

Zawartość:

- pierwsza postać normalna
- druga postać normalna
- trzecia postać normalna

Projektując relacyjną bazę danych deweloper powinien stosować się do minimum trzech podstawowych zasad zwanych

- Pierwszą postacią normalną
- Drugą postacią normalną
- Trzecią postacią normalną

Pierwsza postać normalna

Każda wartość w bazie danych musi być atomowa.

Oznacza to, że każda komórka w tabeli musi zawierać dokładnie jedną wartość z logicznego punktu widzenia. Jeżeli zachodzi konieczność przechowywania w jednej komórce większej ilości wartości oznacza to, że struktura bazy może być źle zaprojektowana.

id	name	price	ingredients
1	Margherita	22	["ser", "sos pomidorowy"]
2	Capriciosa	25	["ser", "sos pomidorowy", "szynka"]
3	Hawajska	26	["ser", "sos pomidorowy", "szynka", "ananas"]

Jakby się ktoś uparł to można to podciągnąć pod jedną wartość w postaci JSONa, ale jeżeli weźmiemy pod uwagę, że w bazie będziemy musieli powtarzać np. „ser” wiele razy to pojawia się problem zbędnej duplikacji danych.

id	name	price
1	Margherita	22
2	Capriciosa	25
3	Hawajska	26

position_id	ingredient_id
1	1
1	2
2	1
...	...

id	name
1	ser
2	sos pomidorowy
3	szynka
4	ananas

Rozwiązanie to wykorzystanie relacji wiele-do-wielu, która jak pamiętamy wymaga trzeciej tabeli wiążącej.

Inny przykład to tabela z uczniami w szkole, która posiada kolumnę „oceny”. Przy takim projekcie nie ma opcji by nie wrzucać kolejnych ocen do tej jednej komórki. A to jest oczywiście złamanie 1PN.

Druga postać normalna

Każda wartość danego rekordu musi zależeć od klucza głównego. Oczywiście musi być też spełniona 1PN.

Na pierwszy rzut oka ciężko to zrozumieć. Zobaczmy przykład:

id	firstname	lastname	address	city	product	price	quantity
1	Jan	Kowalski	Chabrowa 4	Wrocław	Długopis	4	100
2	John	Smith	Towarowa 7	Warszawa	Zeszyt	8	4

Przechowujemy tutaj jakieś dane osobowe (może dane klientów) oraz informacje o jakimś produkcie (np. kupionym przez wcześniej wspomnianego klienta). Jedna tabela przechowuje informacje o klientach i ich adresach oraz produkty wraz z cenami.

Powinniśmy to robić na dwie tabele i połączyć je relacją wiele do wiele.

Takie id jak wyżej to tak zwany klucz sztuczny – wartość sztucznie dodana by jednoznacznie identyfikować rekord.

Wyobraźmy sobie, że autor tabeli zastosował jako klucz główny klucz naturalny w postaci numeru pesel. W takiej sytuacji dosłowne rozumienie treści 2PN miało by nawet sens.

Trzecia postać normalna

Każda kolumna nie będąca kluczem nie może zależeć od innej kolumny nie będącej kluczem. Oczywiście poprzednie dwie zasady muszą być spełnione.

Najlepszym przykładem jest przechowywanie ceny w tabeli z produktami. Wyobraźmy sobie tabelę „Produkty” przechowującą nazwę produktu, kategorię oraz cenę w formacie brutto, netto (każda w osobnej kolumnie).

Taka sytuacja jest złamaniem 3PN, ponieważ cenę brutto da się wyliczyć z ceny netto i odwrotnie znając globalny procent. Nie ma sensu więc przechowywać obu wartości w tabeli.