

DBS – BI-WSI-SI-03

Normalizace relačního schématu a normální formy (první, druhá, třetí, BCNF).

Obsah

1	Normalizace	2
1.1	Funkční závislost	2
1.1.1	Příklad	2
1.1.2	Amstrongova pravidla	2
1.1.3	Klíč relace	2
2	Normální formy	2
2.1	BCNF - Boyce-Coddova normální forma	2
3	Normalizace	3
3.1	Beztrátová dekompozice	3

1 Normalizace

Cílem normalizace je minimalizovat redundanci data a aktualizací anomálie. Měla by vést k vzniku tabulek, které lze snadno udržovat.

1.1 Funkční závislost

X je množina všech atributů, Y a Z jsou $\subseteq X$. Z závisí funkčně na Y , jestliže ke každé Y hodnotě existuje maximálně jedna Z hodnota. Značí se jako $Y \rightarrow Z$.

1.1.1 Příklad

- ke každému kinu existuje nanejvýš jedna adresa $\Rightarrow$ $\text{nazev_kina} \rightarrow \text{adresa}$
 - název kina určuje atribut adresa
 - naopak adresa závisí na názvu kina
- pro každou dvojici kino, film existuje nejvýše jedno datum, kdy se film v kině vysílá $\Rightarrow$ $\text{nazev_kina}, \text{jmeno_filmu} \rightarrow \text{datum}$
 - dvojice jméno kina a jméno filmu určuje atribut datum
 - datum závisí na dvojici jmen

1.1.2 Armstrongova pravidla

- triviální funkční závislost = jestliže Y je $\subseteq X$, pak $X \rightarrow Y$
- tranzitivita = jestliže $X \rightarrow Y$ a $Y \rightarrow Z$, pak $X \rightarrow Z$
- kompozice pravé strany = $X \rightarrow Y$ a $X \rightarrow Z$, pak $X \rightarrow YZ$
- dekompozice pravé strany = $X \rightarrow YZ$, pak $X \rightarrow Y$ a $X \rightarrow Z$

1.1.3 Klíč relace

Řekněme $R(A)$, kde K je $\subseteq A$. Tak K je klíčem schématu $R(A)$, jestliže splňuje dvě vlastnosti:

- $K \rightarrow A$
- neexistuje $K' \subset K$ taková, že $K' \rightarrow A$

2 Normální formy

- 1NF - atributy jsou atomické (neexistují strukturované a vícehodnotové atributy)
- 2NF - žádný neklíčový atribut není závislý na části klíče (vždy závisí na celém)
- 3NF - jsou-li všechny neklíčové atributy závislé na klíči přímo a nikoliv tranzitivně
 - nesmí tranzitivně záviset na jiném klíči
 - nesmí záviset na jiném neklíčovém atributu
 - musí záviset pouze na primárním klíči

2.1 BCNF - Boyce-Coddova normální forma

Schéma relace R je v BNFC, jestliže pro každou netriviální závislost $X \rightarrow Y$ platí, že X obsahuje klíč schématu R . Pokud je schéma v BCNF, pak je i v 3NF.

3 Normalizace

Eliminace aktualizačních anomálií se zajišťuje převedením relačního schématu do 3NF (resp. BCNF). Lze normalizovat pomocí dekompozice. Požadavky dekompozice:

- výsledná schémata by měla mít "stejnou" sémantiku - nové i staré schéma musí mít stejnou funkční závislost
- nová relace by měla obsahovat "stejná" data, jaká by obsahovala původní relace
- o dekompozici lze uvažovat jako o projekcích původní relace, kde každá projekce definuje nové relace, a jejich spojením se musí dostat stejná data jako měla původní relace - **beztrátová dekompozice**

3.1 Beztrátová dekompozice

Musí být splněna pravidla:

- schéma $R(\underline{A}, B, C)$, kde A, B, C jsou disjunktní, a je splněna funkční závislost $B \rightarrow C$
- po rozdělení R na schémata $R1(\underline{B}, C)$ a $R2(\underline{A}, B) \Leftrightarrow$ je rozdělení (dekompozice) beztrátové

Normalizace je důležitá u "write intensive" databází, databáze kde se hodně zapisují a upravují data. Naopak u "read only" databází se vyplatí denormalizace, zavedením redundance se sníží odezva na složitější dotazy, protože denormalizované databáze mají většinou méně tabulek než normalizované. A každé spojování tabulek při dotazování prodlužuje dobu odezvy.