

Jeśli klucze w tabeli opisującej związek składają się z wielu atrybutów lub są długie, należy zastąpić je kluczami sztucznymi.

1.4.7. Normalizacja tabel

Normalizację stosuje się, aby sprawdzić, czy zaprojektowane tabele mają prawidłową strukturę. Proces normalizacji rozpoczynamy, gdy zostanie utworzony wstępny projekt tabel. Pozwala ona określić, czy informacje przewidziane w projekcie bazy zostały przydzielone do właściwych tabel. Natomiast nie da odpowiedzi na pytanie, czy projekt bazy danych jest prawidłowy.

Korzyści płynące z normalizacji tabel są następujące:

- zlikwidowanie problemu powtarzania danych,
- optymalizacja objętości bazy danych,
- optymalizacja efektywności obsługi bazy danych,
- minimalizacja zagrożenia błędami przy wprowadzaniu danych.

Normalizacja bazy danych wymaga rozbicia dużych tabel na mniejsze. Zmniejsza to wydajność bazy, dlatego w niektórych przypadkach nie normalizuje się tabel — szczególnie w systemach niekorzystających z modelu relacyjnego.

Stosowane są cztery reguły normalizacji, ale w większości projektów baz danych wystarczy sprawdzić trzy pierwsze. Zostaną one omówione poniżej.

Dla każdej z nich stosowane są określenia: pierwsza postać normalna (I PN), druga postać normalna (II PN) i trzecia postać normalna (III PN).

Pierwsza postać normalna

DEFINICJA

Tabela jest w pierwszej postaci normalnej (**I PN**), gdy każdy wiersz w tabeli przechowuje informacje o pojedynczym obiekcie, a każde pole tabeli zawiera informację elementarną (atomową).

Oznacza to, że w komórce tabeli nie może wystąpić lista wartości, na przykład w polu *Narodowość autora* nie można umieścić dwóch narodowości — polskiej i angielskiej.

Załóżmy, że w projektowanej bazie danych dla księgarni została zaprojektowana tabela *Realizacja zamówień* z polami: *Nazwisko klienta*, *Imię*, *Adres*, *Telefon*, *PESEL*, *Tytuł książki*, *Liczba egzemplarzy*, *Cena*. W polu *Tytuł książki* będą umieszczane tytuły książek zakupionych przez klienta. Gdy klient kupi dwie książki, w polu *Tytuł książki* należałoby wpisać dwa tytuły. Powstałaby lista wartości (rysunek 1.30). Tak zaprojektowana tabela nie jest w I PN. Nie będzie możliwe prawidłowe przetwarzanie danych zapisanych w tabeli. Rozwiązaniem jest zapisanie informacji o zakupionych książkach

w dwóch wierszach. W pierwszym wierszu w polu *Tytuł książki* należy wpisać tytuł pierwszej książki, w drugim — tytuł drugiej książki, natomiast nazwisko klienta zostanie powtórzone w liczbie wierszy równej liczbie zakupionych książek (rysunek 1.31). Teraz tabela jest w I PN.

Realizacja zamówień							
Nazwisko klienta	Imię	Adres	Telefon	PESEL	Tytuł książki	Liczba	Cena
Nowak	Marek	Toruń	(56) 6589234	79120307431	Dziady	2	20,00 zł
Kowalski	Adam	Warszawa	(22) 3451234	80122401871	Balladyna, Tango	1	15,00 zł
Górecki	Grzegorz	Poznań	(45) 2367897	82061203983	Pan Tadeusz	1	21,00 zł
Zan	Marcin	Gdańsk	(33) 8373635	82020201875			

Rysunek 1.30. Tabela nie jest w I PN, ponieważ w polu *Tytuł książki* pojawiła się lista wartości

Realizacja zamówień							
Nazwisko klienta	Imię	Adres	Telefon	PESEL	Tytuł książki	Liczba	Cena
Nowak	Marek	Toruń	(56) 6589234	79120307431	Dziady	2	20,00 zł
Kowalski	Adam	Warszawa	(22) 3451234	80122401871	Balladyna	1	15,00 zł
Kowalski	Adam	Warszawa	(22) 3451234	80122401871	Tango	2	18,00 zł
Górecki	Grzegorz	Poznań	(45) 2367897	82061203983	Pan Tadeusz	1	21,00 zł

Rysunek 1.31. Tabela jest w I PN, ponieważ w polu *Tytuł książki* występują pojedyncze wartości

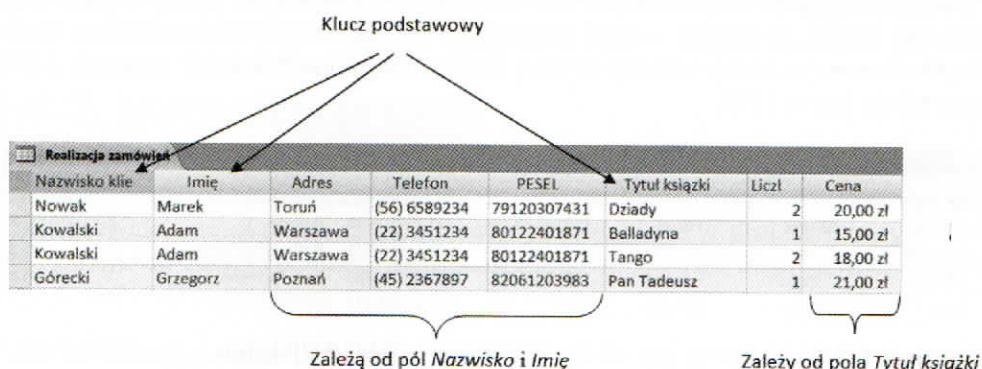
Druga postać normalna

DEFINICJA

Tabela jest w drugiej postaci normalnej (**II PN**), jeżeli jest w pierwszej postaci normalnej (**I PN**) oraz każde z pól niewchodzących w skład klucza podstawowego zależy od całego klucza, a nie od jego części.

Ta reguła i następna służą do sprawdzenia, czy w tabeli i bazie danych nie doszło do redundancji, czyli niepotrzebnego powtarzania danych. Ponieważ **II PN** odnosi się do klucza podstawowego, należy określić ten klucz dla tabeli. Jeżeli klucz podstawowy składa się z jednego pola, tabela jest w **II PN**, ponieważ wszystkie pola, poza polem klucza podstawowego, muszą odnosić się do pola klucza podstawowego.

W zaprojektowanej tabeli *Realizacja zamówień* kluczem podstawowym będzie kombinacja pól *Nazwisko klienta*, *Imię* oraz *Tytuł książki*. Pola niewchodzące w skład klucza podstawowego (*Adres*, *Telefon*, *PESEL*) zależą od pól *Nazwisko* i *Imię*, natomiast nie zależą od pola *Tytuł książki*. Pole *Cena* zależy jedynie od pola *Tytuł książki* (rysunek 1.32). Tylko pole *Liczba egzemplarzy* zależy zarówno od pól *Nazwisko* i *Imię*, jak i od pola *Tytuł książki*. Tabela nie jest w **II PN**.



Rysunek 1.32. Tabela nie jest w II PN, ponieważ tylko pole Liczba egzemplarzy zależy od całego klucza

Normalizacja polega na podzieleniu tabeli na takie tabele, które spełnią warunek II PN. Tabelę *Realizacja zamówień* należy podzielić na trzy tabele: *Klient*, z polami *Nazwisko*, *Imię*, *Adres*, *Telefon*, *PESEL*; *Zamówienia*, z polami *Nazwisko klienta*, *Tytuł książki*, *Liczba egzemplarzy*; oraz *Książki*, z polami *Tytuł książki* i *Cena* (rysunek 1.33). Kluczem podstawowym w tabeli *Klient* są pola *Nazwisko* i *Imię*, w tabeli *Zamówienia* pola *Nazwisko klienta* i *Tytuł książki*, a w tabeli *Książki* pole *Tytuł książki*. Zostało zlikwidowane powtarzanie danych w tabeli *Realizacja zamówień* i wszystkie tabele są w II PN.

Nazwisko klienta	Imię	Adres	Telefon	PESEL
Nowak	Marek	Toruń	(56) 6589234	79120307431
Kowalski	Adam	Warszawa	(22) 3451234	80122401871
Górecki	Grzegorz	Poznań	(45) 2367897	82061203983
Zan	Marcin	Gdańsk	(33) 8373635	82020201875
Bagińska	Anna	Warszawa	(71) 7316252	91110402837
Pol	Aleksander	Szczecin	(23) 6517830	70073003228

Nazwisko klienta	Tytuł książki	Liczba egz.
Nowak	Dziady	2
Kowalski	Balladyna	1
Kowalski	Tango	2
Górecki	Pan Tadeusz	1

Tytuł książki	Cena
Dziady	20,00 zł
Balladyna	15,00 zł
Tango	18,00 zł
Pan Tadeusz	21,00 zł

Rysunek 1.33. Podział tabeli Realizacja zamówień na tabele spełniające warunek II PN

Aby wyświetlić zestawienie dotyczące klienta i kupionych przez niego książek, należy zdefiniować połączenie między tabelami.

Trzecia postać normalna

DEFINICJA

Tabela jest w trzeciej postaci normalnej (III PN), jeżeli jest w pierwszej i w drugiej postaci normalnej oraz każde z pól niewchodzących w skład klucza podstawowego niesie informację bezpośrednio o kluczu i nie odnosi się do żadnego innego pola.

Założmy, że w projektowanej bazie danych dla księgarni została zaprojektowana tabela *Faktura* z polami: *Nazwisko klienta*, *Imię*, *Adres*, *PESEL*, *Numer faktury*, *Sposób płatności* i *Data wystawienia faktury*. Założmy również, że w pola tabeli będą wpisywane tylko

wartości elementarne, czyli tabela jest w I PN. Klucz podstawowy to pole *Numer faktury*. Wszystkie pola niewchodzące w skład klucza zależą od całego klucza, czyli tabela jest w II PN. Sprawdźmy, czy tabela jest w III PN. Pola *Sposób płatności* i *Data wystawienia faktury* odnoszą się do faktury, czyli zawierają informacje o kluczu. Natomiast pola *Adres* i *PESEL* zawierają informacje na temat klienta, a nie faktury (rysunek 1.34), czyli nie niosą informacji bezpośrednio o kluczu. Tabela nie jest w III PN.

Klucz podstawowy

Nazwisko klie	Imię	Adres	PESEL	Numer faktury	Sposób płatn	Data wystaw
Nowak	Marek	Toruń	79120307431		1 gotówka	2013-04-06
Kowalski	Adam	Warszawa	80122401871		2 przelew	2012-07-29
Kowalski	Adam	Warszawa	80122401871		4 gotówka	2012-01-17
Bagińska	Anna	Warszawa	91110402837		5 gotówka	2012-08-29
Pol	Aleksander	Szczecin	70073003228		6 gotówka	2012-03-02
Pol	Aleksander	Szczecin	70073003228		3 przelew	2012-12-10

Odnoszą się do pól *Nazwisko* i *Imię*
Odnoszą się do pola *Numer faktury*

Rysunek 1.34. Tabela nie jest w III PN, ponieważ pola *Adres* i *PESEL* nie niosą informacji o kluczu

Normalizacja, podobnie jak w przypadku II PN, polega na podzieleniu tabeli na takie tabele, które spełnią warunek III PN.

Tabelę *Faktura* należy podzielić na dwie tabele: *Faktura* (z polami *Numer faktury*, *Sposób płatności* i *Data wystawienia faktury*) oraz *Klient* (z polami *Nazwisko klienta*, *Imię*, *Adres*, *PESEL*) (rysunek 1.35). Zostało zlikwidowane powtarzanie danych o kliencie w tabeli *Faktura*. Dane o kliencie będą zapisane tylko raz, w tabeli *Klient*.

Nazwisko klie	Imię	Adres	PESEL
Nowak	Marek	Toruń	79120307431
Kowalski	Adam	Warszawa	80122401871
Górecki	Grzegorz	Poznań	82061203983
Zan	Marcin	Gdańsk	82020201875
Bagińska	Anna	Warszawa	91110402837
Pol	Aleksander	Szczecin	70073003228

Numer faktury	Sposób płatn	Data wystawie	Nazwisko klienta
	1 gotówka	2013-04-06	Nowak
	2 przelew	2012-07-29	Kowalski
	3 przelew	2012-12-10	Pol
	4 gotówka	2012-01-17	Kowalski
	5 gotówka	2012-08-29	Bagińska
	6 gotówka	2012-03-02	Pol

Rysunek 1.35. Podział tabeli *Faktura* na tabele spełniające warunek III PN

Przykład 1.2

Przestrzegając reguł tworzenia tabel, po sprawdzeniu za pomocą normalizacji, czy tabele mają prawidłową strukturę, baza danych dla księgarni mogłaby składać się z następujących tabel: