

Normalizacja baz danych

Doprowadzenie relacji do postaci normalnej nazywa się normalizacją bazy danych. Postać normalna – postać relacji w bazie danych, w której nie występuje redundancja (nadmiarowość), czyli powtarzanie się tych samych informacji.

Pierwsza postać normalna (1NF)

1 Postać Normalna (1NF) to pierwszy poziom normalizacji danych, który zakłada, że:

- każda komórka musi przechowywać pojedynczą, atomową wartość, a nie zestaw wartości lub struktur danych.

product_name	product_price
Laptop, Smartphone	2500.00, 800.00
Tablet, Keyboard	1200.00, 150.00
TV, Monitor	1800, 500.00

W tym przypadku kolumna **product_name** zawiera zestaw produktów, a kolumna **product_price** zawiera zestaw cen. Nie spełnia to wymagań **1NF**, ponieważ komórki nie zawierają pojedynczych wartości.

Po normalizacji 1NF dane zostają podzielone na oddzielne wiersze - każdy wiersz zawiera pojedynczy produkt i jego cenę. Wtedy tabela może wyglądać następująco:

product_name	product_price
Keyboard	150.00
Laptop	2500.00
Monitor	500.00
Smartphone	800.00
Tablet	1200.00
TV	1800.00

Teraz każdy wiersz zawiera pojedynczy produkt i jego cenę, co spełnia warunki **1NF**.

Druga postać normalna (2NF)

2 Postać Normalna (2NF) to drugi poziom normalizacji danych, który:

- spełnia 1NF oraz zakłada, że każda kolumna w tabeli musi zależeć od klucza głównego.

Spójrzmy na poniższą tabelę **orders**:

order_id	customer_name	address	city	product	price	quantity
1	John Smith	123 Main Street	New York	Laptop	1200	2
2	Alice Johnson	456 Elm Avenue	Los Angeles	Smartphone	800	3
3	Maria Garcia	789 Oak Road	Chicago	Tablet	500	1
4	Robert Brown	101 Pine Street	Houston	Monitor	300	4

Powyższa tabela spełnia 1NF, ale nie spełnia 2NF. Mamy tutaj zbyt wiele informacji.

Dane klienta oraz produktu zależą od identyfikatora zamówienia. Dane są zbyt powiązane. Powinniśmy oddzielić dane klienta i produktu od zamówienia.

Tabela z danymi klienta:

customer_id	customer_name	address	city
1	John Smith	123 Main Street	New York
2	Alice Johnson	456 Elm Avenue	Los Angeles
3	Maria Garcia	789 Oak Road	Chicago
4	Robert Brown	101 Pine Street	Houston

Tabela produktów:

product_id	product_name	price
1	Laptop	1200.00
2	Smartphone	800.00
3	Tablet	500.00
4	Monitor	300.00

Tabela zamówień:

order_id	customer_id	product_id	quantity
1	1	1	2
2	2	2	3
3	3	3	1
4	4	4	4

Od teraz wszystkie dane niepowiązane z zamówieniami zostały przeniesione do innych tabel. Tabela **zamówienia** zawiera jedynie klucze obce (dowiązania) do tabel **customers** i **products**.

Trzecia postać normalna (3NF)

3 Postać Normalna (3NF) to trzeci poziom normalizacji danych, który:

- spełnia 2NF oraz zakłada, że niekluczowa kolumna nie może zależeć od innej niekluczowej kolumny.

Przeanalizujmy tabelę produktów:

id	name	size	price
1	Margherita	MEDIUM	15.00
2	Pepperoni	LARGE	20.00
3	Hawaiian	MEDIUM	15.00
4	Vegetarian	SMALL	10.00
5	Supreme	LARGE	20.00

W przypadku trzeciej postaci normalnej (3NF) każda kolumna niekluczowa musi zależeć tylko od klucza głównego, a nie od innych niekluczowych kolumn. W tabeli **products** cena i rozmiar pizzy są kolumnami niekluczowymi, a cena jest zależna od rozmiaru, co jest niezgodne z trzecią postacią normalną (3NF). Aby spełnić tę zasadę, można utworzyć osobną tabelę **product_sizes**, gdzie cena byłaby związana tylko z rozmiarem produktu.

Postaramy się w taki sposób przekształcić dane, aby spełnić 3NF.

W pierwszej kolejności uzależnimy cenę od rozmiaru, ale w oddzielnej tabeli **product_sizes**:

size	price
LARGE	20.00
MEDIUM	15.00
SMALL	10.00

W następnym kroku przebudujemy tabelę produktów:

id	name	size
1	Margherita	MEDIUM
2	Pepperoni	LARGE
3	Hawaiian	MEDIUM
4	Vegetarian	SMALL
5	Supreme	LARGE

Teraz tabela **products** zawiera klucz obcy **size**, który odnosi się do tabeli **product_sizes**, co umożliwia zarządzanie cenami w zależności od rozmiaru pizzy, a jednocześnie zachowuje strukturę produktów w tabeli **products**.