

OBSAH SEMESTRA

SQL Server

<https://db-engines.com/en/ranking> <https://www.sqlservercentral.com/articles/t-sql-language-changes-in-sql-server-2022>

- 1) Funkčné závislosti, uzávery a bezstrat. závislosť**
- 2-3) Normálne formy, najnovšia NF - ETNF**
- 3-4) Úvod do SQL Server; Množ.op.; Doč.t; WHILE, Uložené procedúry, funkcie**
- 5) Pohľady; CTE, rekurzia a transitívny uzáver**
- 6) Transakcie; Kurzory; Pivot;**
- 7) Triggery; B-stromy a indexy**
- 8) XML, JSON; Window funkcie**

MongoDB

- 9) Big data a NoSQL - Úvod do MongoDB**
- 10) CRUD a kurzory**
- 11) Agregácie a indexy**
- 12) Replikácia a sharding**

Funkčná závislosť, uzávery a bezstratová dekompozícia

- Funkčná závislosť (FZ) a nadkľúč (H, R, r)
- Funkčný a atribútový uzáver
- Bezstratová dekompozícia

Funkčná závislosť (FZ) a nadkľúč - H, R, r

Uvažujme relačnú schému $R(\mathcal{A})$, kde $\mathcal{A}$ je množina atribútov

$$\text{NazovTab}(A_1, A_2, \dots, A_n) \equiv \text{NazovTab}(A_1:T_1, A_2:T_2, \dots, A_n:T_n)$$

- **Hlavička H** je množina všetkých atribútov. **Relačná schéma R** = {názov relácie, množina atribútov a ich typy}. **Relácia r** ako **inštancia** R (tabuľka) je množina n-tíc. Hlavičku má R aj r.

Definícia

Nech X a Y sú množiny atribútov H. Hovoríme že,

Y **funkčne závisí od X** vzhľadom na H (X funkčne určuje Y; 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
- t_i je tuple/ntica/riadok tabuľky a $t_i[X]$ je priemet riadku t_i na množinu atribútov X
- AK t je n-tica/tuple/riadok tabuľky r, potom $t[X]$ je obmedzenie t na množinu X. Teda $t[X]$ je tiež n-tica, ktorá sa nazýva **projekcia** n-tice t na X, potom $t_i[X]$ je projekcia riadku t_i na množinu atribútov X.

FZ $X \rightarrow A$ sa považuje za funkciu, ktorá spája s každou hodnotou X nejakú jedinečnú hodnotu A a $t[X]$ jedinečne určuje hodnotu $t(A)$.

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

Interpretácia. Uvažujme $R\{\mathcal{A}, \mathcal{F}\}$ s $\mathcal{A} = \{D, S, P\}$ a $\mathcal{F} = \{DS \rightarrow P\}$,

kde D-dodávateľ, S-súčiastka, P-projekt. **Potom $DS \rightarrow P$ znamená:**

ľubovoľný daný dodávateľ dodáva D danú súčiastku S najviac do jedného projektu P.

Vlastnosti (podrobnejšie pozri nižšie)

$A \rightarrow B_1B_2\dots B_k \Leftrightarrow A \rightarrow B_i, i=1,2,\dots,k$ Treba dokázať v oboch smeroch (nižšie)

rozštiepenie $\Rightarrow$ **$\Leftarrow$ spojovanie**

Transitivnosť: $A_1A_2\dots A_n \rightarrow B_1B_2\dots B_k$ & $B_1B_2\dots B_k \rightarrow C_1C_2\dots C_m \Rightarrow A_1A_2\dots A_n \rightarrow C_1C_2\dots C_m$

Def. 1a Nech X je množina atribútov hlavičky R. X je **kandidátny kľúč** (pre R), ak:

- X funkčne určuje všetky ostatné (neklúčové) atribúty R.
- 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.

Poznámky:

- (**Kandidátny**) kľúč pre reláciu R je **minimálna** množina atribútov X, pre ktorú platí $X \rightarrow \mathcal{A}$ resp. R
- v Def. 1b sa predpokladá, že relácia r ako inštancia R je množina (teda bez duplicitných n-tíc)
- ak tabuľka má kľúč, potom v nej všetky riadky sú rozdielne
- z bodu a) def. 1a vyplýva, že dva rozdielne riadky R sa nemôžu rovnať na X. Skutočne, keby sa rovnali na X, potom by sa museli rovnať aj v iných atr. a preto dva riadky by boli rovnaké – protirečenie.

Relácia môže mať viac kandidátnych kľúčov – z nich sa vyberie jeden – **primárny kľúč**.

Def. 2 Hovoríme, že množina atribútov X je **podkľúč (PK)** / **nadkľúč (NK)**, ak X je podmnožinou / nadmnožinou **nejakého, aspoň jedného** kandidátneho kľúča, $X \subseteq KK_i$ / $KK_k \subseteq X$.

Poznámky:

- **nadkľúč** (superkľúč) je ľubovoľná nadmnožina kľúča.
- Pre nadkľúč X platí: relácia nemá dva rozdielne riadky s rovnakými hodnotami pre atribúty z X .

A	B	C	D
1	Z	10	A
1	Z	20	B
	L	10	B
2	L	30	B
3	L	10	A

Je $\{A,C\}$ nadkľúč?

- $A \rightarrow B$
- $A,C \rightarrow B$
- $A,C \rightarrow B,D$

Príklad. Kompozitný kľúč

Relácia: dielo

Autor	NazovDiela	Rok	Vydavatel	PocetStran	Zaner
J.Austin	Sense and Sensibility	1811	T.Egerton	320	Romance
A.C.Clark	Cradle	1988	Warner Books	293	Scifi
G.Lee	Cradle	1988	Warner Books	293	Scifi
A.Hejlsberg	The C# Programming Language	2003	Addison Wesley	672	Software
S.Wiltamuth	The C# Programming Language	2003	Addison Wesley	672	Software
P.Golde	The C# Programming Language	2003	Addison Wesley	672	Software
A.C.Clark	XZY	1988			
A.C.Clark	Cradle	??			

- Autor môže napísať viac* diel s tým istým názvom (ale nie v tom istom roku!)
- Nové vydanie (dotlač) uvažujeme za to isté dielo
- (Dielo bez autora – Biblia?)

Úloha: Nájdime kľúč pre dielo!

Uvažujme za determinanta najprv **dvojice**.

NazovDiela, Rok jednoznačne určia aj Vydavateľa, aj PocetStran, aj Zanera (spojovanie):

- NazovDiela, Rok $\rightarrow$ Vydavateľ
- NazovDiela, Rok $\rightarrow$ PocetStran $\Leftrightarrow$ NazovDiela, Rok $\rightarrow$ Vydavateľ, PocetStran, Zaner
- NazovDiela, Rok $\rightarrow$ Zaner

Lenže:

- NazovDiela, Rok, **neurčuje** Autora, lebo niektoré diela majú viac autorov

Podobne:

- Rok, Autor **neurčuje** NazovDiela, lebo autor v danom roku môže napísať rôzne diela
- NazovDiela, Autor **neurčuje** Rok, lebo podľa predpokladu * Autor s tým istým názvom môže napísať viac diel (v rôznych rokoch).

Ale **trojica** NazovDiela, Rok, Autor je determinat.

NazovDiela, Rok, Autor - je kľúč.

Skutočne:

Vyššie sme ukazali, že

- trojica NazovDiela, Rok, Autor určuje všetky atribúty - je determinat,
- žiadna jej podmnožina neurčuje funkčne ostatné atribúty.

Existuje aj iný kľúč?

Pravidlá pre funkčné závislosti

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

Uvažujme relačnú schému $R \equiv R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A} = \{A_1, \dots, A_n\}$, $\mathcal{F} = \{FZ_1, \dots, FZ_k\}$.

Vlastnosti funkčných závislostí (**Armstrongove pravidlá**)

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

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

A3) $x \rightarrow y \ \& \ y \rightarrow z \Rightarrow x \rightarrow z$ tranzitívnosť

Pr. I) Nech $\{AB \rightarrow C, CD \rightarrow E\} \Rightarrow ABD \rightarrow E$

Riešenie:

$AB \rightarrow C$ - dané

$ABD \rightarrow CD$ - A2)

$CD \rightarrow E$ - dané

$ABD \rightarrow E$ - A3)

II) $(x \rightarrow y) \ \& \ (x \rightarrow z) \Rightarrow x \rightarrow yz$

III) $(x \rightarrow y) \ \& \ (wy \rightarrow z) \Rightarrow wx \rightarrow wz$

IV) $(x \rightarrow y) \ \& \ (z \subseteq y) \Rightarrow x \rightarrow z$ - dekompozícia

Veta. Armstrongove pravidlá

- sú **korektné** – nové funkčné závislosti F_i

odvodené z A1-A3 platia pre každú inštanciu R

- sú **úplné** – ľubovoľnú funkčnú závislosť F_i je

možné odvodiť z A1-A3 pre každú inštanciu R

- vlastnosti A1-A3 sú **nezávislé** a bez

hoci ktorej z nich úplnosť je narušená

Dôkaz: II) pomocou Armstr.

a) $x \rightarrow y \Rightarrow x \rightarrow xy$ podľa A1), A2)

b) $x \rightarrow z \Rightarrow xy \rightarrow yz$ podľa A2)

c) $x \rightarrow yz$ podľa a), b), A3)

Na cvičení dokázať 5), 6).

Dokážme A2): 1) na základe definície, 2) **sporom**.

Kontrapríklad, že v A2) z pravej strany nevyplýva
ľavá strana: $xz \rightarrow yz$ platí, ale $x \rightarrow y$ nie.

X	Z	Y
1	A	γ
1	A	γ
1	B	α

Ďalší dôkaz

Príklad. Dokážte pomocou definície a aj pomocou Armstrongových pravidiel

$$X \rightarrow YZ \Rightarrow X \rightarrow Y \ \& \ X \rightarrow Z$$

Podľa definície: uvažujme

$$[X_1, Y_1, Z_1]$$

$$[X_2, Y_2, Z_2]$$

Keďže $X_1 = X_2$, tak $Y_1Z_1 = Y_2Z_2$, potom platí, že ak $X_1 = X_2$, tak $Y_1 = Y_2$ a zároveň ak $X_1 = X_2$, tak $Z_1 = Z_2$.

Podľa Armstrongových pravidiel:

$$X \rightarrow YZ \xrightarrow{A1, A3} Y \Rightarrow X \rightarrow Y$$

$$X \rightarrow YZ \xrightarrow{A1, A3} Z \Rightarrow X \rightarrow Z$$

Uzávery (Closures)

Funkčný uzáver F^+

Uvažujme relačnú schému $R(\mathcal{A}, \mathcal{F})$, kde $\mathcal{A}$ a $\mathcal{F}$ sú množiny atribútov a FZí, a nech FZ $F \subseteq \mathcal{F}$.

Funkčný uzáver F^+ je množina všetkých FZí, ktoré vyplývajú z F pomocou Armstrongových pravidiel.

$$F^+ = \{\text{všetky } X \rightarrow Y, \text{ ktoré vyplývajú z } F\}$$

Podobne sa definuje funkčný uzáver $\mathcal{F}^+$: $\mathcal{F}^+ = \{\text{všetky } X \rightarrow Y, \text{ ktoré vyplývajú z } \mathcal{F}\}$

Korektnosť ArmPrav: ak $X \rightarrow Y$ je odvodená z $\mathcal{F}$ (pomocou ArmPrav), potom $X \rightarrow Y \in \mathcal{F}^+$.

Úplnosť ArmPrav: ak $X \rightarrow Y \in \mathcal{F}^+$, potom $X \rightarrow Y$ je odvodená z $\mathcal{F}$ (pomocou ArmPrav).

Atribútový uzáver $A^+ \equiv A^+ | \mathcal{F}$

Uvažujme relačnú schému $R(\mathcal{A}, \mathcal{F})$ a množinu atribútov $A \subseteq \mathcal{A}$.

Uzáver $A^+ \equiv A^+ | \mathcal{F}$ **atribútov** A vzhľadom na množinu FZí $\mathcal{F}$ je množina všetkých atribútov $X \subseteq \mathcal{A}$, pre ktoré FZ $A \rightarrow X$ logicky vyplýva (teda napr. na základe Arm. pravidiel) z $\mathcal{F}$, $A^+ = \{A | F \models X \rightarrow A\}$.

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 $C \notin X$

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

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

Príklad 1a). 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 $\{A, B\}^+$.

Riešenie.

a) Namiesto $X = \{AB\}$ môžeme uvažovať aj $X = \{A, B\}$, lebo $AB \rightarrow A$ a $AB \rightarrow B$.

b) Pretože ľavá strana $AB \rightarrow C$ je v X , preto pravú stranu dodáme do X , $X = \{A, B, C\}$.

Pretože ľavá strana $BC \rightarrow AD \rightarrow D$ je v X , preto časť pravej strany dodáme do X , $X = \{A, B, C, D\}$.

$D \rightarrow E \Rightarrow X = \{A, B, C, D, E\}$.

c) $CF \rightarrow B$ nic. $\Rightarrow \{A, B\}^+ = \{A, B, C, D, E\}$.

Všeobecný princíp, pomocou ktorého môžeme stanoviť, či z jednej FZ₁ vyplýva druhá FZ₂:

Uvažujme reláciu $R(\mathcal{A}, \mathcal{F})$, $F_1 \subseteq \mathcal{F}$ a $F_2: A \rightarrow B$, kde $A \in \mathcal{A}$.

Ak $B \in A^+$, potom z F_1 vyplýva F_2 .

Teda

$$\boxed{\begin{array}{l} B \in A^+ \\ \Rightarrow A \rightarrow B \end{array}}$$

Platí aj silnejšia ekvivalencia:

$B \in A^+ \Leftrightarrow$ z F_1 vyplýva F_2 .

$\Leftrightarrow$

Lema: $x \rightarrow y$ sa dá odvodiť z F pomocou Armstrongových axióm práve vtedy keď

$y \subseteq x^+$ vzhľadom na F .

Pre každý atribút $a \in y \subseteq x^+$. Platí $x \rightarrow a$ podľa definície uzáveru x^+ . Podľa vlastnosti 4) platí aj $x \rightarrow y$.

Naopak nech $x \rightarrow y$ sa dá odvodiť. Potom pre každé $a \in y$ platí $x \rightarrow a$ podľa 6) a $a \in x^+$.

Príklad 1b). Pokračovanie predchádzajúceho príkladu. Vyplývajú z $\mathcal{F}$ FZi: $D \rightarrow A$ a $AB \rightarrow D$?

Riešenie.

Závislosť $D \rightarrow A$ nevyplýva z $\mathcal{F}$, pretože $\{D\}^+ = \{D, E\}$ a A nie je z $\{D\}^+$.

$AB \rightarrow D$ vyplýva, lebo D je z $\{A, B\}^+ = \{A, B, C, D, E\}$.

https://en.wikipedia.org/wiki/Lossless_join_decomposition

$$r_1 \bowtie r_2,$$

$$\Leftrightarrow \text{SELECT } * \text{ FROM } r1 \text{ JOIN } r2 \text{ ON } r1.id = r2.id$$

relačnými schémami R_1 a R_2), ak $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$. Používa sa aj označenie $\{\mathcal{A}_1, \mathcal{A}_2\}$, či $R = R_1 \cup R_2$.

spĺňa $\mathcal{F}$, platí

teda $\pi_{A1}(r) \propto \pi_{A2}(r) = r$, kde $A_i \equiv \mathcal{A}_i$, $i=1, 2$,

čo jednoducho označujeme ako

$$\mathbf{r}_1 \boxtimes \mathbf{r}_2 = \mathbf{r},$$

či $R_1 \bowtie R_2 = R$ alebo ako aj $\mathcal{A}_1 \bowtie \mathcal{A}_2 = \mathcal{A}$.

Formálne sa **bezstratovosť** definuje pomocou relácie r , ale význam je nezávislý od konkrétnych dát, musí to platiť pre všetky možné relácie nad schémou R .

Kým $t[X]$ ako n -tica sa nazýva **projekcia** n -tice t na X , $r[\mathcal{A}_i]$ sa nazýva **projekcia** relácie r na $\mathcal{A}_i$.

Def 3. 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. Dekompozícia je bezstratová, ak aspoň jedna z nasledujúcich funkčných závislostí je z $\mathcal{F}^+$

- $(\mathcal{A}_1 \cap \mathcal{A}_2) \rightarrow \mathcal{A}_1 \in \mathcal{F}^+$
- $(\mathcal{A}_1 \cap \mathcal{A}_2) \rightarrow \mathcal{A}_2 \in \mathcal{F}^+$

teda spoločný(é) atribút(y) je (sú) nadkľúčom buď pre R_1 alebo R_2 .

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 $(\mathcal{A}_1 \cap \mathcal{A}_2) = A \rightarrow \mathcal{A}_1$, lebo $\{A \rightarrow A, A \rightarrow B\} \in \mathcal{F}^+$.

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!

Příklad 3 na cvičení: Uvažujme rozklad $\{T_1, T_2\}$ tabulky T $\begin{array}{c|c} 1 & 1 \\ \hline 1 & 2 \end{array}$, kde $T_1 = \{a\}$ a $T_2 = \{a, b\}$.

	a	b
a	1	1
b	1	2
c	2	1

- a) Presvedčte sa, že rozklad bude stratový (SQL kódom a potom aj pomocou kritéria).
b) Zmeňte jednu hodnotu v tabuľke T, aby daný rozklad bol bezstratový (je prienik NK?).

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

Príklad bezstratovej a stratovej dekompozície tabuľky

(<https://onecompiler.com/sqlserver>)

```
-- SQL Server
-- Uvazujme tabuľku ABC
drop table if exists ABC
create table ABC(A int, B int, C int)
insert ABC values
( 1, 2, 3),
(31, 2, 33)
select * from ABC
```

Navrhните

```
-- a) Lossless/bezstratovu dekompoziciu tabuľky ABC
-- b) Lossy/stratovu dekompoziciu tabuľky ABC
```

-- Riešenie

```
-- a) Lossless/bezstratovu dekompoziciu tabuľky ABC
```

```
--  $R(ABC) = R(AB) + R(AC)$ 
```

```
drop table if exists AB
select A,B into AB from ABC
--select * from AB
```

```
drop table if exists AC
select A,C into AC from ABC
--select * from AC
```

```
select "Prázdna tabuľka:"
select x.A a, B, C from AB x JOIN AC y ON x.A=y.A
except
select * from ABC -- prázdna tabuľka!
select "- prázdna tabuľka."
```

```
-- b) Lossy/stratovu dekompoziciu tabuľky ABC
```

```
--  $R(ABC) \neq R(AB) + R(BC)$ 
```

```
drop table if exists BC
select B,C into BC from ABC
--select * from BC
```

```
select "Neprázdna tabuľka:"
select A, x.B b, C from AB x JOIN BC y ON x.B=y.B
except
select * from ABC -- tabuľka nie je prázdna
select "- neprázdna tabuľka."
```