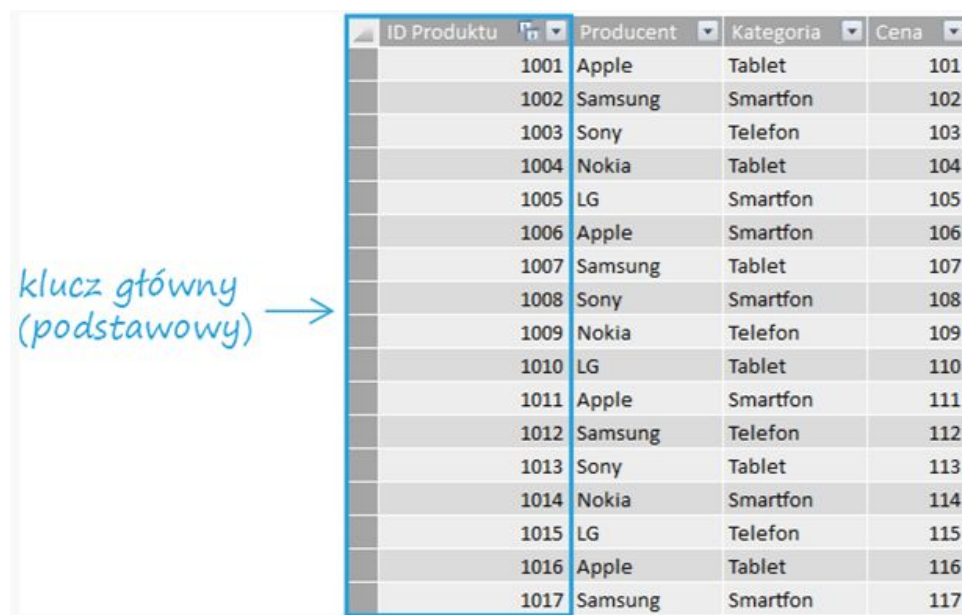


# Bazy danych

## 2. Relacyjne bazy danych (model danych, klucze, postaci normalne, SQL

Relacyjna baza danych to opisany i zorganizowany zbiór tabel połączonych relacjami – związkami między sobą. Ten sposób przechowywania informacji pozwala na uniknięcie powtarzania się danych oraz przeprowadzanie analiz na podstawie wielu tabel. Każda tabela składa się z wierszy. Poszczególne rekordy składają się z pól, przechowujących jedną daną. Aby istniała możliwość utworzenia z tabel relacyjnego modelu danych, przynajmniej w jednej z nich musi występować klucz główny (podstawowy). Jest to po prostu kolumna służąca do identyfikacji poszczególnych rekordów tabeli. Wartości w kluczu podstawowym muszą być unikalne, aby istniała możliwość przypisania jednego wiersza tabeli do jednej wartości klucza.



ID Produktu	Producent	Kategoria	Cena
1001	Apple	Tablet	101
1002	Samsung	Smartfon	102
1003	Sony	Telefon	103
1004	Nokia	Tablet	104
1005	LG	Smartfon	105
1006	Apple	Smartfon	106
1007	Samsung	Tablet	107
1008	Sony	Smartfon	108
1009	Nokia	Telefon	109
1010	LG	Tablet	110
1011	Apple	Smartfon	111
1012	Samsung	Telefon	112
1013	Sony	Tablet	113
1014	Nokia	Smartfon	114
1015	LG	Telefon	115
1016	Apple	Tablet	116
1017	Samsung	Smartfon	117

W podstawowym użyciu jest to tzw. klucz jednopolowy – identyfikacja rekordu odbywa się przy pomocy jednego pola w wierszu.

Istnieją także klucze złożone (wielopolowe) – w ich przypadku identyfikacja odbywa się przy pomocy więcej niż jednej kolumny. Unikalne jest zestawienie komórek tworzących klucz w wierszu.

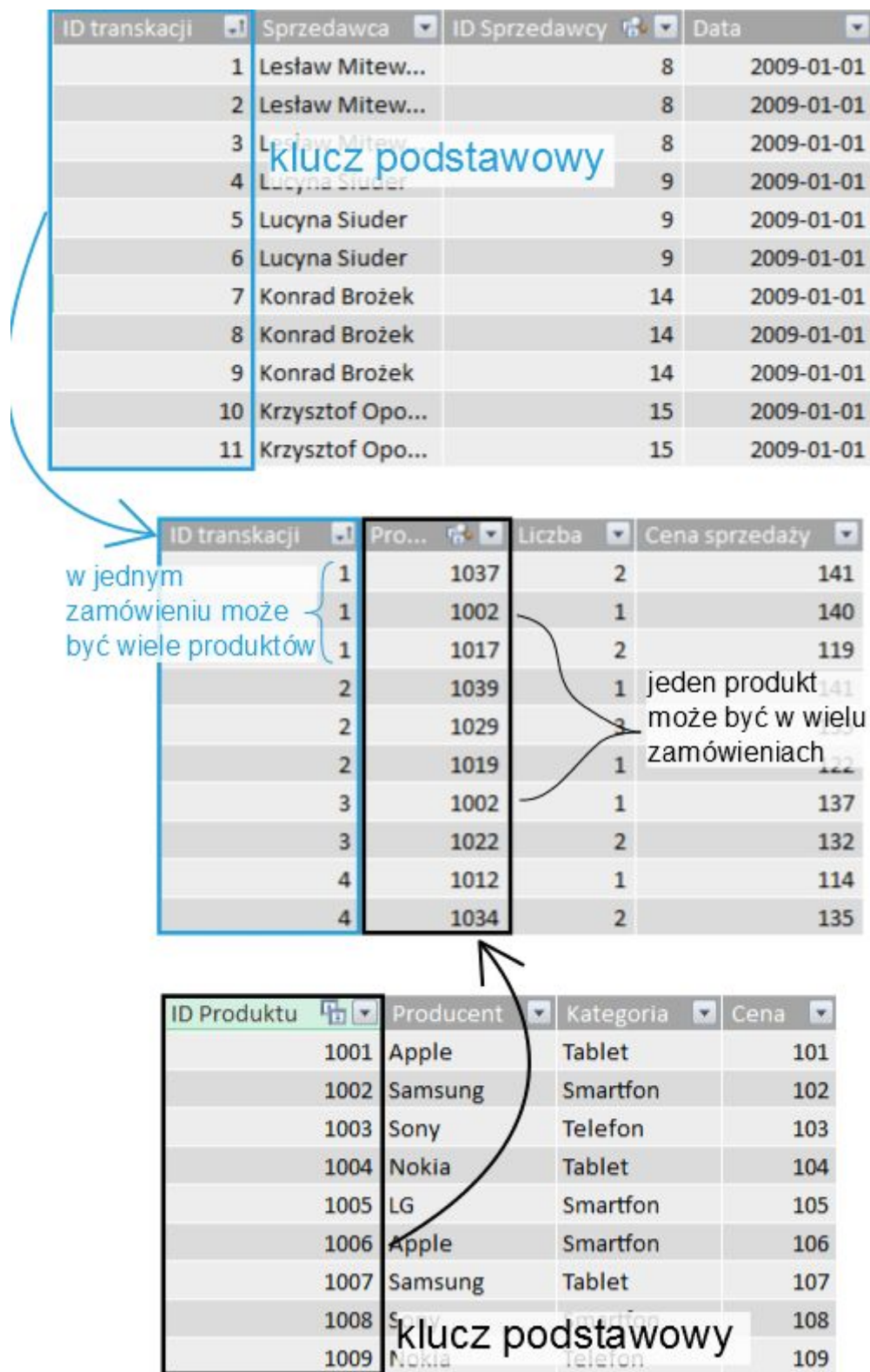
	nr producenta	typ produktu	Id
	30	1001	301001
	30	1003	301003
	30	1004	301004
	31	1001	311001
	31	1003	311003
	31	1005	311005
	32	1002	321002
	32	1005	321005
	33	1001	331001
	33	1002	331002
	33	1004	331004

Wyróżniamy trzy rodzaje relacji:

– **jeden do jednego** – w tego typu relacji jednemu rekordowi z tabeli A odpowiada tylko jeden wiersz z tabeli B. Rodzaj ten występuje stosunkowo rzadko, ponieważ wszystkie informacje przechowywane w ten sposób można zamieścić w jednej tabeli.

– **jeden do wielu** – jednemu rekordowi z tabeli A odpowiada wiele rekordów z tabeli B. Jest to najpowszechniejszy typ relacji.

– **wiele do wielu** – rekord w tabeli A może mieć wiele dopasowanych wiele wierszy z tabeli B oraz odwrotnie – rekord z tabeli B może mieć dopasowanych wiele wierszy z tabeli A. Taki typ jest możliwy do zdefiniowania tylko poprzez dodanie do modelu trzeciej tabeli (zwanej tabelą łączą), w której będą znajdowały się wartości kluczy podstawowych tabel A oraz B



**Klucz obcy** – kombinacja jednego lub wielu atrybutów tabeli, które wyrażają się w dwóch lub większej liczbie relacji. Wykorzystuje się go do tworzenia relacji pomiędzy parą tabel, gdzie w jednej tabeli ten zbiór atrybutów jest kluczem obcym, a w drugiej kluczem głównym.

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

### **Pierwsza postać normalna (1NF)**

Relacja jest w pierwszej postaci normalnej, jeśli:

- opisuje jeden obiekt,
- wartości atrybutów są elementarne (atomowe, niepodzielne) – każda kolumna jest wartością skalarną (atomową), a nie macierzą lub listą czy też czymkolwiek, co posiada własną strukturę,
- nie zawiera kolekcji (powtarzających się grup informacji)
- kolejność wierszy może być dowolna (znaczenie danych nie zależy od kolejności wierszy).

#### **Tabela przed normalizacją**

<b>Płeć</b>	<b>Imię</b>
Męska	Jan, Piotr, Zenon
Żeńska	Anna, Maria, Zofia

#### **Pierwsza postać normalna**

<b>Płeć</b>	<b>Imię</b>
Męska	Jan
Męska	Piotr
Męska	Zenon
Żeńska	Anna
Żeńska	Maria
Żeńska	Zofia

## Druga postać normalna (2NF)

Relacja jest w drugiej postaci normalnej wtedy i tylko wtedy, gdy jest w I postaci normalnej i żadna kolumna niekluczowa nie jest częściowo funkcyjnie zależna od jakiegokolwiek klucza potencjalnego.

### Przykład tabeli „Pracownicy” przed normalizacją

Imię	Nazwisko	Płeć	Stanowisko	Stawka za godzinę
Antoni	Anonim	Męska	Stolarz	10 zł
Natalia	Niewiadoma	Żeńska	Sekretarka	20 zł
Alina	Enigma	Żeńska	Sekretarka	20 zł

Klucz potencjalny składa się tu z dwóch pól: "Imię" oraz "Nazwisko". Przy założeniu, że każde imię ma przypisaną jedną płeć, czyli, że płeć zależy tylko od jednego z atrybutów klucza potencjalnego, tabela nie spełnia warunków na drugą postać normalną.

### Przykład tabeli „Pracownicy” po normalizacji do 2NF

Imię	Nazwisko	Stanowisko	Stawka za godzinę
Antoni	Anonim	Stolarz	10 zł
Natalia	Niewiadoma	Sekretarka	20 zł
Alina	Enigma	Sekretarka	20 zł

Każdy atrybut niekluczowy zależy od całego klucza potencjalnego.

### Nowa tabela „Płeć imienia” po normalizacji

Imię	Płeć
Antoni	Męska
Natalia	Żeńska
Alina	Żeńska

### Trzecia postać normalna (3NF)

Relacja jest w trzeciej postaci normalnej wtedy i tylko wtedy, gdy jest w II postaci normalnej i żaden atrybut niekluczowy nie jest zależny funkcyjnie od innych atrybutów niekluczowych.

#### Przykład tabeli „Pracownicy” przed normalizacją

Imię	Nazwisko	Stanowisko	Stawka za godzinę
Antoni	Anonim	Stolarz	10 zł
Natalia	Niewiadoma	Sekretarka	20 zł
Alina	Enigma	Sekretarka	20 zł

Klucz potencjalny składa się tu z dwóch pól: "Imię" oraz "Nazwisko". Oba atrybuty niekluczowe: "Stanowisko" oraz "Stawka za godzinę" są zależne od całego klucza potencjalnego- tzn. dany pracownik ma przyporządkowane jedno stanowisko i jedną stawkę godzinową.



Jeśli założymy, że każde stanowisko jest tak samo płatne, to wartości w kolumnie „Stawka za godzinę” są zależne jedynie od pola „Stanowisko”, a tylko pośrednio od klucza potencjalnego. Prowadzi to do powtarzania się wartości „20 zł”, co powoduje redundancję (nadmiarowość danych). Może to też prowadzić do anomalii i niespójności danych, gdy np. pani Alinie zmienimy stawkę na 25 zł, a zapomnimy zmienić stawki drugiej sekretarce, pani Natalii.

Sprowadzenie do III postaci normalnej będzie polegać na przeniesieniu stawek do osobnej tabeli, a w tabeli pracowników pozostawienie jedynie nazwy stanowiska.

**Tabela „Pracownicy” po normalizacji**

Imię	Nazwisko	Stanowisko
Antoni	Anonim	Stolarz
Natalia	Niewiadoma	Sekretarka
Alina	Enigma	Sekretarka

**Tabela „Stawki godzinowe” po normalizacji**

Stanowisko	Stawka
Stolarz	10 zł
Sekretarka	20 zł