

# Normalizacja

# Pierwsza postać normalna

- Jedynymi relacjami dozwolonymi w modelu relacyjnym są relacje spełniające następujący warunek:
- każda wartość w relacji, tj. każda wartość atrybutu w każdej krotce, jest wartością atomową (wartością nie rozkładalną)
- (1PN) oznacza, że tabela nie zawiera powtarzających się grup informacji, co znaczy, że każda kolumna jest wartością skalarną (atomową), a nie macierzą lub listą czy też czymkolwiek, co posiada własną strukturę

# Relacja nie znormalizowana

Pracownik	Języki
Jan Kowalski	angielski – słabo, niemiecki - dobrze
Adam Kot	rosyjski – bardzo dobrze

# Relacja nie znormalizowana

Pracownik	Znajomość języków	
	Język	Poziom
Jan Kowalski	angielski	słabo
	niemiecki	dobrze
Adam Kot	rosyjski	bardzo dobrze

# Relacja nie znormalizowana

- Relację "przed normalizacją" zdefiniowano na dwóch dziedzinach: *Pracownik* i *Znajomość Języków*
- Elementami dziedziny *Znajomość Języków* są również relacje (zdefiniowane na dziedzinach *Język* i *Poziom*)
- Relacja jest z punktu widzenia definicji relacją dwuczłonową, ale nie wszystkie jej dziedziny są proste (dziedzina prosta to taka, której wszystkie elementy są atomowe)

# Relacja znormalizowana

Pracownik	Język	Poziom
Jan Kowalski	angielski	słabo
Jan Kowalski	niemiecki	dobrze
Adam Kot	rosyjski	bardzo dobrze

# Relacja znormalizowana

- Relacja jest relacją trójczłonową, której wszystkie dziedziny są proste, jest więc znormalizowana
- Powodem tego jest uproszczenie struktury danych, które z kolei powoduje uproszczenie operatorów w subjęzyku danych
- Uproszczenia te nie ograniczają w niczym możliwości reprezentowania obiektów

# Relacja znormalizowana - nieporozumienia

Pracownik	Imię dziecka	Data ur. dziecka
Kowalski	Ania	01.01.2000
	Jaś	15.03.2001
Kot	Patrycja	20.10.2001
	Filemon	30.07.2003



# Relacja znormalizowana - nieporozumienia

Pracownik	Imię dziecka1	Data ur. dziecka1	Imię dziecka2	Data ur. dziecka2
Kowalski	Ania	01.01.2000	Jaś	15.03.2001
Kot	Patrycja	20.10.2001	Filemon	30.07.2003

# Pierwsza postać normalna

Przedmiot	Prowadzący	Student	Ocena
matematyka	prof. Lis	Jak Kot	2,0
matematyka	prof. Lis	Ewa Osa	3,0
matematyka	prof. Lis	Adam Struś	5,0

# Pierwsza postać normalna

- Powtarzająca się grupa danych to podzbiór relacji zawierający co najmniej dwa atrybuty, posiadająca własny klucz prosty, w którym istnieją powtarzające się krotki
- Powtarzanie się takich samych krotek wymuszone jest faktem, że mamy do czynienia z grupą dla której część atrybutów jest strukturą a nie wartością skalarną

# Pierwsza postać normalna

Przedmiot	Student	Ocena
matematyka	Jak Kot	2,0
matematyka	Ewa Osa	3,0
matematyka	Adam Struś	5,0

Przedmiot	Prowadzący
matematyka	prof. Lis

# Anomalie przy usuwaniu, wstawianiu i aktualizacji – baza klientów systemu CRM

Id Firmy	Nazwa	Nazwisko	Imie	Miejscowosc	Nazwa branży	Opis
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Handel	Skup i sprzedaż
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Gastronomia	Restauracje, bary szybkiej obsługi
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Samochody	Naprawa samochodów
3	NULL	Marchewka	Antoni	Olkusz	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
4	NULL	Słomka	Barbara	Brzesko	Gastronomia	Restauracje, bary szybkiej obsługi

# Anomalie przy usuwaniu

- Po usunięciu informacji o firmie P H U „Żagiel” tracimy informacje o branży „Handel” (Opis)

Id Firmy	Nazwa	Nazwisko	Imie	Miejscowosc	Nazwa branży	Opis
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Handel	Skup i sprzedaż
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Gastronomia	Restauracje, bary szybkiej obsługi
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Samochody	Naprawa samochodów
3	NULL	Marchewka	Antoni	Olkusz	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
4	NULL	Słomka	Barbara	Brzesko	Gastronomia	Restauracje, bary szybkiej obsługi

# Anomalie przy wstawianiu

- Wstawienie informacji nowym kliencie wymaga wpisania opisu branży mimo, że opis już istnieje

Id Firmy	Nazwa	Nazwisko	Imie	Miejscowosc	Nazwa branży	Opis
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Handel	Skup i sprzedaż
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Gastronomia	Restauracje, bary szybkiej obsługi
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Samochody	Naprawa samochodów
3	NULL	Marchewka	Antoni	Olkusz	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
4	NULL	Słomka	Barbara	Brzesko	Gastronomia	Restauracje, bary szybkiej obsługi
5	Dodana Firma	NULL	NULL	Dąbrowa G.	Usługi krawieckie	Szycie, fastrygowanie, przeróbki

# Anomalie przy aktualizacji

- Zmiana opisu „Usług krawieckich” musi być dokonana w czterech miejscach

Id Firmy	Nazwa	Nazwisko	Imie	Miejscowosc	Nazwa branży	Opis
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, poprawki
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Handel	Skup i sprzedaż
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, poprawki
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Gastronomia	Restauracje, bary szybkiej obsługi
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Samochody	Naprawa samochodów
3	NULL	Marchewka	Antoni	Olkusz	Usługi krawieckie	Szycie, fastrygowanie, poprawki
4	NULL	Słomka	Barbara	Brzesko	Gastronomia	Restauracje, bary szybkiej obsługi
5	Dodana Firma	NULL	NULL	Dąbrowa G.	Usługi krawieckie	Szycie, fastrygowanie, poprawki



## Druga postać normalna

- Relacja jest w drugiej postaci normalnej, jeśli każdy atrybut tej relacji nie wchodzący w skład żadnego klucza potencjalnego jest w pełni funkcyjnie zależny wyłącznie od wszystkich podrelacji klucza głównego

# Ta relacja nie jest w drugiej postaci normalnej

klucz

Id Firmy	Nazwa	Nazwisko	Imie	Miejscowosc	Nazwa branzy	Opis
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
1	P.H.U. "Żagiel"	NULL	NULL	Kraków	Handel	Skup i sprzedaż
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Gastronomia	Restauracje, bary szybkiej obsługi
2	Samo Zdrowie S.A.	NULL	NULL	Kraków	Samochody	Naprawa samochodów
3	NULL	Marchewka	Antoni	Olkusz	Usługi krawieckie	Szycie, fastrygowanie, przeróbki
4	NULL	Słomka	Barbara	Brzesko	Gastronomia	Restauracje, bary szybkiej obsługi
5	Dodana Firma	NULL	NULL	Dąbrowa G.	Usługi krawieckie	Szycie, fastrygowanie, przeróbki

zależność

brak zależności

## Ta relacja nie jest w drugiej postaci normalnej bo:

- kluczem w relacji jest podzbiór atrybutów *Id Firmy* i *Nazwa branży* bo powtórzenie krotki o dwóch identycznych wartościach tych atrybutów wskazywałoby na powtórne przypisanie tej samej branży tej samej firmie
- atrybut *Nazwa* zależy funkcjonalnie od atrybutu *Id Firmy* a nie zależy od atrybutu *Nazwa branży* (może być wiele firm w każdej z branż)

# Relacja po dekompozycji

Id Firmy	Nazwa branży
1	Usługi krawieckie
1	Handel
2	Usługi krawieckie
2	Gastronomia
2	Samochody
3	Usługi krawieckie
4	Gastronomia
5	Usługi krawieckie

Id Firmy	Nazwa	Nazwisko	Imie	Miejscowosc
1	P.H.U. "Żagiel"	NULL	NULL	Kraków
2	Samo Zdrowie S.A.	NULL	NULL	Kraków
3	NULL	Marchewka	Antoni	Olkusz
4	NULL	Słomka	Barbara	Brzesko
5	Dodana Firma	NULL	NULL	Dąbrowa G.

Nazwa branży	Opis
Usługi krawieckie	Szycie, fastrygowanie, przeróbki
Handel	Skup i sprzedaż
Gastronomia	Restauracje, bary szybkiej obsługi
Samochody	Naprawa samochodów

# Trzecia postać normalna

- Relacja jest w trzeciej postaci normalnej, jeśli:
  - jest w drugiej postaci normalnej
  - żaden atrybut nie będący kluczem nie jest funkcjonalnie związany z żadnym innym atrybutem nie będącym również kluczem

## Ta relacja nie jest w trzeciej postaci normalnej

Pracownik	PESEL	KodPocztowy	Miejscowość	Województwo
Jan Kowalski	12345678911	32-082	Bolechowice	małopolskie
Adam Kot	98977796666	30-150	Kraków	małopolskie
Ewa Lis	76281976372	32-082	Bolechowice	małopolskie



# Zależność funkcjonalna przechodnia

- Niech  $X$ ,  $Y$  i  $Z$  będą trzema rozłącznymi podzbiorami atrybutów danej relacji
- $Z$  jest przechodnio funkcjonalnie zależny od  $X$ , jeśli  $Z$  jest funkcjonalnie zależny od  $Y$  i  $Y$  jest funkcjonalnie zależny od  $X$  natomiast  $X$  nie jest zależny od  $Y$  i  $Y$  nie jest zależny od  $Z$

# Forma normalna Boyce-Codd'a

- Jest uzupełnieniem trzeciej postaci normalnej i jest niezbędna w przypadku gdy atrybuty będące kandydatami na klucze są:
  - wielokrotne,
  - złożone,
  - nakładające się na siebie



# Forma normalna Boyce'a-Codd'a

- Relacja jest w postaci Boyce-Codd'a jeżeli dla każdej nietrywialnej zależności między podzbiarami relacji zbiór będący wyznacznikiem jest zbiorem identyfikującym tej relacji
- Zależność  $X \rightarrow Y$  jest trywialna jeżeli  $Y$  jest podzbiorem  $X$
- Definicja BCNF zastępuje definicje, pierwszej, drugiej i trzeciej formy normalnej dodatkowo je poszerzając

# Przykład 1.

<i>IdPracownika</i>	<i>Zawód</i>	<i>Wykształcenie</i>	<i>Stawka</i>
1	ślusarz	podstawowe	5,20
1	tokarz	zawodowe	5,30
2	ślusarz	zawodowe	5,50
3	tokarz	zawodowe	5,30
4	ślusarz	podstawowe	5,20

kluczem w relacji jest podzbiór *IdPracownika, Zawód*  
zależność *Zawód, Wykształcenie*  $\rightarrow$  *Stawka* jest  
funkcjonalna i nietrywialna  
a *Zawód, Wykształcenie* nie jest zbiorem identyfikującym

# Przykład 1. - rozwiązanie

IdPracownika	Zawód	Wykształcenie
1	ślusarz	podstawowe
1	tokarz	zawodowe
2	ślusarz	zawodowe
3	tokarz	zawodowe
4	ślusarz	podstawowe

Zawód	Wykształcenie	Stawka
ślusarz	podstawowe	5,20
tokarz	zawodowe	5,30
ślusarz	zawodowe	5,50

## Przykład 2.

IdStudenta	Seminarium	Opiekun
1	marketing	Kowalski
1	kadry	Kozłowski
2	kadry	Janowski
3	marketing	Kowalski
4	informatyka	Macioł

ponieważ opiekun może mieć tylko jedno seminarium to kluczem w relacji jest podzbiór *IdStudenta, Seminarium* lub *IdStudenta, Opiekun* zależność *Opiekun*  $\rightarrow$  *Seminarium* jest funkcjonalna i nietrywialna a *Opiekun* nie jest zbiorem identyfikującym

## Przykład 2. - rozwiązanie

IdStudenta	Opiekun
1	Kowalski
1	Kozłowski
2	Janowski
3	Kowalski
4	Macioł

Seminarium	Opiekun
marketing	Kowalski
kadry	Kozłowski
kadry	Janowski
informatyka	Macioł

# Czwarta forma normalna

- Relacja jest w czwartej formie normalnej wtedy i tylko wtedy, gdy jest w trzeciej postaci normalnej i nie zawiera wielowartościowej zależności atrybutów

# Zależność wielowartościowa

Nazwisko	Imię dziecka	Znajomość języków
Kot	Ania	niemiecki
Kot	Jaś	angielski
Kot	Ania	angielski
Kot	Jaś	niemiecki
Słoń	Ola	niemiecki
Słoń	Ola	angielski

# Zależność wielowartościowa - rozwiązanie

Nazwisko	Imię dziecka
Kot	Ania
Kot	Jaś
Słoń	Ola

Nazwisko	Znajomość języków
Kot	niemiecki
Kot	angielski
Słoń	niemiecki
Słoń	angielski