

Normalizace databází

- Motivace - snížit redundanci dat a předcházet tak problémům při provozování databáze, zejména nekonzistenci dat
- **Normalizace databáze** - proces úpravy návrhu databáze tak, aby vyhovovala daným podmínkám
- **Normální formy (NF)** - jednotlivé stupně normalizace dané definovanými podmínkami na strukturu tabulek

Definice pojmů

Funkční závislost

Atribut B je funkčně závislý na atributu A, jestliže s každou hodnotou atributu A je asociována právě jedna hodnota atributu B

Píšeme $A \rightarrow B$

Příklad: Místo trvalého pobytu je funkčně závislé na identifikátoru osoby, protože každá osoba má právě jedno místo trvalého pobytu.

Atribut může být funkčně závislý i na skupině více atributů
 $\{A, B\} \rightarrow C$

$\{A, B\}$ pak označujeme jako **determinující množinu**

Definice pojmů

Triviální funkční závislost

Funkční závislost $\{A\} \rightarrow B$ se nazývá triviální, pokud je atribut B podmnožinou množiny atributů $\{A\}$

Příklady: $\{A, B\} \rightarrow A, B \rightarrow B$

Definice pojmů

Úplná funkční závislost

Atribut B je úplně funkčně závislý na množině atributů $\{A\}$, pokud $\{A\} \rightarrow B$ a zároveň B není funkčně závislý na žádné části (vlastní podmnožině) množiny $\{A\}$

Příklad:

$\{id_ctenar, id_kniha\} \rightarrow \{ctenar_adresa\}$

není úplná, protože

$id_ctenar \rightarrow ctenar_adresa$

Definice pojmů

Nadklíč

Je atribut K nebo skupina atributů $\{K\} = \{K_1, \dots, K_n\}$, který/-á jednoznačně určuje řádek v dané tabulce

Nadklíč $\{K\}$ je tedy determinantem všech atributů tabulky. Pro každý atribut (sloupec) A z tabulky potom platí $\{K\} \rightarrow A$

Definice pojmů

Kandidátský klíč

Je minimální nadklíč, žádná jeho podmnožina již není nadklíčem

Neklíčový atribut

Není součástí žádného kandidátského klíče v tabulce

Primární klíč

Je jeden z kandidátských klíčů vybrán návrhářem databáze pro účely identifikace řádků v tabulce

Normální formy (NF)

- Hierarchická posloupnost skupin kritérií na strukturu tabulek v databázi
- Čím je NF vyšší, tím méně pravděpodobněji dojde k problémům s nekonzistencí při operacích s daty
- Edgar F. Codd původně definoval první tři normální formy (1NF, 2NF a 3NF)
- Teoreticky jsou popsány i další, v praxi se ovšem většinou nepoužívají

První normální forma (1NF)

- Tabulka neobsahuje opakující se záznamy, tzn. musí mít nejméně jeden kandidátský klíč
- Každý atribut musí obsahovat pouze dále nedělitelné (atomární) hodnoty

Porušení 1NF

zakaznik	datum	transakce
Petr	15. 4. 2007	+1000 +670 -76
Pavel	15. 4. 2007	-2450 +120

Porušení 1NF

zakaznik	datum	transakce
Petr	15. 4. 2007	+1000 +670 -76
Pavel	15. 4. 2007	-2450 +120

Tabulka nemá kandidátský klíč,
nelze jednoznačně identifikovat

její řádky

Obsahuje dále dělitelné hodnoty

Normalizace do 1NF

id_transakce	zakaznik	datum	transakce
1	Petr	15. 4. 2007	+1000
2	Petr	15. 4. 2007	+670
3	Petr	15. 4. 2007	-76
4	Pavel	15. 4. 2007	-2450
5	Pavel	15. 4. 2007	+120

Druhá normální forma (2NF)

- Tabulka musí být v 1NF
- Žádný z neklíčových atributů není funkčně závislý na nějaké části (vlastní podmnožině) kandidátského klíče
- Jinými slovy, všechny funkční závislosti neklíčových atributů na kandidátských klíčích jsou úplné funkční závislosti
- Pokud není ani jeden z kandidátských klíčů složený (tzn. je tvořen jediným atributem), je tabulka v 1NF automaticky též v 2NF.

Porušení 2NF

id_oddeleni	id_zamestnance	prijmeni_zamestnance
1	36	Novotná
3	36	Novotná
3	12	Halingerová
2	7	Bond

Porušení 2NF

Kandidátský klíč

{id_oddeleni, id_zamestnance} →
prijmeni_zamestnance

id_oddeleni	id_zamestnance	prijmeni_zamestnance
1	36	Novotná
3	36	Novotná
3	12	Halingerová
2	7	Bond

Funkční závislost na části kandidátského klíče
id_zamestnance → prijmeni_zamestnance

Normalizace do 2NF

id_oddeleni	id_zamestnance
1	34
3	36
3	12
2	7

id_zamestnance	prijmeni_zamestnance
34	Horáček
36	Novotná
12	Halingerová
7	Bond

Třetí normální forma (3NF)

- Tabulka musí být v 2NF
- Mezi neklíčovými atributy neexistují žádné netriviální funkční závislosti
- Porušení 3NF se projevuje tím, že nějaký neklíčový atribut je funkčně závislý na jiném neklíčovém atributu
- Hovoří se též o **nepřímé (tranzitivní)** závislosti na kandidátském klíči: $\{K\} \rightarrow A \rightarrow B$, kde $\{K\}$ je kandidátský klíč a B je na něm závislý pouze nepřímo (tranzitivně)
- Při 3NF jsou všechny atributy funkčně přímo závislé pouze na celém klíči a na ničem jiném (až na triviální případy typu $A \rightarrow A$)

Porušení 3NF

id_objednavky	zakaznik	pocet_kusu	cena_kus	celkem
1011	KB System	20	10	200
1012	Nawado, a.s.	15	10	150
1013	OEPACK	10	12	120
1014	UK v Praze	500	7	3500
1015	Jan Král	10	12	120

Porušení 3NF

Primární klíč

id_objednavky	zakaznik	pocet_kusu	cena_kus	celkem
1011	KB System	20	10	200
1012	Nawado, a.s.	15	10	150
1013	OEPACK	10	12	120
1014	UK v Praze	500	7	3500
1015	Jan Král	10	12	120

Funkční závislost na neklíčovém atributu
{pocet_kusu, cena_kus} → celkem

Normalizace do 3NF

id_objednavky	zakaznik	pocet_kusu	cena_kus
1011	KB System	20	10
1012	Nawado, a.s.	15	10
1013	OEPACK	10	12
1014	UK v Praze	500	7
1015	Jan Král	10	12

```
SELECT id_objednavky,  
       zakaznik,  
       pocet_kusu,  
       cena_kus,  
       pocet_kusu * cena_kus AS celkem  
FROM objednavky;
```

Další normální formy

- Boyce-Coddova (BCNF)
 - Čtvrtá normální forma (4NF)
 - Pátá normální forma (5NF, PJ/NF)
 - Šestá normální forma (6NF)
-
- http://en.wikipedia.org/wiki/Database_normalization
 - http://student.cvut.cz/cwut/index.php/Normální_formy