OBSAH SEMESTRA

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

0) Relačná algebra
1) Funkčné závislosti, uzávery a bezstrat. závislosť
2) Normálne formy, najnovšia NF - ETNF
3) Úvod do SQL Server; Množ.op.; Doč.t; WHILE
4) 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
- Funkčný a atribútový uzáver
- Bezstratová dekompozícia

Funkčná závislosť (FZ) a nadkľúč

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)

- Relačná schéma = názov relácie, množina atribútov a ich typy.

Definícia

Nech X a Y sú <u>množiny</u> atribútov. Hovoríme že,

Y funkčne závisí od X (X funkčne určuje Y; X je determinant) $X \rightarrow Y$,

ak pre ľubovoľné dva riadky, v torých 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
- 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. Ak je t n-tica, potom 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

- 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) $C\rightarrow B$

Vlastnosti

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

rozštiepenie => <= spojovanie

Tranzitívnosť: $A_1A_2...A_n \rightarrow B_1B_2...B_k$ & $B_1B_2...B_k \rightarrow C_1C_2...C_m$ => $A_1A_2...A_n \rightarrow C_1C_2...C_m$

Kandidátny, nad a primárny kľúč

Def. 1a Nech X je množina atribútov. 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. 1a sa predpokladá, že relácia r ako inštancia R je množina (teda bez duplicitných riadkov)
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 $X \rightarrow R$
4. z a) 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 – <u>primárny kľúč</u>.

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.
- (Nech je daná relácia **R**(**A**). Potom množinu atribútov $\mathbf{X} \subseteq \mathbf{A}$ takú, že $\mathbf{X} \rightarrow \mathbf{A}$, nazývame nadkľúč.)
- Pre nadkľúč X platí: relácia nemá dva rozdielne riadky s rovnakými hodnotami pre atribúty z X.

A	B	C	D
1	Z	10	Aa
1	Z	20	Bb
	L	10	Cc
2	L	30	Dd
3	L	10	Cc

{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 $\rightarrow$ Vydavatel
- NazovDiela, Rok $\rightarrow$ PocetStran $\Leftrightarrow$ NazovDiela, Rok $\rightarrow$ Vydavatel, PocetStran, Zaner
- NazovDiela, Rok $\rightarrow$ 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 $\mathscr{A} = \{A_1, \ldots, A_n\}$, $\mathscr{F} = \{FZ_1, \ldots, FZ_k\}$.

Vlastnosti funkčných závislostí (Armstrongove pravidlá)

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

Pr. 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)

4) $(\mathbf{x} \rightarrow \mathbf{y})$ & $(\mathbf{x} \rightarrow \mathbf{z}) \Rightarrow \mathbf{x} \rightarrow \mathbf{yz}$
5) $(\mathbf{x} \rightarrow \mathbf{y})$ & $(\mathbf{wy} \rightarrow \mathbf{z}) \Rightarrow \mathbf{wx} \rightarrow \mathbf{wz}$
6) $(\mathbf{x} \rightarrow \mathbf{y})$ & $(\mathbf{z} \subseteq \mathbf{y}) \Rightarrow \mathbf{x} \rightarrow \mathbf{z}$ - 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) $\mathbf{x} \rightarrow \mathbf{y} \Rightarrow \mathbf{x} \rightarrow \mathbf{xy}$ podľa A1), A2)
b) $\mathbf{x} \rightarrow \mathbf{z} \Rightarrow \mathbf{xy} \rightarrow \mathbf{yz}$ podľa A2)
c) $\mathbf{x} \rightarrow \mathbf{yz}$ 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 – $\mathbf{xz} \rightarrow \mathbf{yz}$ platí, ale $\mathbf{x} \rightarrow \mathbf{y}$ nie.

X	Z	Y
1	A	γ
1	A	γ
1	B	α

Dôkazy

Def.: Nech máme schému $R(A_1, \ldots, A_n)$, kde $A_1, \ldots, A_n$ sú atribúty. Nech máme dve množiny atribútov $X, Y \in \mathcal{A}$, $\mathcal{A} = \{A_1, \ldots, A_n\}$. Hovoríme, že X funkčne určuje Y (Y funkčne určuje X), ak platí pre ľubovoľné dva riadky tabuľky R, že ak hodnoty x sú rovnaké, tak potom aj hodnoty y budú rovnaké.
Ozn. $X \longrightarrow Y$

Pre funkčnú závislosť platia ARMSTRONGOVÉ PRAVIDLÁ :
A1, $X \subseteq Y \implies Y \longrightarrow X$ reflexivita
A2, $X \longrightarrow Y, \forall Z : XZ \longrightarrow YZ$
A3, $X \longrightarrow Y \wedge Y \longrightarrow Z \implies X \longrightarrow Z$ tranzitivita

Príklady:
Zadanie: Dokážte pomocou definície:

1. $X \longrightarrow Y \wedge X \longrightarrow Z \implies X \longrightarrow YZ$

   Uvažujem ľubovoľné dva riadky:

$$[X_1\,,\ Y_1\,,\ Z_1\,,\ U_1]$$
$$[X_2\,,\ Y_2\,,\ Z_2\,,\ U_2]$$

Ak $X_1 = X_2$, tak $Y_1 = Y_2$ a zároveň ak $X_1 = X_2$, tak $Z_1 = Z_2$. Potom ale platí aj $Y_1Z_1 = Y_2Z_2$ za predpokladu $X_1 = X_2$, čo bolo potrebné dokázať.

(Podľa Armstrongových pravidiel : $X = XX \xrightarrow[X\rightarrow Y+A2,]{} YX \xrightarrow[X\rightarrow Z\,+A2,]{} YZ$)

2. $X \longrightarrow YZ \Longrightarrow X \longrightarrow Y \wedge X \longrightarrow Z$

   Podľa definície: uvažujme

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

   Keďže $X_1 = X_2$, tak $Y_1Z_1 = Y_2Z_2$, 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:

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

3. $A \longrightarrow B \wedge CB \longrightarrow D \Longrightarrow AC \longrightarrow D$

   Podľa definície: uvažujme

$$[A_1\,,\ C_1\,,\ B_1\,,\ D_1,\ U_1]$$
$$[A_2\,,\ C_2\,,\ B_2\,,\ D_2,\ U_2]$$

   Keďže $A_1 = A_2$, tak $B_1 = B_2$ a zároveň ak $C_1B_1 = C_2B_2$, tak $D_1 = D_2$ potom platí ak $A_1C_1 = A_2C_2$, tak $D_1 = D_2$

   Podľa Armstrongových pravidiel:

$$A \longrightarrow B \underset{A2,}{\Longrightarrow} AC \longrightarrow BC \longrightarrow D \underset{A3,}{\Longrightarrow} AC \longrightarrow D$$

Uzávery

Funkčný uzáver F^+

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

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

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

Korektnosť ArmPrav: ak X→Y je odvodená z $\mathscr{F}$, potom X→Y $\in \mathscr{F}^+$.
Úplnosť ArmPrav: ak X→Y $\in \mathscr{F}^+$, potom X→Y je odvodená z $\mathscr{F}$.

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

Uvažujme reláciu $R(\mathscr{A},\mathscr{F})$ a množinu atribútov $A \subseteq \mathscr{A}$.

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

Algoritmus atribútového uzáveru:
a) Nech je X = {A}.
b) Cyklus:

hľadáme FZ $B_1, ..., B_m \to C$ takú, že $C \in \mathscr{A}$, $B_i \in X$, $i = \overline{1,m}$, ale $C \notin X$

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

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

Príklad 1a). Uvažujme reláciu $R(\mathscr{A},\mathscr{F})$, kde $\mathscr{A}$ = (A,B,C,D,E,F) a

$\mathscr{F}$ ={AB→C, BC→AD, D→E, CF→B}. Určme $\{A,B\}^+$.

Riešenie.
a) Namiesto X={AB} môže byť definícia 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 $R(\mathscr{A},\mathscr{F})$, $F_1 \subseteq \mathscr{F}$ a $F_2 : A \to B$, kde $A \in \mathscr{A}$.

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

Teda

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

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 $\mathscr{F}$ 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

Natural Join relačnej algebry - $r1 \bowtie r2$

$$r1 \bowtie r2 \quad \Leftrightarrow \quad r1 \bowtie_{r1.id = r2.id} r2$$

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

Kvôli jednoduchosti množinu atribútov $\mathscr{A}$ relačnej schémy R budeme označovať ako A, $\mathscr{A} \equiv$ A.

Dekompozícia relačnej schémy R dvomi relačnými schémami $\{R_1, R_2\}$ je taký rozklad, že $A = A_1 \cup A_2$. Používajú sa aj označenia $\{A_1, A_2\}$, či $R = R_1 \cup R_2$.

Dekompozícia {R1, R2} je **bezstratová** (rozklad bezstratovým spojením - lossless-join decomposition) vzhľadom na množinu FZí $\mathscr{F}$, ak pre každú inštancia **r** z R, ktorá spĺňa $\mathscr{F}$ platí:

$$\pi_{A1}(r) \bowtie \pi_{A2}(r) = r, \text{ teda } \; r[A1] \bowtie r[A2] = r \,,$$

čo jednoducho označujeme ako $r_1 \bowtie r_2 = r$, $R_1 \bowtie R_2 = R$ alebo ako aj $A_1 \bowtie A_2 = A$.

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 $\mathscr{F}^+$

- $(A_1 \cap A_2) \to A_1 \in \mathscr{F}^+$
- $(A_1 \cap A_2) \to A_2 \in \mathscr{F}^+$

teda spoločný atribút(y) je nadkľúč buď pre R_1 alebo R_2 (podľa dohody $\mathscr{A}_i \equiv A_i$).

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 $\{R_1, R_2\}$, kde $R_1=\{A, B\}$ a $R_2=\{A, C\}$, je bezstratová (tu A, B, C sú štandardné atribúty).
Pretože $R_1 \cap R_2 = A$ a A je KK v R_1, máme $R_1 \cap R_2 \to R_1 \in \mathscr{F}^+$.

Príklad 2:
Nech $R \equiv R(\mathscr{A})$, kde $\mathscr{A} = \{a, b\}$. Uvažujme rozklad $\{T_1, T_2\}$, kde $T_1=\{a\}$ a $T_2=\{b\}$.
Dá sa tu rozprávať o bezstratovej dekompozícii?
`Nie, lebo kvôli prirodzenému spojeniu musí existovať jeden spoločný názov stĺpca!`

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

a	b
1	1
2	1
1	1

, kde $T_1=\{a\}$ a $T_2=\{a, b\}$.
a) Presvedčte sa, že rozklad bude stratový.
b) Zmeňte jednu hodnotu v tabuľke T, aby daný rozklad bol bezstratový.

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