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 $R(\mathscr{A})$, kde $\mathscr{A}$ je množina atribútov

NazovTab(A_1, A_2, ..., A_n) ≡ NazovTab(A_1:T_1, A_2:T_2, ..., 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.

$t_1[X] = t_2[X] \quad \Rightarrow \quad t_1[Y] = t_2[Y]$.

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

AK t je n-tica, 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.
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).

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

Príklad. Nech daná tabuľka reprezentuje súčasný stav relácie R(A,B,C). Ktoré z nasledujúcich FZí platia v R na základe jej súcasného stavu:
a) $A \rightarrow B$, b) $B \rightarrow A$, c) $A \rightarrow C$, d) $C \rightarrow A$ e) $B \rightarrow C$, f) $C \rightarrow B$

Interpretácia. Uvažujme $R\left\{\mathscr{A}, \mathscr{F}\right\}$ s $\mathscr{A} = \{D, S, P\}$ a $\mathscr{F}$ ={ $DS \rightarrow P$ }, 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)
$A \rightarrow B_1B_2...B_k \quad \Leftrightarrow \quad A \rightarrow B_i$, i=1,2,...,k   Treba dokázať v oboch smeroch (nižššie)
rozštiepenie => <= spojovanie
Tranzitívnosť: $A_1A_2...A_n \rightarrow B_1B_2...B_k$ & $B_1B_2...B_k \rightarrow C_1C_2...C_m$ => $A_1A_2...A_n \rightarrow C_1C_2...C_m$

Def. 1a Nech X je množina atribútov hlavičky R. X je kandidátny kľúč (pre R), ak:
a) X funkčne <u>určuje všetky ostatné</u> (nekľúčové) atribúty R.
b) žiadna <u>podmnožina</u> X <u>neurčuje</u> 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:
1. v def. 1b sa predpokladá, že relácia r ako inštancia R je množina (teda bez duplicitných n-tíc)
2. <u>ak tabuľka má kľúč, potom v nej všetky riadky sú rozdielne</u>
3. <u>(Kandidátny) kľúč</u> pre reláciu R je <u>minimálna</u> množina atribútov X, pre ktorú platí $X \rightarrow \mathscr{A}$ resp. R
4. z bodu a) def. 1a vyplýva, že <u>dva rozdielne riadky R sa nemôžu rovnať na X</u>. 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 mat 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, $X \subseteq KK_i$ / $KK_i \subseteq X$.

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

{A,C} je 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) uvažovať za to isté dielo
- (Dielo bez autora – Biblia?)

Nájdime kľúč! Uvažujme za determinanta najprv dvojice.
NazovDiela, Rok jednoznačne určia aj Vydavatela, aj PocetStran, aj Zaner (spojovanie):
- NazovDiela, Rok → Vydavatel
- NazovDiela, Rok → PocetStran ⇔ NazovDiela, Rok → Vydavatel, PocetStran, Zaner
- NazovDiela, Rok → Zaner

Lenže:
- NazovDiela, Rok, **neurčuje** Autor, 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).

Riešením je trojica:
NazovDiela, Rok, Autor - je kľúč.
- najprv ukážeme, že trojica určuje všetky ostatné. Predpokladajme, že dva riadky na trojici sú <u>rovnaké</u> => sú rovnaké aj na dvojici NazovDiela, Rok – lenže ona, ako sme videli vyššie, urči Vydavatel, PocetStran, Zaner (ale Autor nie) a preto dva riadky sú <u>rovnaké</u> (vo všetkých atribútoch), čo je protirečenie s 3.
- treba ešte ukázať, že žiadna podmnožina trojice neurčuje funkčne ostatné atribúty (to sme urobili vyššie).

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 $R \equiv R(\mathscr{A}, \mathscr{F})$, kde $\mathscr{A} = \{A_1,...,A_n\}$, $\mathscr{F} = \{FZ_1,...,FZ_k\}$.

Vlastnosti funkčných závislostí (Armstrongove pravidlá)

A1) $\mathbf{x} \subseteq \mathbf{y} \Rightarrow \mathbf{y} \to \mathbf{x}$ reflexívnosť (triv.závislosť)
A2) $\mathbf{x} \to \mathbf{y} \Rightarrow \forall \mathbf{z}: \mathbf{xz} \to \mathbf{yz}$ augmentation
A3) $\mathbf{x} \to \mathbf{y}$ & $\mathbf{y} \to \mathbf{z} \Rightarrow \mathbf{x} \to \mathbf{z}$ tranzitívnosť

Pr. Nech $\{AB \to C, CD \to E\} \Rightarrow ABD \to E$
Riešenie:
$AB \to C$ - dané
$ABD \to CD$ - A2)
$CD \to E$ - dané
$ABD \to E$ - A3)

4) $(\mathbf{x} \to \mathbf{y})$ & $(\mathbf{x} \to \mathbf{z}) \Rightarrow \mathbf{x} \to \mathbf{yz}$
5) $(\mathbf{x} \to \mathbf{y})$ & $(\mathbf{wy} \to \mathbf{z}) \Rightarrow \mathbf{wx} \to \mathbf{wz}$
6) $(\mathbf{x} \to \mathbf{y})$ & $(\mathbf{z} \subseteq \mathbf{y}) \Rightarrow \mathbf{x} \to \mathbf{z}$ - dekompozícia

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

Dôkaz: 4) pomocou Armstr.
a) $\mathbf{x} \to \mathbf{y} \Rightarrow \mathbf{x} \to \mathbf{xy}$ podľa A1), A2)
b) $\mathbf{x} \to \mathbf{z} \Rightarrow \mathbf{xy} \to \mathbf{yz}$ podľa A2)
c) $\mathbf{x} \to \mathbf{yz}$ 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 – $\mathbf{xz} \to \mathbf{yz}$ platí, ale $\mathbf{x} \to \mathbf{y}$ nie.

X	Z	Y
1	A	γ
1	A	γ
1	B	α

Dôkazy

Príklad. Dokážte pomocou definície a Armstrongových pravidiel

$$X \longrightarrow YZ \Longrightarrow X \longrightarrow Y \ \& \ X \longrightarrow Z$$

Podľa definície: uvažujme

$$[X_1, Y_1, Z_1]$$
$$[X_2, Y_2, Z_2]$$

Keďže $X_1 = X_2$, tak $Y_1 Z_1 = Y_2 Z_2$, 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:

$$X \longrightarrow YZ \underset{A1,}{\longrightarrow} Y \underset{A3,}{\Longrightarrow} X \longrightarrow Y$$
$$X \longrightarrow YZ \underset{A1,}{\longrightarrow} Z \underset{A3,}{\Longrightarrow} X \longrightarrow Z$$

Uzávery

Funkčný uzáver F^+

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

Funkčný uzáver F^+ je množina všetkých FZí, ktoré vyplývajú z F pomocou Armstrongových pravidiel.
$F^+ = \{$všetky $X \to Y$, ktoré vyplývajú z $F\}$

Podobne: $\mathcal{F}^+ = \{$všetky $X \to Y$, ktoré vyplývajú z $\mathcal{F}\}$

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

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

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

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

Algoritmus atribútového uzáveru:
a) Nech je $X = \{A\}$.
b) Cyklus: nech $B_i \in X$, $i = \overline{1, m}$,
hľadáme FZ $B_{1}, \ldots, B_m \to C$ takú, že $C \in \mathcal{F}$, ale $C \notin X$
=> dodáme C do X, tzn. $X = X \cup C$, X={ A, $B_1, \ldots, B_m$, 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→C, BC→AD, D→E, CF→B}. Určme $\{A,B\}^+$.
Riešenie.
a) Namiesto X={AB} môžeme uvažovať aj X={A,B}, lebo z AB→A a z AB→B.
b) Pretože ľavá strana AB→C je v X, preto pravú stranu dodáme do X, X={A,B,C}.
Pretože ľavá strana BC→AD→D je v X, preto časť pravej strany dodáme do X, X={A,B,C,D}.
D→E => X={A,B,C,D,E}.
c) CF→B nic. => $\{A,B\}^+$={A,B,C,D,E}.

Všeobecný princíp, pomocou ktorého môžeme stanoviť, či z FZ_1 vyplýva FZ_2:
Uvažujme reláciu $R(\mathcal{A}, \mathcal{F})$, $F_1 \subseteq \mathcal{F}$ a $F_2 : A \to B$, kde $A \in \mathcal{A}$.
Ak $B \in A^+$, potom z F_1 **vyplýva** F_2.
Teda

$$\boxed{\begin{array}{rl} & B \in A^+ \\ \Rightarrow & A \to B \end{array}}$$

Platí aj silnejšia ekvivalencia: $B \in A^+$ ⇔ z F_1 **vyplýva** F_2.
⇔
Lema: $\mathbf{x} \to \mathbf{y}$ sa dá odvodiť z $\mathbf{F}$ pomocou Armstrongových axióm práve vtedy keď
$\mathbf{y} \subseteq \mathbf{x}^+$ vzhľadom na $\mathbf{F}$.
Pre každý atribút a $\in \mathbf{y} \subseteq \mathbf{x}^+$. Platí $\mathbf{x} \to$ a podľa definície uzáveru $\mathbf{x}^+$. Podľa vlastnosti 4) platí aj $\mathbf{x} \to \mathbf{y}$.
Naopak nech $\mathbf{x} \to \mathbf{y}$ sa dá odvodiť. Potom pre každé a $\in \mathbf{y}$ platí $\mathbf{x} \to$ a podľa 6) a $\in \mathbf{x}^+$.
Príklad 1b). Pokračovanie predchádzajúceho príkladu. Vyplývajú z $\mathcal{F}$ FZi: D→A a AB→D?
Riešenie.
AB→D vyplýva, lebo D je z $\{A,B\}^+$={A,B,C,D,E}.
D→A nevyplýva, lebo A nie je z $\{D\}^+$={D,E}.

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

Pripomíname, že **natural join/prirodzené spojenie** v relačnej algebre je r1 ⋈ r2, kde

r1 ⋈ r2 ⇔ r1 $\bowtie_{r1.id = r2.id}$ r2
⇔ SELECT * FROM r1 JOIN r2 ON r1.id = r2.id

Def 1. Hovoríme, že pre relačnú schému $R \equiv R(\mathscr{A})$ dvojica $\{R_1, R_2\}$ je **dekompozícia/rozklad** (s dvomi relačnými schémami R_1 a R_2), ak $\mathscr{A} = \mathscr{A}_1 \cup \mathscr{A}_2$. Používa sa aj označenie $\{\mathscr{A}_1, \mathscr{A}_2\}$, či $R = R_1 \cup R_2$.

Def 2. Dekompozícia $\{R_1, R_2\}$, či $\{\mathscr{A}_1, \mathscr{A}_2\}$, je **bezstratová (rozklad s bezstratovým spojením -** lossless-join decomposition) vzhľadom na množinu FZí $\mathscr{F}$, ak pre každú inštanciu r schémy R, ktorá spĺňa $\mathscr{F}$, platí

$$r[\mathscr{A}_1] \bowtie r[\mathscr{A}_2] = r, \text{ teda } \pi_{A1}(r) \bowtie \pi_{A2}(r) = r, \text{ kde } Ai \equiv \mathscr{A}_i,\ i=1, 2,$$

čo jednoducho označujeme ako $r_1 \bowtie r_2 = r$ (či $R_1 \bowtie R_2 = R$ alebo ako aj $\mathscr{A}_1 \bowtie \mathscr{A}_2 = \mathscr{A}$).

Kým **dekompozícia** sa týka relačnej **schémy** R aj **relácie** r, **bezstratová** dekompozícia **relácie**/tabuľky r.
Kým t[X] ako n-tica sa nazýva **projekcia** n-tice t na X, r[$\mathscr{A}_i$] sa nazýva **projekcia** relácie r na $\mathscr{A}_i$.

Def 3. Dekompozícia $\{R1, R2\}$ 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 $\mathscr{F}^+$

- $(\mathscr{A}_1 \cap \mathscr{A}_2) \to \mathscr{A}_1 \in \mathscr{F}^+$
- $(\mathscr{A}_1 \cap \mathscr{A}_2) \to \mathscr{A}_2 \in \mathscr{F}^+$

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

Príklad 1:
Nech $R \equiv R(\mathscr{A}, \mathscr{F})$, kde $\mathscr{A} = \{A, B, C\}$ a $\mathscr{F} = \{A \to B\}$.
Ukážme, že rozklad $\{\mathscr{A}_1, \mathscr{A}_2\}$, kde $\mathscr{A}_1 = \{A, B\}$ a $\mathscr{A}_2 = \{A, C\}$, je bezstratový (vzhľadom na $\mathscr{F}$).
Pretože $\mathscr{A}_1 \cap \mathscr{A}_2 = A$, a A je KK v $\mathscr{A}_1$, máme $(\mathscr{A}_1 \cap \mathscr{A}_2) = A \to \mathscr{A}_1$, lebo $\{A \to A, A \to B\} \in \mathscr{F}^+$.

Príklad 2:
Nech $R \equiv R(\mathscr{A})$, kde $\mathscr{A} = \{A, B\}$. Uvažujme rozklad $\{\mathscr{A}_1, \mathscr{A}_2\}$, kde $\mathscr{A}_1 = \{A\}$ a $\mathscr{A}_2 = \{B\}$.
Dá sa tu rozprávať o bezstratovej dekompozícii?
Nie, lebo kvôli prirodzenému spojeniu musí existovať jeden spoločný stĺpec!

Príklad 3 na cvičení: Uvažujme rozklad $\{T_1, T_2\}$ tabuľky T

a	b
1	1
1	2
2	1

, kde $T_1 = \{a\}$ a $T_2 = \{a, b\}$.

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 $R_1 \bowtie R_2$.