


Dědičnost v db modelu

Mise

Co je cílem přednášky?



Rozbít jednoznačnou představu o tom,
co je “správně” dělat db tabulky.

Data

- parent - atributy p1...pX
- childA - atributy a1...aY
- childB - atributy b1..bZ

Pro potřeby přednášky budu ignorovat vztahy 1:n nebo n:m

A vím, že existuje [3. normální forma](#).

Varianta “správně”

Parent

p1...pX

Child A

PK, a1...aY

Child B

PK, b1... bZ

- Kdy to tak děláme?
- Kdy to tak neděláme?
- Má to nevýhody?



Poznámky

- Výhody
 - Optimální data (teda pouze pokud je $PK < a1...aY \ \&\& \ PK < b1...bZ$)
- Nevýhody
 - Tři tabulky jsou hodně
 - Další potomek bude vyžadovat další tabulku
 - Potřeba joinování
 - Každý si myslí, že to tak vždycky dělá, ale nedělá

Varianta jedna tabulka

Parent A B

p1...pX

a1...aY

b1...bZ

- Kdybyste to takhle udělali?



Poznámky

- Hodí se, když je málo sloupců $a_1 - a_Y$, $b_1 - b_Z$ (například 0)
- Může být rychlejší než předchozí varianta
 - Žádné joiny
 - Společné indexy
- Nový potomek = nové sloupce

Varianta dvě tabulky

Parent A

p1...pX

a1...aY

Parent B

p1...pX

b1...bZ

- Má tohle někdy smysl?



Poznámky

- Hodí se když je málo p1...pX (např. nula)
- Často se využívá pro sloupce typu: created, deleted, owner... (já říkal, že to vždycky “správně” neděláte)
- Když se vždy pracuje buď s p1..pX a a1..aY nebo p1...pX a b1...bZ najednou.

Varianta generická



- Je tomu rozumět?
- Jde o tabulka zvláštních událostí (datum, co, v kolik) a narození (jméno, příjmení, místo).

Sloupec = řádek

ID	name	value
84	p1	11.10.2029
84	a1	výbuch
84	a2	17:00
23	p1	12.3.1824
23	b1	Bedřich
23	b2	Smetana
23	b3	Litomyšl
...		

Poznámky

- Vlastně může nahradit skoro celou databázi
- Přidání nových struktur pouze inserty
- Nejde používat cizí klíče na hodnoty
- Omezené využití indexů
- Špatně se dělají selecty

Rychlý dotaz	Pomalý dotaz
Pevná délka záznamu	Dynamická délka záznamu
Hledám podle indexu	Hledám podle neindexovaného výrazu
Všechna data čtu z indexu	Podle indexu hledám data
Všechna data, se kterými pracuji se vejdou do paměti najednou	Všechna data, se kterými pracuji se nevejdou do paměti najednou
Nejoinuji tabulky	Joinuji tabulky
Nedělám výpočty nad výsledky	Nad výsledky dělám nějakou operaci (group, having)
Pořadí výsledků podle zpracování (indexu)	Výsledky musím nakonec setřídít
Vrátím pouze předem známou stránku výsledků	Vrátím všechno

Co je "správně" při návrhu tabulek?

- Relační vztahy jsou jen jeden aspekt
- Další jsou velikost dat
- A nároky na vyhledávání
- A náročnost úprav v aplikaci

Možnost, jak dělat optimální strukturu (true story)

- Vždy vše cpát do jedné tabulky
- Koukat se, které operace jsou pomalé
- Vydělovat sloupce do tabulek, aby to byly rychlé

Přesvědčil jsem někoho, že
“správně” může být i jinak?

