

Optimalizácia a normalizácia databázy



Optimalizácia návrhu databázy

- Zabudli sme na nejaké stĺpce?
- Sú niektoré stĺpce zbytočné, pretože ich možno vypočítať z existujúcich polí?
- Zadávame opakovane duplicitné informácie do niektorej tabuľky?
- Máme tabuľky s mnohými poľami, obmedzeným počtom záznamov a mnohými prázdnyimi poľami v jednotlivých záznamoch?
- Obsahuje každý stĺpec údaj týkajúci sa predmetu tabuľky?
- Sú všetky vzťahy medzi tabuľkami vyjadrené buď spoločnými poľami, alebo treťou tabuľkou?

Použitie pravidiel normalizácie

- Pomocou týchto pravidiel zistíme, či sú tabuľky správne štruktúrované.
- Proces uplatnenia pravidiel na návrh databázy sa nazýva normalizácia databázy alebo jednoducho normalizácia.
- Normalizácia je najužitočnejšia vtedy, keď sme už vyjadrili všetky informačné položky a dospeli k predbežnému návrhu.
- Pravidlá uplatňujeme postupne a v každom kroku overujeme, či je návrh v súlade s tzv. normálnou formou.
- Všeobecne je akceptovaných päť normálnych foriem.

Prvá normálna forma

- Prvá normálna forma stanovuje, že v každom prieniku riadka a stĺpca v tabuľke existuje jedna hodnota a nikdy nie viac
- Ak si všetky jednotlivé prieniky riadkov a stĺpcov predstavíme ako bunky, každá bunka smie obsahovať len jednu hodnotu.



Každé pole má obsahovať jednu hodnotu daného významu a nemáme vytvárať opakujúce sa skupiny stĺpcov

Nevhodné	Vhodnejší princíp
Telefony = „0901..., 0902...“	Samostatné záznamy telefónov, ak ich potrebujeme evidovať viac.
Produkt1, Produkt2, Produkt3	Položky objednávky ako samostatné riadky.
„Notebook; Myš; Monitor“ v jednom poli	Každý výskyt vzťahu samostatne.

Druhá normálna forma

- Druhá normálna forma vyžaduje, aby bol každý neklúčový stípec plne závislý od hlavného kľúča a nie len od jeho časti.
- Toto pravidlo platí pre hlavný kľúč pozostávajúci minimálne z dvoch stípcov.



Nekľúčové pole má závisieť od celého kľúča, nie iba od jeho časti.

IDObjednavky	IDProduktu	NazovProduktu	Mnozstvo
1001	5	Notebook	1
1001	9	Myš	2

Ak je kľúčom dvojica IDObjednavky + IDProduktu, názov produktu závisí iba od IDProduktu, nie od celej dvojice. Preto patrí do tabuľky Produkty. Množstvo však opisuje konkrétny produkt v konkrétnej objednávke a patrí do položky objednávky.

Tretia normálna forma

- Tretia normálna forma vyžaduje nielen to, aby bol každý neklúčový stĺpec závislý od celého hlavného kľúča, ale aj to, aby boli neklúčové stĺpce od seba nezávislé.
- Inak povedané, každý neklúčový stĺpec musí byť závislý, a to výhradne od hlavného kľúča.

3NF zjednodušene požaduje, aby neklúčové údaje nezáviseli od iných neklúčových údajov, ale od kľúča.

IDZakaznika	PSC	Mesto
1	94901	Nitra
2	94901	Nitra
3	91701	Trnava

Ak pravidlá systému hovoria, že konkrétne PSČ jednoznačne určuje mesto, ukladanie mesta pri každom zákazníkovi môže vytvárať závislosť PSC → Mesto. Či ho oddelíme, závisí aj od reálnych pravidiel dát a potrieb systému; príklady normalizácie treba vždy posudzovať podľa domény.

Normalizácia nemusí byť cieľ

- Dobrá databáza má podporovať reálne potreby aplikácie.
- Normalizovaný návrh býva bezpečným východiskom, ale v špecifických systémoch sa môže vedome použiť denormalizácia – kontrolované uloženie odvodených alebo opakovaných údajov pre výkon, históriu alebo jednoduchosť čítania.

Indexy a výkon

- Index je pomocná dátová štruktúra, ktorá databázovému systému uľahčuje vyhľadanie záznamov podľa hodnoty poľa.
- Hlavný kľúč je v Access indexovaný automaticky.
- Indexovať možno aj ďalšie polia, najmä tie, podľa ktorých sa často vyhľadáva, triedi alebo prepája.

Index môže pomôcť	Index má aj cenu
rýchlejšie vyhľadávanie	zaberá ďalší priestor
rýchlejšie triedenie a spájanie	pri INSERT/UPDATE/DELETE treba index aktualizovať
vynútenie jedinečnosti pri No Duplicates	príliš veľa indexov môže spomaľovať zápisy
Pravidlo: Neindexujeme „všetko“. Index vyberáme podľa spôsobu používania dát. Optimalizácia je kompromis medzi rýchlosťou čítania, zápisu, priestorom a jednoduchosťou správy.	

Optimalizácia

- Častou chybou je optimalizovať databázu skôr, než máme správny návrh a reálne meranie problému. Nesprávne uložené dáta sa nedajú zachrániť množstvom indexov.
 1. Ujasnenie účelu a pravidiel databázy.
 2. Návrh entít, kľúčov a vzťahov.
 3. Odstránenie nežiaducej redundancie a anomálie.
 4. Overenie funkčnosti na reálnych dotazoch a objeme dát.
 5. Identifikácia pomalých operácií.
 6. Pridanie vhodných indexov alebo úprava dotazov/návrhu.
 7. Nové overenie po nastavení predošlých bodov.



Ďakujem za pozornosť!