



## Bazy Danych

### Projektowanie. Normalizacja.

Piotr Macioł  
WMiUR, KSiM,  
pmaciol@agh.edu.pl  
85, pok. 606



## Projektowanie bazy danych:

- **Projektowanie logicznej struktury bazy:**
  - » **Etap I:** określenie encji i zdefiniowanie atrybutów opisujących encje
    - przyporządkowanie encji do zjawisk
    - standaryzacja nazw i formatów
    - identyfikacja źródeł danych
  - » **Etap II:** określenie związków między encjami
    - identyfikacja typu związków (relacji) (1-1, 1-M, N-M)
  - » **Etap III:** normalizacja relacji
    - obniżenie redundancji i wyeliminowanie anomalii (usuwanie, wstawiania i aktualizacji)
- **Projektowanie fizycznej struktury bazy:**
  - » nałożenie struktury logicznej na fizyczne urządzenia

KSiM, WMiUR, AGH

3



## Etapy projektowania baz danych:

1. **Specyfikacja wymagań użytkownika** - określenie zjawisk, dostępności i użyteczności danych, ich formatu i sposobów obliczeń, cele, zakres i kontekst systemu
2. **Projektowanie koncektualne** - projektowanie schematu E-R bazy. Użycie modelu E-R wpływa również na realizację pozostałych faz.
3. **Specyfikacja wymagań funkcjonalnych** - dokładny opis wymagań klienta i wszystkich przyszłych użytkowników systemu
4. **Projektowanie logiczne i fizyczne**
5. **Implementacja**

KSiM, WMiUR, AGH

2



## Wybór schematu relacji:

- Projekt bazy danych polega na znalezieniu właściwych schematów relacji tworzących bazę danych
- **Niewłaściwy projekt** (niewłaściwe schematy) mogą prowadzić do
  - » Redundancji (powtarzania informacji).
  - » Niemożliwości reprezentowania pewnych informacji.
  - » Anomalii związanych z operowaniem danymi (głównie modyfikacją danych)
- **Cele:**
  - » Unikanie redundancji danych
  - » Zapewnienie reprezentowania związków między danymi
  - » Zachowanie warunków integralności (umożliwienie kontroli warunków integralności podczas modyfikacji danych).

KSiM, WMiUR, AGH

4



## Modelowanie koncektualne:

- Model pojęciowy (koncektualny, podstawowy):**
- Zapis wymagań w postaci sformalizowanej, abstrahujący od problemów implementacyjnych („co” a nie „jak”)
  - Zawiera
    - » modele graficzne
    - » tekstowe (ale też sformalizowane) uzupełnienia modeli graficznych
  - Jest ważny, ponieważ:
    - » reprezentuje istotę wymagań
    - » nie zależy od zmiennych możliwości implementacji
    - » jest dobrą podstawą do projektowania systemów mających funkcjonować przez wiele lat

KSiM, WMiUR, AGH

5



## Wymagania:

- Spełnienie wymagań użytkownika jest naszym celem działania
- Wymagania muszą więc być dokładnie znane
- Poprawne i kompletne sformułowanie wymagań na ogół nie jest proste
- Należy odróżniać *funkcje* systemu od *mechanizmów*, czyli sposobów implementacji funkcji
- Modelować dane niezależnie od przyszłej implementacji - **modelowanie E/R**
- Tworząc model pojęciowy nie można formułować zastrzeżeń co do realizowalności wymagań
- Zaprojektować dobry schemat - **normalizacja**

KSiM, WMiUR, AGH

6



## „Student lubi piwo...”

Lp.	Imię	Nazwisko	Marka piwa	Nazwa piwa	Gatunek	Alkohol [%]
1.	Stefan	Zeromski	Okocim	Palone	lager	7
2.	Stefan	Zeromski	Zywiec	Porter	porter baltycki	9,5
3.	Stefan	Zeromski	Cornelius	ALE ALE	ale	5,8
4.	Eliza	Orzechowska	Ciechan	Miodowe	piłzner	5,7
5.	Eliza	Orzechowska	Redd's	Cranberry	piłzner	4,5
6.	Adam	Mickiewicz	Okocim	Porter	porter baltycki	8,3
7.	Friedrich	Nietzsche	Paulaner	Hefe-Weissbier Naturtrüb	pszeniczne	5,5
8.	Friedrich	Nietzsche	Ciechan	Pszeniczne	pszeniczne	5,6
9.	Tadeusz	Konwicki	Zywiec	Zywiec Beer	piłzner	5,6
10.	Rafał	Wojczak	Zywiec	Zywiec Beer	piłzner	5,6
11.	Witława	Szymborska	Zywiec	Zywiec Beer	piłzner	5,6
12.	Johann	Goethe	Zywiec	Zywiec Beer	piłzner	5,6
13.	James	Joyce	Carlsberg	Carlsberg	lager	5
14.	Francis	Fitzgerald	Haineken	Haineken Lager Beer	lager	5
15.	Maria	Skłodowska-Curie	Zywiec	Desperados	piłzner	6
16.	Zofia	Nałkowska	komla	Miodowe	piłzner	5,7
17.	Fiodor	Dostojewski	Tykie	Gronie	piłzner	5,6
18.	Fryderyk	Chopin	Tykie	Gronie	piłzner	6,6
19.	Stefan	Batory	Tykie	Gronie	piłzner	7,6

KSIM, WIMiP, AGH



## „Student lubi piwo...”

STUDENT			PIWO			
L.p.	Imię	Nazwisko	Marka piwa	Nazwa piwa	Gatunek	Alkohol [%]
1.	Stefan	Zeromski	Okocim	Palone	lager	7
2.	Stefan	Zeromski	Zywiec	Porter	porter baltycki	9,5
3.	Stefan	Zeromski	Cornelius	ALE ALE	ale	5,8
4.	Eliza	Orzechowska	Ciechan	Miodowe	piłzner	5,7
5.	Eliza	Orzechowska	Redd's	Cranberry	piłzner	4,5
6.	Adam	Mickiewicz	Okocim	Porter	porter baltycki	8,3
7.	Friedrich	Nietzsche	Paulaner	Hefe-Weissbier Naturtrüb	pszeniczne	5,5
8.	Friedrich	Nietzsche	Ciechan	Pszeniczne	pszeniczne	5,6
9.	Tadeusz	Konwicki	Zywiec	Zywiec Beer	piłzner	5,6
10.	Rafał	Wojczak	Zywiec	Zywiec Beer	piłzner	5,6
11.	Witława	Szymborska	Zywiec	Zywiec Beer	piłzner	5,6
12.	Johann	Goethe	Zywiec	Zywiec Beer	piłzner	5,6
13.	James	Joyce	Carlsberg	Carlsberg	lager	5
14.	Francis	Fitzgerald	Haineken	Haineken Lager Beer	lager	5
15.	Maria	Skłodowska-Curie	Zywiec	Desperados	piłzner	6
16.	Zofia	Nałkowska	komla	Miodowe	piłzner	5,7
17.	Fiodor	Dostojewski	Tykie	Gronie	piłzner	5,6
18.	Fryderyk	Chopin	Tykie	Gronie	piłzner	6,6
19.	Stefan	Batory	Tykie	Gronie	piłzner	7,6

KSIM, WIMiP, AGH



## „Student lubi piwo...”

STUDENT			PIWO		GATUNEK	PIWO
Lp.	Imię	Nazwisko	Marka piwa	Nazwa piwa	Gatunek	Alkohol [%]
1.	Stefan	Zeromski	Okocim	Palone	lager	7
2.	Stefan	Zeromski	Zywiec	Porter	porter baltycki	9,5
3.	Stefan	Zeromski	Cornelius	ALE ALE	ale	5,8
4.	Eliza	Orzechowska	Ciechan	Miodowe	piłzner	5,7
5.	Eliza	Orzechowska	Redd's	Cranberry	piłzner	4,5
6.	Adam	Mickiewicz	Okocim	Porter	porter baltycki	8,3
7.	Friedrich	Nietzsche	Paulaner	Hefe-Weissbier Naturtrüb	pszeniczne	5,5
8.	Friedrich	Nietzsche	Ciechan	Pszeniczne	pszeniczne	5,6
9.	Tadeusz	Konwicki	Zywiec	Zywiec Beer	piłzner	5,6
10.	Rafał	Wojczak	Zywiec	Zywiec Beer	piłzner	5,6
11.	Witława	Szymborska	Zywiec	Zywiec Beer	piłzner	5,6
12.	Johann	Goethe	Zywiec	Zywiec Beer	piłzner	5,6
13.	James	Joyce	Carlsberg	Carlsberg	lager	5
14.	Francis	Fitzgerald	Haineken	Haineken Lager Beer	lager	5
15.	Maria	Skłodowska-Curie	Zywiec	Desperados	piłzner	6
16.	Zofia	Nałkowska	komla	Miodowe	piłzner	5,7
17.	Fiodor	Dostojewski	Tykie	Gronie	piłzner	5,6
18.	Fryderyk	Chopin	Tykie	Gronie	piłzner	6,6
19.	Stefan	Batory	Tykie	Gronie	piłzner	7,6

KSIM, WIMiP, AGH



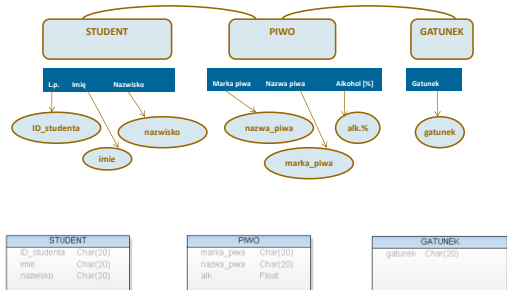
## „Student lubi piwo...”

STUDENT			PIWO		GATUNEK	PIWO
Lp.	Imię	Nazwisko	Marka piwa	Nazwa piwa	Gatunek	Alkohol [%]
1.	Stefan	Zeromski	Okocim	Palone	lager	7
2.	Stefan	Zeromski	Zywiec	Porter	porter baltycki	9,5
3.	Stefan	Zeromski	Cornelius	ALE ALE	ale	5,8
4.	Eliza	Orzechowska	Ciechan	Miodowe	piłzner	5,7
5.	Eliza	Orzechowska	Redd's	Cranberry	piłzner	4,5
6.	Adam	Mickiewicz	Okocim	Porter	porter baltycki	8,3
7.	Friedrich	Nietzsche	Paulaner	Hefe-Weissbier Naturtrüb	pszeniczne	5,5
8.	Friedrich	Nietzsche	Ciechan	Pszeniczne	pszeniczne	5,6
9.	Tadeusz	Konwicki	Zywiec	Zywiec Beer	piłzner	5,6
10.	Rafał	Wojczak	Zywiec	Zywiec Beer	piłzner	5,6
11.	Witława	Szymborska	Zywiec	Zywiec Beer	piłzner	5,6
12.	Johann	Goethe	Zywiec	Zywiec Beer	piłzner	5,6
13.	James	Joyce	Carlsberg	Carlsberg	lager	5
14.	Francis	Fitzgerald	Haineken	Haineken Lager Beer	lager	5
15.	Maria	Skłodowska-Curie	Zywiec	Desperados	piłzner	6
16.	Zofia	Nałkowska	komla	Miodowe	piłzner	5,7
17.	Fiodor	Dostojewski	Tykie	Gronie	piłzner	5,6
18.	Fryderyk	Chopin	Tykie	Gronie	piłzner	6,6
19.	Stefan	Batory	Tykie	Gronie	piłzner	7,6

KSIM, WIMiP, AGH



## „Student lubi piwo...”

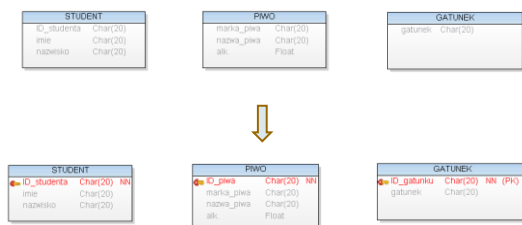


KSIM, WIMiP, AGH

11



## „Student lubi piwo...”



KSIM, WIMiP, AGH

12



## „Student lubi piwo...”

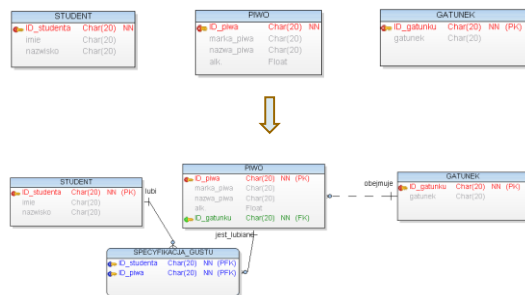
STUDENT		PIWO		GATUNEK		
Lp.	Imię	Nazwisko	Marka piwa	Nazwa piwa	Gatunek	Alkohol [%]
1.	Stefan	Zeromski	Okocim	Palone	lager	7
2.	Stefan	Zeromski	Zywiec	Porter	porter	9,5
3.	Stefan	Zeromski	Coronet	Alc-ice	ale	5,8
4.	Eliza	Oroszkowa	Ciechan	Miodowe	piłzner	5,7
5.	Eliza	Oroszkowa	Redd's	Cranberry	piłzner	4,5
6.	Adam	Michalewicz	Okocim	Porter	porter	8,3
7.	Friedrich	Nietzsche	Coronet	Stefan's Natural	piłzner	5,5
8.	Friedrich	Nietzsche	Ciechan	Poznańskie	piłzner	5,6
9.	Tadeusz	Konwici	Zywiec	Zywiec Beer	piłzner	5,6
10.	Rafał	Wojasicki	Zywiec	Zywiec Beer	piłzner	5,6
11.	Władysław	Szybonski	Sawier	Zywiec Beer	piłzner	5,6
12.	Johann	Goethe	Zywiec	Zywiec Beer	piłzner	5,6
13.	James	Joyce	Carlsberg	Carlsberg	lager	5
14.	Francis	Fitzgerald	Haineken	Haineken Lager Beer	lager	5
15.	Maria	Skłodowska-Curie	Zywiec	Desperados	piłzner	6
16.	Zofia	Kurkowska	Komła	Miodowe	piłzner	5,7
17.	Fredrik	Dostopewski	Tyskie	Gronia	piłzner	5,6
18.	Fryderyk	Chopin	Tyskie	Gronia	piłzner	6,6
19.	Stefan	Batory	Tyskie	Gronia	piłzner	7,6

KSIM, WIMiP, AGH

13



## „Student lubi piwo...”

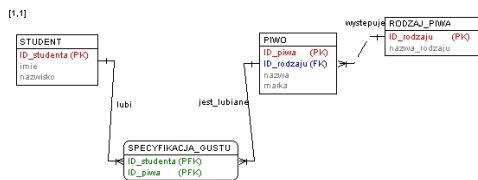


KSIM, WIMiP, AGH

14



## starsza wersja...



KSIM, WIMiP, AGH

15



## Normalizacja.

KSIM, WIMiP, AGH

16



## Przykład wad bazy danych:

- **Rozważmy schemat relacji:**
  - » `pracownik oddzialu = <nazwa_oddzialu, numer_oddzialu, adres_oddzialu, pesel, imie, nazwisko, stanowisko, wynagrodzenie>`
- **Redundancja:**
  - » Dane `nazwa_oddzialu, numer_oddzialu, adres_oddzialu` są pamiętane dla każdego pracownika
  - » Marnowanie miejsca (przestrzeni potrzebnej dla przechowywania danych)
  - » Komplikacje (aktualizacja, błąd, usuwanie...)
- **Brak możliwości reprezentowania pewnych informacji**
  - » Nie można reprezentować informacji o oddziałach, których powołanie dopiero jest planowane i nie zatrudniły jeszcze pracowników
  - » Rozwiązaniem mogło być zastosowanie wartości NULL, ale takie rozwiązanie także stwarza pewne problemy
  - » Nie możemy zaprezentować w prosty sposób wszystkich pracowników z danego miasta, ponieważ miejscowość nie jest wyodrębnionym atrybutem

KSIM, WIMiP, AGH

17



## Rodzaje anomalii:

- **Anomalia dołączania** - wada jest to, że musimy wpisywać wszystko albo nic,
  - » np. planuje się powołanie nowego oddziału, nie możemy jednak zapisać w bazie miejscowości, adresu i nazwy zakupionego budynku, dopóki nowy oddział nie zatrudni pracowników
- **Anomalia aktualizacji** - np. oddział firmy został przeniesiony, przez co musimy aktualizować jego adres dla każdego pracownika, a przez przypadek omijamy Jana Kowalskiego
- **Anomalia usuwania** – zamknięto oddział firmy, w związku z czym usunięto z bazy wszystkie rekordy zawierające jego nazwę, tym samym usunięto wszystkie dane dotyczące pracowników

KSIM, WIMiP, AGH

18



## Dekompozycja:

- Można **zdekomponować** schemat relacji – *pracownik\_oddzialu*:
  - » *oddzial* = *<nazwa oddzialu, numer oddzialu, miejscowosc\_oddzialu, adres\_oddzialu>*
  - » *pracownik* = *<numer\_oddzialu, pesel, imie, nazwisko, stanowisko, wynagrodzenie>*
- Wszystkie atrybuty oryginalnego schematu (*R*) muszą się pojawić w dekompozycji (*R*<sub>1</sub>, *R*<sub>2</sub>):

$$R = R_1 \cup R_2$$

- W przypadku gdy relacja nie posiada właściwej postaci należy dokonać dekompozycji relacji *R* na *<R*<sub>1</sub>, *R*<sub>2</sub>, ..., *R*<sub>n</sub>>
- Proces „dochodzenia” do „właściwej” postaci określa się mianem **normalizacji**

KSIM, WIMiP, AGH

29



## Postacie normalne:

- Pierwsza (1PN)
- Druga (2PN)
- Trzecia (3PN)
- Boyce’a-Codd’a (PNBC)
- Czwarta (4PN)
- Piąta (5PN)

KSIM, WIMiP, AGH

30



## Pierwsza postać normalna

- Relacja jest w pierwszej postaci normalnej, jeśli każda wartość atrybutu w każdej krotce tej relacji jest **wartością elementarną**, czyli nierozkładalną.
- Relacja jest w pierwszej postaci normalnej, jeśli nie ma **powtarzających się grup**.

KSIM, WIMiP, AGH

21



## Przykład:

pracownik_oddzialu							
numer_oddzialu	nazwa_oddzialu	miejscowosc_oddzialu	adres_oddzialu	pesel	imie	nazwisko	stanowisko
A1	Afflaton	Kraków	Słoneczna 2/6	21108106713	Jan	Głuchomski	Sprzedawca
A1	Afflaton	Kraków	Słoneczna 2/6	1505768456	Wiktoria	Sakowski	Kierowca
A2	Afflaton	Kraków	Fabryczna 13/7	13015778123	Magdalena	Waligorska	Sprzedawca
B1	Betatrix	Hel	Leśna 10	30127623435	Ignacy	Kuchta	Kierownik sekcji
B1	Betatrix	Hel	Leśna 10	20035934567	Marcyna	Sepiel	Kierowca
B2	Betatrix	Hel	Leśna 10	31169675846	Jadwiga	Makusz	Sprzedawca
B3	Betatrix	Gdynia	Widokowa 12/26	21016922387	Stanisław	Marekowska	Kierowca
B3	Betatrix	Gdynia	Portowa 23	17037467575	Zuzanna	Wiał	Sprzedawca

oddzial			
numer	nazwa	miejscowosc	adres
A1	Afflaton	Kraków	Słoneczna 2/6
A2	Afflaton	Kraków	Fabryczna 13/7
B1	Betatrix	Hel	Leśna 10
B2	Betatrix	Gdynia	Widokowa 12/26
B3	Betatrix	Gdynia	Portowa 23

pracownik						
pesel	imie	nazwisko	stanowisko	wynagrodzenie	numer_oddzialu	adres
21108106713	Jan	Głuchomski	Sprzedawca	2 700,00 zł	A1	Słoneczna 2/6
1505768456	Wiktoria	Sakowski	Kierowca	1 800,00 zł	A1	Słoneczna 2/6
13015778123	Magdalena	Waligorska	Sprzedawca	2 700,00 zł	A2	Fabryczna 13/7
30127623435	Ignacy	Kuchta	Kierownik sekcji	3 500,00 zł	B1	Leśna 10
20035934567	Marcyna	Sepiel	Kierowca	1 800,00 zł	B1	Leśna 10
31169675846	Jadwiga	Makusz	Sprzedawca	2 700,00 zł	B2	Leśna 10
21016922387	Stanisław	Marekowska	Kierowca	1 800,00 zł	B3	Widokowa 12/26
17037467575	Zuzanna	Wiał	Sprzedawca	2 700,00 zł	B3	Portowa 23

KSIM, WIMiP, AGH

22



## Zależności funkcyjne i wielowartościowe:

- Zależność funkcyjna** oznacza, że znając wartość jednego atrybutu, zawsze możemy określić wartość innego. Symbolem stosowanym w teorii relacji jest strzałka umieszczona pomiędzy dwoma atrybutami, na przykład:

$$X \rightarrow Y \quad (X \text{ określa } Y)$$

przykład: gdy znamy numer PESEL naszego pracownika, możemy określić jego nazwisko

- Zależność wielowartościowa** oznacza, że znając wartość jednego atrybutu, możemy zawsze określić wartości zbioru innego atrybutu. W teorii relacji używa się symbolu zależności wielowartościowej w postaci podwójnej strzałki, na przykład:

$$X \twoheadrightarrow Y \quad (X \text{ określa wiele } Y)$$

przykład: znając numer oddziału możemy określić nazwiska wszystkich zatrudnionych pracowników

KSIM, WIMiP, AGH

23



## Teoria postaci normalnych relacji:

Formalnie **zależność danych** można zdefiniować następująco. Schemat relacji oznaczamy przez  $R\langle A_1 \dots A_N \rangle$ , gdzie  $A_1 \dots A_N$  są atrybutami relacji. Niech  $X$  i  $Y$  będą podzbiarami zbioru atrybutów.

$$X \subset \{A_1 \dots A_N\}, Y \subset \{A_1 \dots A_N\}.$$

Zależność danych zapisujemy w postaci:

$$X \rightarrow Y$$

i mówimy, że podzbiór atrybutów  $Y$  **zależy funkcyjnie** od podzbioru atrybutów  $X$ , jeżeli nie jest możliwe, by relacja  $R$  zawierała dwie krotki mające składowe zgodne (tzn. identyczne dla wszystkich atrybutów ze zbioru  $X$ ) i jednocześnie co najmniej jedną niezgodną składową dla atrybutów ze zbioru  $Y$ .

KSIM, WIMiP, AGH

24



## Klucz:

- **Kluczem relacji** nazywamy taki zbiór atrybutów tej relacji, których kombinacje wartości jednoznacznie identyfikują każdą krotkę tej relacji a żaden podzbiór tego zbioru nie posiada tej własności. W kluczu nie może zawierać się wartość NULL.
- Klucz jest **kluczem prostym**, jeżeli powyżej opisany zbiór jest jednoelementowy - w przeciwnym razie mówimy o **kluczu złożonym**.
- W ogólności, w relacji można wyróżnić wiele kluczy, które nazywamy **kluczami potencjalnymi**. Wybrany klucz spośród kluczy potencjalnych nazywamy **kluczem głównym**.

KSIM, WIMiP, AGH

25



## Druga i trzecia postać normalna:

- Relacja jest w **drugiej** postaci normalnej, jeśli jest w 1PN oraz każdy atrybut tej relacji **nie wchodzący** w skład żadnego klucza potencjalnego jest w **pełni funkcyjnie zależny** od wszystkich kluczy potencjalnych tej relacji.
- Relacja jest w **2PN** jeżeli każdy atrybut nie wchodzący w skład klucza **zależy od klucza a nie od jego części**.
- Relacja będąca w pierwszej postaci normalnej, jest równocześnie w **drugiej** postaci normalnej, jeśli wszystkie jej klucze potencjalne są **kluczami prostymi**.
- Dana relacja jest w **trzeciej** postaci normalnej, jeśli jest ona w drugiej postaci normalnej i każdy jej atrybut nie wchodzący w skład żadnego klucza potencjalnego **nie jest przechodnio funkcyjnie zależny** od żadnego klucza potencjalnego tej relacji.

KSIM, WIMiP, AGH

26



## Druga i trzecia postać normalna:

Inaczej mówiąc, wszystkie niekluczowe kolumny są określane

**kluczem, całym kluczem i tylko kluczem,  
„...tak nam pomóż Codd”**

**The key, the whole key, and nothing but the  
key, so help me Codd**

KSIM, WIMiP, AGH

27



## Przykład:

Protokoły zaliczeń i egzaminów									
Rodzaj studiów	Rok studiów	Rok akademicki	Przedmiot	Przewodzący	Forma	Termin	Student	Ocena	Data
				Imię i nazwisko			nr albumu	Nazwisko i imię	
stacjonarne	I	2005/2006	matematyka	prof. Jan Kot	egzamin	I	12345	Pies Jan	2,0 20.01.2006
stacjonarne	I	2005/2006	matematyka	prof. Jan Kot	egzamin	I	12346	Lis Ewa	3,0 20.01.2006
zawoczne	I	2005/2006	matematyka	prof. Jan Kot	egzamin	II	12347	Stoń Adam	3,0 20.01.2006
stacjonarne	I	2005/2006	wychowanie fizyczne	mgr Janina Myz	zaliczenie	I	12345	Pies Jan	5,0 26.01.2006
stacjonarne	I	2005/2006	wychowanie fizyczne	mgr Janina Myz	zaliczenie	I	12346	Lis Ewa	3,0 26.01.2006
zawoczne	I	2005/2006	wychowanie fizyczne	mgr Janina Myz	zaliczenie	II	12347	Stoń Adam	3,0 26.01.2006

Wartości atrybutów nie są elementarne



## Przykład (c.d.):

**Klucz potencjalny:** <rodzaj\_studiów, rok\_studiów, rok\_akademicki, przedmiot, forma, termin, student\_nr\_albumu>

### Zależności funkcyjne

Protokoły zaliczeń i egzaminów

rodzaj_studiów	rok_studiów	rok_akademicki	przedmiot	przewodzący	przewodzący	przewodzący	forma	termin	student_nr_albumu	ocena	data
				Imię	Nazwisko	Nazwisko					
stacjonarne	I	2005/2006	matematyka	prof. Jan Kot	egzamin	I	12345	Jan Pies	2,0	20.01.2006	
stacjonarne	I	2005/2006	matematyka	prof. Jan Kot	egzamin	I	12346	Ewa Lis	3,0	20.01.2006	
zawoczne	I	2005/2006	matematyka	prof. Jan Kot	egzamin	II	12347	Adam Stoń	3,0	20.01.2006	
stacjonarne	I	2005/2006	wychowanie fizyczne	mgr Janina Myz	zaliczenie	I	12345	Jan Pies	5,0	26.01.2006	
stacjonarne	I	2005/2006	wychowanie fizyczne	mgr Janina Myz	zaliczenie	I	12346	Ewa Lis	3,0	26.01.2006	
zawoczne	I	2005/2006	wychowanie fizyczne	mgr Janina Myz	zaliczenie	II	12347	Adam Stoń	3,0	26.01.2006	

Powtarzające się grupy

KSIM, WIMiP, AGH

29



Przykład relacji i kluczy

przewodzący	nr_albumu	Imię	Nazwisko
20050504557	prof.	Jan	Kot
31106675946	mgr	Janina	Myz
20057665456	dr hab. inż.	Wojciech	Kuna

id_przedmiotu	rodzaj_studiów	rok_studiów	rok_akademicki	przedmiot	przewodzący	forma
1	stacjonarne	I	2005/2006	matematyka	20050504557	egzamin
2	zawoczne	II	2005/2006	matematyka	15057665456	egzamin
3	stacjonarne	I	2005/2006	wychowanie fizyczne	31106675946	zaliczenie
4	zawoczne	II	2005/2006	wychowanie fizyczne	31106675946	zaliczenie

id_przedmiotu	termin	nr_albumu	Imię	Nazwisko	ocena	data
1	I	12345	Jan	Pies	2,0	20.01.2006
2	I	12346	Ewa	Lis	3,0	20.01.2006
3	II	12347	Adam	Stoń	3,0	20.01.2006
4	I	12345	Jan	Pies	5,0	26.01.2006
5	I	12346	Ewa	Lis	3,0	26.01.2006
6	I	12347	Adam	Stoń	3,0	26.01.2006

**Klucz główny relacji protokoły:**  
<id\_przedmiotu, termin, nr\_albumu>

Student	nr_albumu	Imię	Nazwisko
12345	Jan	Pies	
12346	Ewa	Lis	
12347	Adam	Stoń	

KSIM, WIMiP, AGH

30



### 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



KSIM, WIMiP, AGH

31



### Zależność funkcjonalna przechodnia

- Niech  $X$ ,  $Y$  i  $Z$  będą trzema rozłącznymi podzbiórmi 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  $Z$

KSIM, WIMiP, AGH

32



### 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

KSIM, WIMiP, AGH

33



### Forma normalna Boyce'a-Codd'a

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

KSIM, WIMiP, AGH

34



### Forma normalna Boyce-Codd'a

IdStudenta	Seminarium	Opiekun
1	metalurgia	Kowalski
1	inżynieria wiedzy	Kozłowski
2	inżynieria wiedzy	Janowski
3	metalurgia	Kowalski
4	informatyka	Zajac

- opiekun może mieć tylko jedno seminarium, więc kluczem w relacji może być: podzbiór **IdStudenta + Seminarium** lub **IdStudenta + Opiekun**
- Przy tej drugiej opcji nie spełniona jest III postać normalna, ponieważ klucz **IdStudenta + Opiekun** nie determinują seminarium, która zależy tylko i wyłącznie od opiekuna. Z tego powodu jako kandydata na klucz główny powinno się wziąć pod uwagę **IdStudenta + Seminarium**
- zależność **Opiekun  $\rightarrow$  Seminarium** jest funkcjonalna i nietrywialna a **Opiekun** nie jest zbiorem identyfikującym

KSIM, WIMiP, AGH

35



### Forma normalna Boyce-Codd'a

- Załóżmy, że chcemy przydzielić nazwiska prowadzących do odpowiednich seminariów przed rozpoczęciem zapisów studentów na konkretny przedmiot.
- Nie możemy tego zrobić, ponieważ nazwisko studenta jest częścią klucza głównego i z tego powodu, przy wstawianiu nowego wiersza tabeli trzeba wstawić nazwisko studenta (anomalia dołączania).
- Również może zaistnieć anomalia w przypadku, gdy wystąpi potrzeba zmiany nazwiska prowadzącego przedmiot. Wtedy będzie trzeba wstawić nowe nazwisko do wielu wierszy jednocześnie.
- Aby zaradzić tym problemom należy sprowadzić tabelę do postaci Boyce-Codda. W tym celu zostaną utworzone 3 tabele:

KSIM, WIMiP, AGH

36



### Forma normalna Boyce-Codd'a

IdStudenta	Opiekun	Seminarium	Opiekun
1	Kowalski	metalurgia	Kowalski
1	Kozłowski	inżynieria wiedzy	Kozłowski
2	Janowski	inżynieria wiedzy	Janowski
3	Kowalski	informatyka	Zając
4	Zając		

IdStudenta	Seminarium
1	metalurgia
1	inżynieria wiedzy
2	inżynieria wiedzy
3	metalurgia
4	informatyka

KSIM, WMiP, AGH

37



### Forma normalna Boyce-Codd'a

IdPracownika	Zawód	Wykształcenie	Stawka
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** lub **IdPracownika, Wykształcenie**  
 zależność **Zawód, Wykształcenie** → **Stawka** jest funkcjonalna i nietrywialna  
 a **Zawód, Wykształcenie** nie jest zbiorem identyfikującym

KSIM, WMiP, AGH

38



### Podsumowanie:

- Normalizacja ma na celu takie przekształcenie relacji, by uniknąć **redundancji i anomalii**.
- Przekształcenie relacji do kolejnych postaci normalnych wiąże się najczęściej ze **zmniejszeniem ilości pamięci** potrzebnej do **przechowywania informacji**.
- Unikanie powtórzeń pozwala na **łatwiejszą i szybszą aktualizację** danych.
- Doprowadzenie bazy do wysokiej postaci normalizacji może **spowolnić odczyt** w dużych bazach ze względu na skomplikowany schemat danych.
- W większości przypadków po znormalizowaniu bazy danych przychodzi kolej na rozważenie możliwości wykonania **odwrotnej operacji (denormalizacji)**, polegającej na połączeniu niektórych znormalizowanych tabel, a to z myślą o przyspieszeniu dostępu do pewnych danych.

KSIM, WMiP, AGH

39



### Czwarta forma normalna

- Relacja jest w czwartej formie normalnej (IV PN) wtedy i tylko wtedy, gdy jest w trzeciej postaci normalnej i nie zawiera nietrywialnej wielowartościowej zależności atrybutów

KSIM, WMiP, AGH

40



### Zależność wielowartościowa

- Podzbiór atrybutów **Y** jest wielowartościowo funkcjonalnie zależny od podzbioru **X** w schemacie **R**, jeżeli dla dowolnej relacji **r** w schemacie **R** i dla dowolnej pary krotek **t<sub>1</sub>** i **t<sub>2</sub>** z relacji **r** istnieje taka para krotek że:

$$s_1[X] = s_2[X] = t_1[X] = t_2[X] \text{ i}$$

$$s_1[Y] = t_1[Y] \text{ i } s_2[R-X-Y] = t_2[R-X-Y] \text{ i}$$

$$s_2[Y] = t_2[Y] \text{ i } s_2[R-X-Y] = t_1[R-X-Y]$$

KSIM, WMiP, AGH

41



### Zależność wielowartościowa

	X	Y	R-X-Y
krotka	Nazwisko	Imię dziecka	Znajomość języków
t <sub>1</sub>	Kot	Ania	niemiecki
t <sub>2</sub>	Kot	Jaś	angielski
s <sub>1</sub>	Kot	Ania	angielski
s <sub>2</sub>	Kot	Jaś	niemiecki
	Słoń	Ola	niemiecki
	Słoń	Ola	angielski

KSIM, WMiP, AGH

42



## Zależność wielowartościowa

$t_1[X] = t_2[X] = s_1[X] = s_2[X] = \{Kot\}$   
 $s_1[Y] = t_1[Y] = \{Ania\}$  i  
 $s_1[R-X-Y] = t_2[R-X-Y] = \{angielski\}$  i  
 $s_2[Y] = t_2[Y] = \{Iaś\}$  i  
 $s_2[R-X-Y] = t_1[R-X-Y] = \{niemiecki\}$

KSIM, WMiP, AGH

43



## Piąta postać normalna

– Relacja jest w piątej formie normalnej (V PN) wtedy i tylko wtedy, gdy jest w czwartej postaci normalnej i nie istnieje jej rozkład odwracalny na zbiór mniejszych tabel.

KSIM, WMiP, AGH

44



## Klucz sztuczny

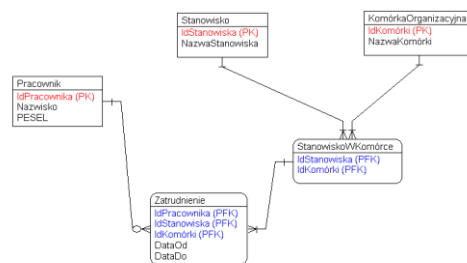
– Klucz stworzony wyłącznie dla potrzeb więzi w celu zastąpienia złożonego klucza głównego

KSIM, WMiP, AGH

45



## Klucz złożony...

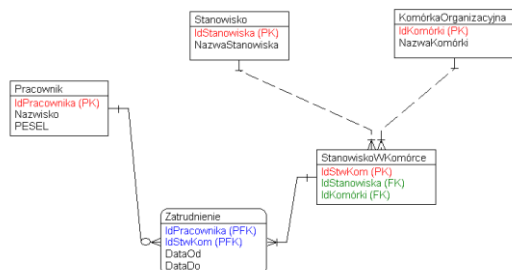


KSIM, WMiP, AGH

46



## ...zastąpiony kluczem sztucznym



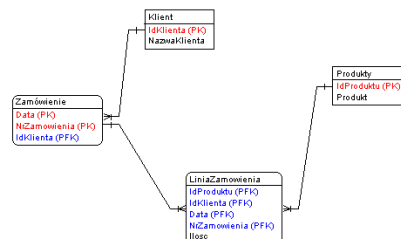
KSIM, WMiP, AGH

47



## Klucz złożony...

[1..1]



KSIM, WMiP, AGH

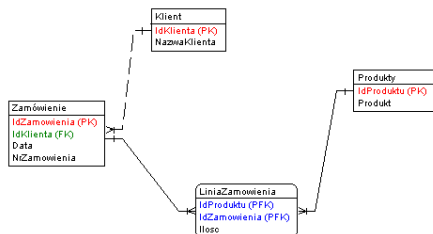
48





...zastąpiony kluczem sztucznym

[1,1]



KSM, WNIPI, AGH

49



Klucz sztuczny

- Klucz sztuczny może być wykorzystany do kodowania atrybutów tekstowych (w niektórych przypadkach także liczbowych) o powtarzających się wartościach, dla których można utworzyć listę
- Użycie klucza sztucznego wymaga stworzenia dodatkowej tabeli (słownika) pozwalającego na „rozkodowanie” klucza

KSM, WNIPI, AGH

50



Nr faktury	Za okres		Identyfikator nabywcy
	od	do	
21113332437	1.11.2007	30.11.2007	1
21218909871	1.12.2007	31.12.2007	1
21113332442	1.11.2007	30.11.2007	2
21218909900	1.12.2007	31.12.2007	2

Nr faktury	Liczba jednostek (czas połączenia)	Id usługi
2111332437	1	1
2111332437	1	2
2111332437	30	3
2111332437	28	4
2111332437	15	5
2111332437	1	6
2111332437	20	7
2111332437	15	8

Id rodzaju usługi	Usługa	Stawka VAT
1	Abonament	22
2	Polecenia krajowe	22
3	SMS	22

Id usługi	Id rodzaju usługi	Strefa czasowa	Kierunek	Cena
1	1			70,00
2	2	Szczyt	Era	2,00
3	2	Szczyt	Plus GSM	1,10
4	2	Poza szczytem	Plus GSM	1,00
5	3			0,50

Strefa czasowa	Kierunek	Cena
Szczyt		
Poza szczytem		
Kierunek		
Plus GSM		

KSM, WNI

51



Nr faktury	Za okres		Identyfikator nabywcy
	od	do	
21113332437	1.11.2007	30.11.2007	1
21218909871	1.12.2007	31.12.2007	1
21113332442	1.11.2007	30.11.2007	2
21218909900	1.12.2007	31.12.2007	2

wartość netto \* stawka

liczba jednostek \* cena

wartość netto + kwota VAT

tego nie trzeba pamiętać

Nr faktury	Usługa	Strefa czasowa	Kierunek	Liczba jednostek (czas połączenia)	Wartość netto	Stawka VAT	Kwota VAT	Wartość brutto
21113332437	Abonament			1	70,00	22,00	15,40	85,40
21113332437	Polecenia krajowe	Szczyt	Era	25,3	50,60	22,00	11,13	61,73
21113332437	Polecenia krajowe	Szczyt	Plus GSM	30	33,00	22,00	7,26	40,26
21113332437	Polecenia krajowe	Poza szczytem	Plus GSM	15	15,00	22,00	3,30	18,30
21113332437	SMS			20	10,00	22,00	2,20	12,20
21113332437	Razem				186,10			186,10
21113332442	Abonament			1	70,00	22,00	15,40	85,40
21113332442	Polecenia krajowe	Szczyt	Era	15	30,00	22,00	6,60	36,60
21113332442	Polecenia krajowe	Szczyt	Plus GSM	28	30,80	22,00	6,78	37,58
21113332442	Polecenia krajowe	Poza szczytem	Plus GSM	12	12,00	22,00	2,64	14,64
21113332442	SMS			15	7,50	22,00	1,65	9,15
21113332442	Razem				186,30		30,77	186,37

KSM, WNIPI, AGH



Nr faktury	Usługa	Strefa czasowa	Kierunek	Liczba jednostek (czas połączenia)	Cena	Stawka VAT
21113332437	Abonament			1	70,00	22
21113332437	Polecenia krajowe	Szczyt	Era	25,3	2,00	22
21113332437	Polecenia krajowe	Szczyt	Plus GSM	30	1,10	22
21113332437	Polecenia krajowe	Poza szczytem	Plus GSM	15	1,00	22
21113332437	SMS			20	0,50	22
21113332442	Abonament			1	70,00	22
21113332442	Polecenia krajowe	Szczyt	Era	15	2,00	22
21113332442	Polecenia krajowe	Szczyt	Plus GSM	28	1,10	22
21113332442	Polecenia krajowe	Poza szczytem	Plus GSM	12	1,00	22
21113332442	SMS			15	0,50	22

to też jest powtarzająca się grupa danych, bo tabela może wyglądać tak:

Nr faktury	Usługa	Strefa czasowa	Kierunek	Liczba jednostek (czas połączenia)	Cena	Stawka VAT
21113332437	Abonament			1	70,00	22
21113332442	Polecenia krajowe	Szczyt	Era	25,3	2,00	22
21113332442	Polecenia krajowe	Szczyt	Plus GSM	30	1,10	22
21113332437	Polecenia krajowe	Poza szczytem	Plus GSM	15	1,00	22
21113332437	SMS			20	0,50	22
21113332442	SMS			15	0,50	22

KSM, WNIPI, AGH

53



Nr faktury	Liczba jednostek (czas połączenia)	Id usługi	Usługa	Strefa czasowa	Kierunek	Cena	Stawka VAT
21113332437	1	1	Abonament			70,00	22
21113332442	1	1					
21113332437	25,3	2	Polecenia krajowe	Szczyt	Era	2,00	22
21113332437	30	3	Polecenia krajowe	Szczyt	Plus GSM	1,10	22
21113332442	28	3					
21113332437	15	4	Polecenia krajowe	Poza szczytem	Plus GSM	1,00	22
21113332442	12	4					
21113332437	20	5					
21113332442	15	5	SMS			0,50	22

brak dobrego kandydata na klucz grupy i dlatego wprowadzamy klucz sztuczny i wykonujemy dekompozycję

Nr faktury	Liczba jednostek (czas połączenia)	Id usługi	Usługa	Strefa czasowa	Kierunek	Cena	Stawka VAT
21113332437	1	1					
21113332442	1	1					
21113332437	25,3	2					
21113332442	15	2					
21113332437	30	3					
21113332442	28	3					
21113332437	15	4					
21113332442	12	4					
21113332437	20	5					
21113332442	15	5					



Id usługi	Usługa	Strefa czasowa	Kierunek	Cena	Stawka VAT
1	Abonament			70,00	22
2	Połączenia krajowe	Szczyt	Era	2,00	22
3	Połączenia krajowe	Szczyt	Plus GSM	1,10	22
4	Połączenia krajowe	Poza szczytem	Plus GSM	1,00	22
5	SMS			0,50	22

to też jest powtarzająca się grupa danych, w której kluczem jest rodzaj usługi i dlatego trzeba tablicę zdekomponować:

Id usługi	Id rodzaju usługi	Strefa czasowa	Kierunek	Cena	Id rodzaju usługi	Usługa	Stawka VAT
1	1			70,00	1	Abonament	22
2	2	Szczyt	Era	2,00	2	Połączenia krajowe	22
3	2	Szczyt	Plus GSM	1,10	3	SMS	22
4	2	Poza szczytem	Plus GSM	1,00			
5	3			0,50			

KSM, WMSIP, AGH

55



Id usługi	Id rodzaju usługi	Strefa czasowa	Kierunek	Cena
1	1			70,00
2	2	Szczyt	Era	2,00
3	2	Szczyt	Plus GSM	1,10
4	2	Poza szczytem	Plus GSM	1,00
5	3			0,50

to nie są powtarzające się grupy danych, ale powtarzające się dane, które warto przechowywać w słownikach:

Id usługi	Id rodzaju usługi	Id Strefy	Id Kierunku	Cena	Id Strefy	Strefa czasowa
1	1			70,00	1	Szczyt
2	2	1	1	2,00	2	Poza szczytem
3	2	1	2	1,10		
4	2	2	2	1,00		
5	3			0,50		

Id Kierunku	Kierunek
1	Era
2	Plus GSM

KSM, WMSIP, AGH

56



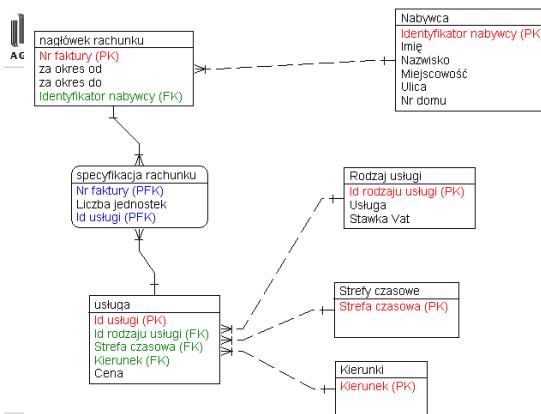
Id usługi	Id rodzaju usługi	Strefa czasowa	Kierunek	Cena	Strefa czasowa
1	1			70,00	Szczyt
2	2	Szczyt	Era	2,00	Poza szczytem
3	2	Szczyt	Plus GSM	1,10	
4	2	Poza szczytem	Plus GSM	1,00	
5	3			0,50	

Kierunek	Kierunek
Era	
Plus GSM	

można nie używać sztucznych kluczy ale należy wówczas zadbać o integralność poprzez zapewnienie kaskadowej aktualizacji:  
**on update cascade on delete cascade**

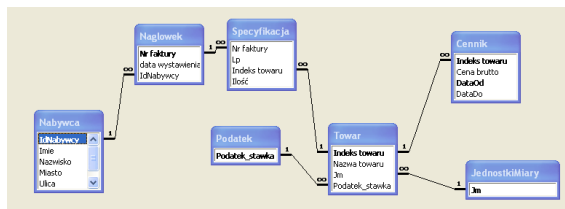
KSM, WMSIP, AGH

57



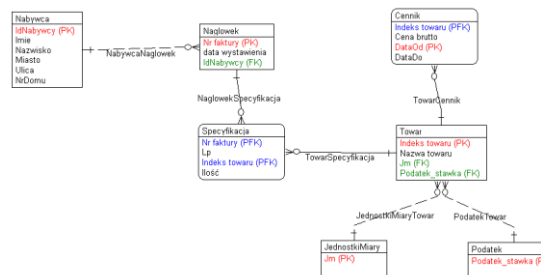
KSM, WMSIP, AGH

58



KSM, WMSIP, AGH

59

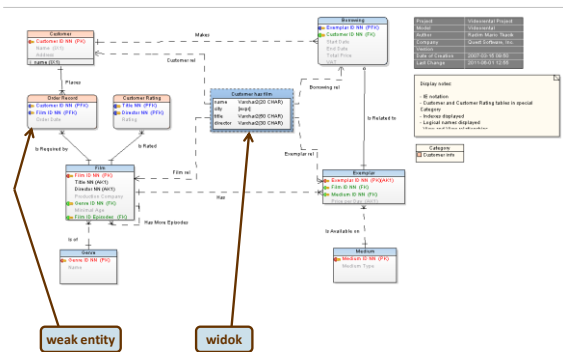


KSM, WMSIP, AGH

60



Toad



KSIM, WIMAP, AGH

61