

Normalizacja

Normalizacja jest procesem eliminacji powtarzających się danych w modelu relacyjnym.

Prosta zasada normalizacji: trzymamy dane w jednym miejscu (tabeli) i w razie potrzeby linkujemy (klucz obcy).

Dodatkowym celem normalizacji jest zapobieganie anomaliom:

- redundancji
- anomalii wprowadzania
- anomalii modyfikacji
- anomalii usuwania

Opis tych anomalii znajdziecie w głównym pliku wykładu.

Normalizacja pozwala także wyeliminowanie relacji wieloznacznych, które przechowują w jednym wierszu szereg danych o różnej tematyce i zakresie.

Normalizacja

Istnieją trzy postaci (reguły) normalizacji zwane postaciami normalnymi, które oznaczamy jako (1NF, 2NF, 3NF).

Uważa się, że trzecia postać normalna jest postacią wystarczającą do poprawnego działania większości aplikacji. Wobec tego, gdy będziecie proszeni o znormalizowanie schematu - w domyśle chodzi o postać trzecią.

Sprowadzanie schematu bazy danych do poszczególnych postaci normalnych prawie zawsze będzie powodowało zwiększenie ilości tabel.

Normalizacja może powodować także wydłużenie dostępu do danych ze względu na stosowanie złączeń pomiędzy tabelami.

Normalizacja

Pierwsza postać normalna 1NF:

- tylko atomowe wartości w komórkach (brak struktur danych takich jak tablice, listy itd.) oraz podział atrybutów na części logicznie niepodzielne jak 'imię i nazwisko', które dzielimy na dwa osobne atrybuty 'imię', 'nazwisko',
- eliminacja powtarzających się grup w poszczególnych tabelach
- utworzenie osobnej tabeli dla każdego zestawu powiązanych danych
- Identyfikowanie każdego zestawu powiązanych danych za pomocą klucza podstawowego

Przykładem może być tabela dostępnych kursów (przedmiotów) na uczelni. W momencie, gdy dany przedmiot prowadzi więcej niż jeden wykładowca, tabela, która nie jest znormalizowana zawierałaby pola *prowadzący1*, *prowadzący2*. W przypadku, gdy prowadzących byłoby więcej, musielibyśmy dodać kolejne pola.

Rozwiązaniem jest utworzenie osobnej tabeli z prowadzącymi i powiązanie je za pomocą klucza obcego.

Normalizacja

Kolejność wierszy nie ma znaczenia!

1NF

Brak atomowości

Zdublowane nazwy kolumn

Imię i nazwisko	Certyfikaty	Certyfikaty
Adam Nowak	MBA, CAE, IELTS	MBA
Justyna Kowalska	CAE	CAE
Mariusz Ogórek	332	
Paulina Noga	MBA	MBA, inż.

Różny typ danych

Zdublowany rekord

Normalizacja

Tabelę, która wymaga normalizacji dzielimy wertykalnie na tabele potomne. Podział ten wymaga jednak, aby tabele wynikowe, po złączeniu, pozwoliły na odtworzenie tabeli pierwotnej.

Przykład:

NrPacjenta	NrZabiegu	DataWizyty	Gabinet	Lekarz
P22034	Z244	10-10-2019	113	Jaśkiewicz
P22434	Z553	11-10-2019	124	Nowak
P22433	Z244	11-10-2019	234	Kowalska
P45452	Z200	11-10-2019	113	Nowak
P92833	Z201	12-10-2019	134	Jaśkiewicz

Normalizacja

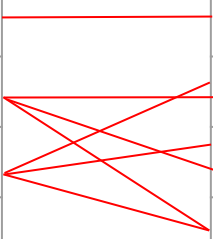
Dekompozycja stratna:

NrPacjenta	NrZabiegu	DataWizyty	DataWizyty	Gabinet	Lekarz
P22034	Z244	10-10-2019	10-10-2019	113	Jaśkiewicz
P22434	Z553	11-10-2019	11-10-2019	124	Nowak
P22433	Z244	11-10-2019	11-10-2019	234	Kowalska
P45452	Z200	11-10-2019	11-10-2019	113	Nowak
P92833	Z201	12-10-2019	12-10-2019	134	Jaśkiewicz

Normalizacja

Dekompozycja stratna:

NrPacjenta	NrZabiegu	DataWizyty		DataWizyty	Gabinet	Lekarz
P22034	Z244	10-10-2019		10-10-2019	113	Jaśkiewicz
P22434	Z553	11-10-2019		11-10-2019	124	Nowak
P22433	Z244	11-10-2019		11-10-2019	234	Kowalska
P45452	Z200	11-10-2019		11-10-2019	113	Nowak
P92833	Z201	12-10-2019		12-10-2019	134	Jaśkiewicz



Normalizacja

Błędnie złączone wiersze:

NrPacjenta	NrZabiegu	DataWizyty	DataWizyty	Gabinet	Lekarz
P22433	Z244	11-10-2019	11-10-2019	124	Nowak
P45452	Z200	11-10-2019	11-10-2019	234	Kowalska

Normalizacja

Druga postać normalna 2NF:

- Tabela jest 1NF
- Wszystkie atrybuty niekluczowe zależą funkcyjnie od pełnego klucza
- Dopuszczalne są zależności funkcyjne pomiędzy atrybutami niekluczowymi

Atrybuty niekluczowe, które zależą funkcyjnie od części klucza złożonego (na przykład jednej z dwóch kolumn, które stanowią klucz złożony) wykluczają istnienie 2NF.

Wszystkie tabele, które są w 1NF i posiadają prosty klucz główny (złożony z jednej kolumny) są automatycznie w 2NF.

Normalizacja

Rozważmy tabelę:

<u>IDNauczyciela</u>	<u>Przedmiot</u>	Wiek
111	Matematyka	64
111	Fizyka	64
224	Historia	46
323	Geografia	34
323	WOS	34

Normalizacja

Rozważmy tabelę:

<u>IDNauczyciela</u>	<u>Przedmiot</u>	Wiek
111	Matematyka	64
111	Fizyka	64
224	Historia	46
323	Geografia	34
323	WOS	34

- tabela jest 1NF - atrybuty mają wartości atomowe
- atrybut niekluczowy *wiek* jest zależny funkcyjnie tylko od części klucza głównego *IDNauczyciela*
- 2NF nie jest spełnione!

Normalizacja

Sprowadzenie tabel do 2NF wymaga dekompozycji na dwie tabele:

<u>IDNauczyciela</u>	Przedmiot
111	Matematyka
111	Fizyka
224	Historia
323	Geografia
323	WOS

<u>IDNauczyciela</u>	Wiek
111	64
111	64
224	46
323	34
323	34

Normalizacja

Inny przykład:

Notes(imię , nazwisko , imieniny , urodziny , znak zodiaku)

<u>imię</u>	<u>nazwisko</u>	Imieniny*	urodziny	znak zodiaku
Adam	Nowak	24.12	8.08	lew
Agnieszka	Kowalska	21.01	2.06	bliźnięta

urodziny -> znak zodiaku - {urodziny} nie wchodzi w skład klucza, więc zależność nie wyklucza postaci 2NF.

imię -> imieniny - {imię} jest podkluczem, więc zależność wyklucza postać 2NF.

* zakładamy imieniny dla danego imienia raz w roku (np. Adam), inaczej byłby to przykład zależności funkcyjnej wielowartościowej

Normalizacja

Trzecia postać normalna:

- Tabela jest 2NF
- nie są możliwe zależności funkcyjne między atrybutami niekluczowymi
- nie występują zależności przechodnie

Dla przykładu: **Notes(imię , nazwisko , imieniny , urodziny , znak zodiaku)**

urodziny -> znak zodiaku - {urodziny} nie wchodzi w skład klucza, więc warunek 2NF jest spełniony, ale 3NF nie.

Normalizacja

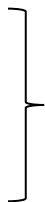
Zależności przechodnie między atrybutami relacji zachodzą, jeżeli spełniona jest zależność:

Jeżeli $A \rightarrow B$ i $B \rightarrow C$ to $A \rightarrow C$

Przykład:

numer produktu \rightarrow numer dystrybutora

numer dystrybutora \rightarrow numer magazynu

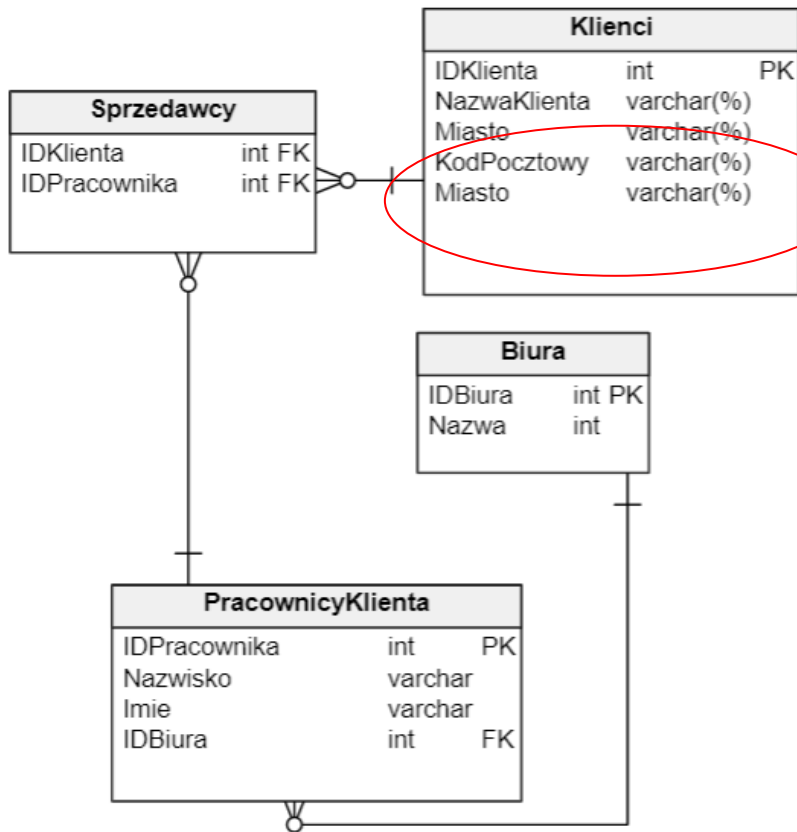


numer produktu \rightarrow numer magazynu

Normalizacja

Klucz główny	Kolumna A	Kolumna B	Zależność przechodnia?
ID_pracownika	Imię	Nazwisko	Nie. Nazwisko jest dziedziczone.
ID_klubowicza	Wskaznik_BMI	CzyNadwaga	Tak. Zakładając $BMI > 25 =$ nadwaga. Zatem: $ID \rightarrow BMI$; $BMI \rightarrow$ Nadwaga; $ID \rightarrow$ Nadwaga
VIN	Model	Producent	Tak. $VIN \rightarrow$ Producent

Normalizacja

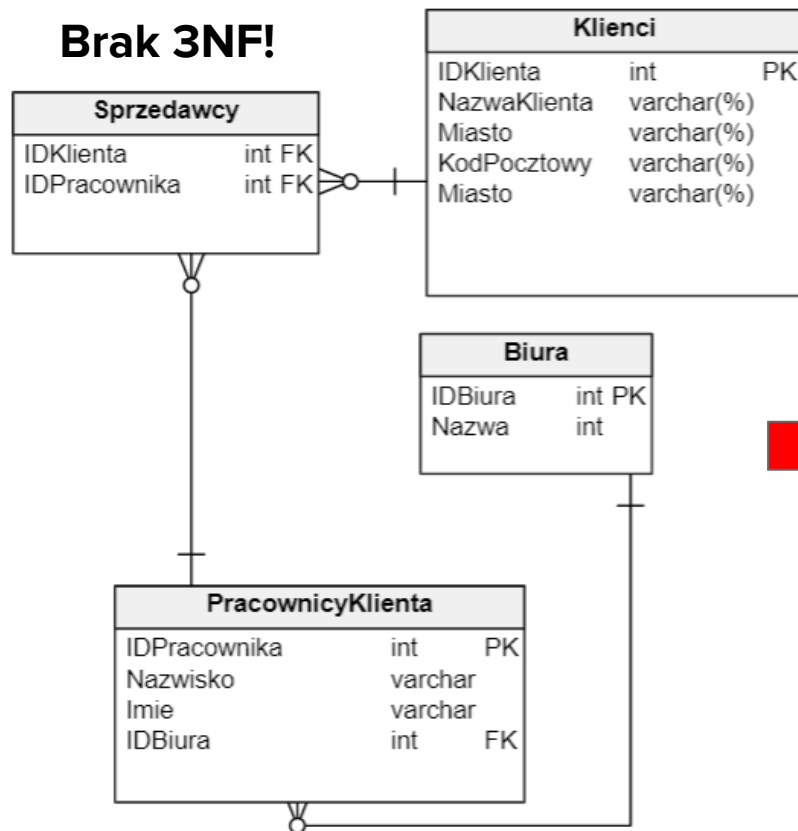


Miasto jest zależne od kodu pocztowego

Brak 3NF

Normalizacja

Brak 3NF!



3NF

