz}$ augmentation

A3) $\mathbf{x \rightarrow y \ \& \ y \rightarrow z \Rightarrow x \rightarrow z}$ tranzitívnosť

Armstrongove pravidlá/axiómy

- sú **korektné** – funkčné závislosti F_i odvodené z F platia pre každú inštanciu R
- sú **úplné** – ľubovoľnú funkčnú závislosť F_i je možné odvodiť z F pre každú inštanciu R
- vlastnosti $A1, A2, A3$ sú **nezávislé** a bez hociktorej z nich úplnosť je narušená

Armstrongove pravidlá/axiómy

- **Dokážte:** $\{AB \rightarrow C, CD \rightarrow E\} \Rightarrow ABD \rightarrow E$
- **Riešenie:**
- $AB \rightarrow C$ - dané
- $ABD \rightarrow CD$ - A2)
- $CD \rightarrow E$ - dané
- $ABD \rightarrow E$ - A3)

Armstrongove pravidlá/axiómy

- 4) $(x \rightarrow y) \ \& \ (x \rightarrow z) \Rightarrow x \rightarrow yz$
- 5) $(x \rightarrow y) \ \& \ (wy \rightarrow z) \Rightarrow wx \rightarrow wz$
- 6) $(x \rightarrow y) \ \& \ (z \dot{\rightarrow} y) \Rightarrow x \rightarrow z$ – dekompozícia
- Dokážte, že platia.

Uzávery (Closures)

- **Funkčný uzáver (F^+):** Množina *všetkých* funkčných závislostí, ktoré logicky vyplývajú z množiny F pomocou Armstrongových pravidiel.
- **Atribútový uzáver (A^+):** Množina všetkých atribútov X , pre ktoré platí, že z A vieme odvoziť X (t.j. $A \rightarrow X$).
- **Praktické využitie:** Pomocou atribútového uzáveru vieme overiť, či daná FZ vyplýva z množiny pravidiel (testujeme, či cieľový atribút patrí do uzáveru východiskových atribútov).

Algoritmus atribútového uzáveru

Algoritmus atribútového uzáveru:

a) Nech je $X = \{A\}$.

b) Cyklus: nech $B_i \in X, i = \overline{1, m}$,

hľadáme FZ $B_1, \dots, B_m \rightarrow C$ takú, že $C \in \mathcal{F}$, ale $C \notin X$

$\Rightarrow$ dodáme C do X , tzn. $X = X \cup C$, $X = \{A, B_1, \dots, B_m, C\}$.

c) Na konci $A^+ \equiv X$.

Príklad

Uvažujme relačnú schému $R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = (A, B, C, D, E, F)$ a $\mathcal{F} = \{AB \rightarrow C, BC \rightarrow AD, D \rightarrow E, CF \rightarrow B\}$. Určme $\{A, B\}^+$.

Bezstratová dekompozícia (Lossless-Join)

- **Čo je dekompozícia?** Rozdelenie jednej schémy R na dve menšie schémy $\{R_1, R_2\}$, pričom $R = R_1 \cup R_2$.
- **Bezstratová dekompozícia:** Pri spätnom prirodzenom spojení (Natural Join) tabuliek nestratíme ani nepridáme žiadne pôvodné dáta ($r[R_1] \bowtie r[R_2] = r$).
- Dekompozícia $\{R_1, R_2\}$ je **stratová**, ak $R_1 \bowtie R_2 \supset R$.
- Poznámka: prípad $R_1 \bowtie R_2 \subset R$ nemôže nastať.
- **Kritérium bezstratovosti:** Dekompozícia je bezstratová práve vtedy, ak prienik schém tvorí nadkľúč pre aspoň jednu z rozdelených schém:
 - $(R_1 \cap R_2) \rightarrow R_1 \in F^+$ alebo $(R_1 \cap R_2) \rightarrow R_2 \in F^+$

Príklady

Príklad 1:

Nech $R \equiv R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A, B, C\}$ a $\mathcal{F} = \{A \rightarrow B\}$.

Ukážme, že rozklad $\{\mathcal{A}_1, \mathcal{A}_2\}$, kde $\mathcal{A}_1 = \{A, B\}$ a $\mathcal{A}_2 = \{A, C\}$, je bezstratový (vzhľadom na $\mathcal{F}$).

Pretože $\mathcal{A}_1 \cap \mathcal{A}_2 = A$, a A je KK v $\mathcal{A}_1$, máme $(\mathcal{A}_1 \cap \mathcal{A}_2) = A \rightarrow \mathcal{A}_1 \in \mathcal{F}^+$, lebo $A \rightarrow A$ a $A \rightarrow B$, teda $A \rightarrow AB \Leftrightarrow A \rightarrow \mathcal{A}_1$, a $\mathcal{F}^+ = \mathcal{F}$.

Príklad 2:

Nech $R \equiv R(\mathcal{A})$, kde $\mathcal{A} = \{A, B\}$. Uvažujme rozklad $\{\mathcal{A}_1, \mathcal{A}_2\}$, kde $\mathcal{A}_1 = \{A\}$ a $\mathcal{A}_2 = \{B\}$.

Dá sa tu rozprávať o bezstratovej dekompozícii?

Nie, lebo kvôli prirodzenému spojeniu musí existovať jeden spoločný stĺpec!

Interpretácia

- Interpretujte $R_1 \cup R_2$, $R_1 \times R_2$ a $R_1 \bowtie R_2$.

Normalizácia do 2NF, 3NF, HBCNF

Definícia 1

Uvažujme relačnú schému $R(\mathcal{A}, \mathcal{F})$. Ak pre ľubovoľnú (aj implicitnú) **netriviálnu** funkčnú závislosť

$$X \rightarrow Y,$$

vyplývajúcu z $\mathcal{F}$, kde $X \subset \mathcal{A}$, $Y \subset \mathcal{A}$, platí

jedna vlastnosť z troch - X je nadkľúč - NK, - Y je podkľúč – $Y \subseteq \text{KK}_i$, - X nie je podkľúč – $X \not\subseteq \text{KK}_j$, potom a len potom je R v 2NF ,	jedna vlastnosť z dvoch - X je nadkľúč, - Y je podkľúč – $Y \subseteq \text{KK}_i$, potom a len potom je R v 3NF ,	Že - X je nadkľúč, potom a len potom je R v HBCNF .
---	---	--

Príklad

Uvažujme štyri príklady s tromi atribútmi (v DB teórii AB je množina zjednotení atribútov A, B).

Určte v ktorej NF sú jednotlivé R:

- 1) $R = R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A, B, C\}$ a $\mathcal{F} = \{A \rightarrow B\}$;
- 2) $R = R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A, B, C\}$ a $\mathcal{F} = \{A \rightarrow B, B \rightarrow C\}$;
- 3) $R = R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A, B, C\}$ a $\mathcal{F} = \{AC \rightarrow B, B \rightarrow C\}$;
- 4) $R = R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A, B, C\}$ a $\mathcal{F} = \{A \rightarrow BC\}$.

Príklad

Nech $R \equiv R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A, B, C, D\}$, $\mathcal{F} = \{ACD \rightarrow B, AC \rightarrow D, AC \rightarrow B, D \rightarrow C\}$.

2a) Skontrolujte, že $\{AC\}^+ = \mathcal{A}$ a preto AC je KK. Existuje ďalší dvojatribútový KK?

2b) Bude R v HBCNF?

2c) Bude R v 3NF?