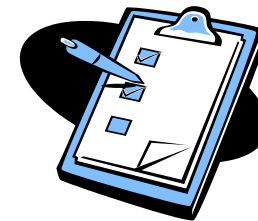


BAZY DANYCH

WYKŁAD NR 2

Uniwersytet Gdański

Agenda



2

Wprowadzenie do normalizacji

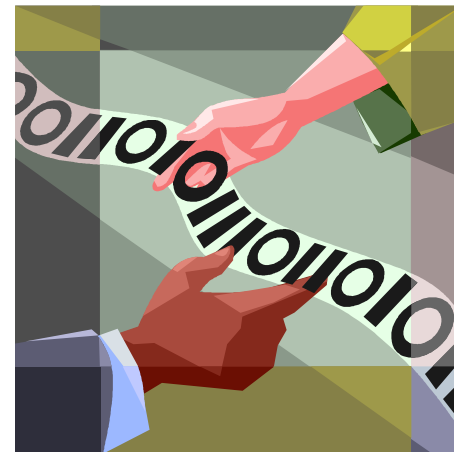
Rodzaje postaci normalnych

Przykłady zależności

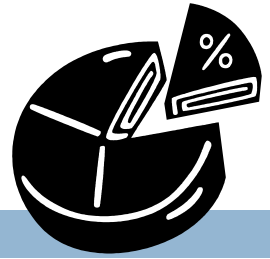
Aspekty teoretyczne baz danych

Normalizacja danych

Założenia, rodzaje, przykłady



Modelowanie danych



4

Podejście zstępujące
(ang. top down)
– diagramy związków encji

Podejście wstępujące
(ang. bottom up)
– normalizacja danych

Czym jest normalizacja?

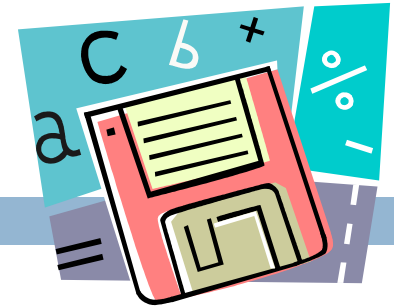


5

- Normalizacja to proces organizowania danych w bazie danych. Polega na:
 - ▣ identyfikacji logicznych związków pomiędzy elementami. Mogą to być związki:
 - funkcyjne
 - niefunkcyjne
 - ▣ eliminacji niespójnych zależności i nadmiarowości

Brak normalizacji

6



Nr indeksu	Imię i nazwisko	Adres	Numer książki	Tytuł i autor książki	Data wypożyczenia i zwrotu
------------	-----------------	-------	---------------	-----------------------	----------------------------

- Problemy:
 - aktualizacja
 - usuwanie
 - wstawianie
 - redundancja (powtórzenia, nadmiarowość)

Postać normalna



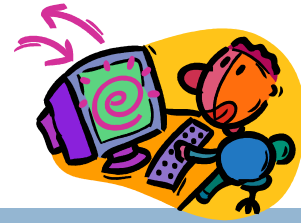
7

Twórcą pierwszej, drugiej i trzeciej postaci normalnej jest E. F. „Ted” Codd.

Form normalnych czwartej, piątej i szóstej nie używa się w praktyce.

Normalizacja prowadzi do zwiększania się liczby tabel co może znacząco zmniejszyć wydajność bazy danych.

Cel normalizowania danych



8

Eliminacja wielokrotnie powtarzanych informacji

Zmniejszenie rozmiarów tabel

Poprawienie wydajności podczas wyszukiwania danych

Umożliwienie wyszukiwania danych według szczegółowych kryteriów
(np. poprzez rozbicie pola adres na miejscowość, ulica, nr domu)

Proces normalizacji



9

Dane nieznormalizowane – ciąg danych

1NF – atrybuty mają być w postaci elementarnej, nierozkładalnej – np. liczba, data, a nie ciąg liczbowy, ciąg dat i liczb)

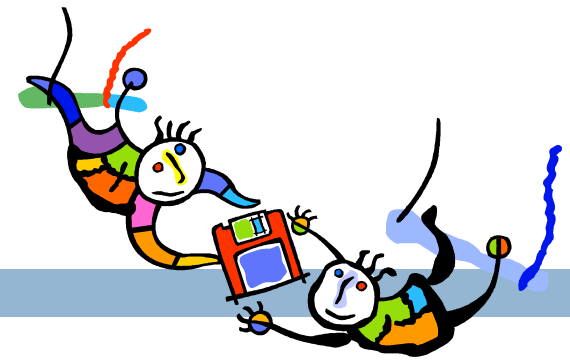
2NF – każdy atrybut zależy funkcyjnie od klucza

3NF – nie istnieją żadne zależności przechodnie

4NF, 5NF oraz 6NF są często jedynie poddawane teoretycznym rozważaniom i tematyka ta nie będzie podejmowana na wykładzie

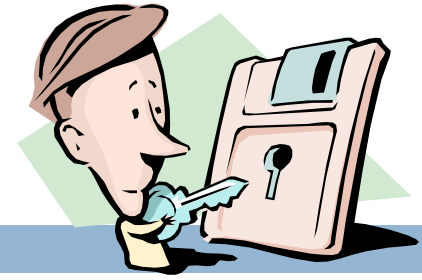
Warunki przeprowadzania normalizacji

10



- Podczas normalizacji nie może dojść do sytuacji w której:
 - ▣ poszczególne atrybuty (kolumny) są usuwane
 - ▣ informacje są tracone
 - ▣ zależności pozostają niezidentyfikowane

Rodzaje kluczy



11

Klucze potencjalne (ang. candidate keys)

- mogą, ale nie muszą zostać wybrane jako klucze główne

Klucze główne (ang. primary keys)

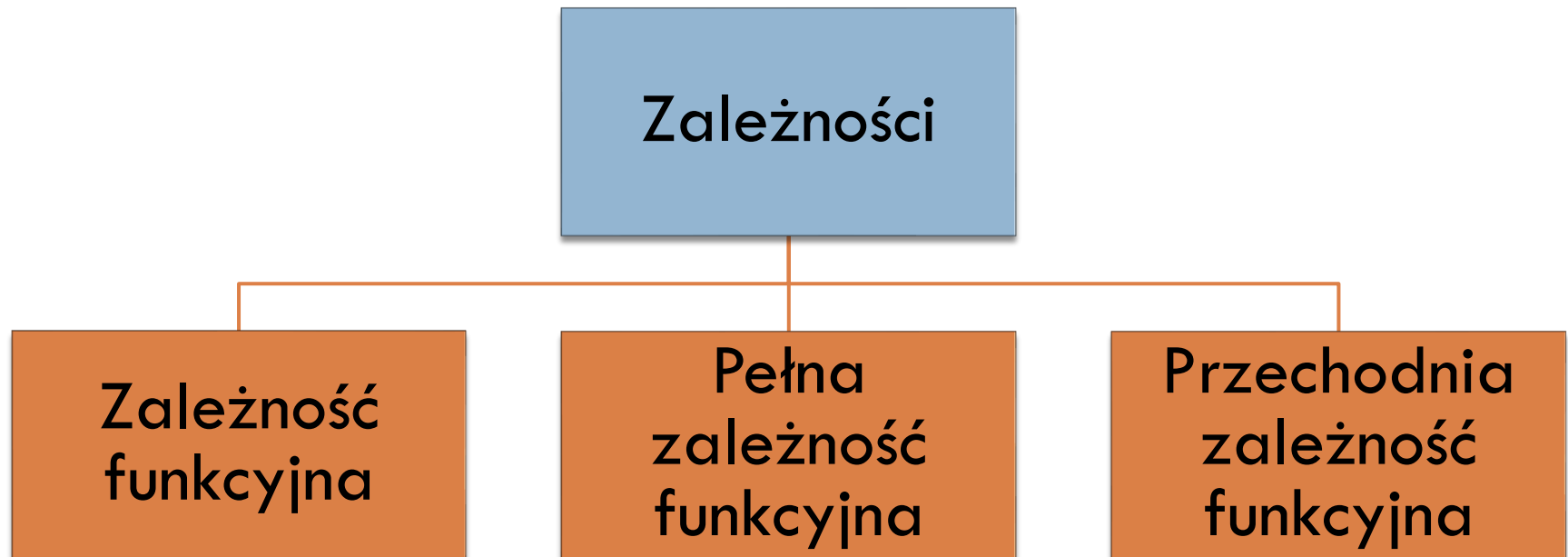
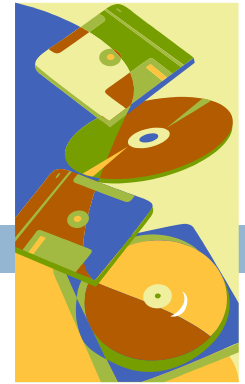
- identyfikują jednoznacznie wiersz, są unikalne, mogą być zbiorem wielu kolumn

Klucze obce (ang. foreign keys)

- odwołują się do kluczy głównych innych tabel

Rodzaje zależności

12

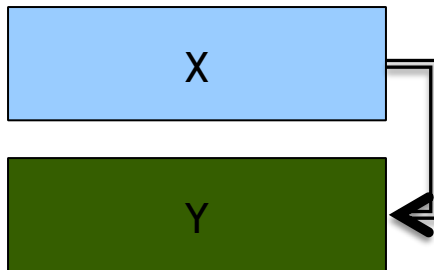


Zależność funkcyjna



13

- Atrybut Y jest **funkcyjnie zależny** od atrybutu X, jeżeli dowolnej wartości atrybutu X odpowiada nie więcej niż jedna wartość atrybutu Y. Wówczas X identyfikuje Y.

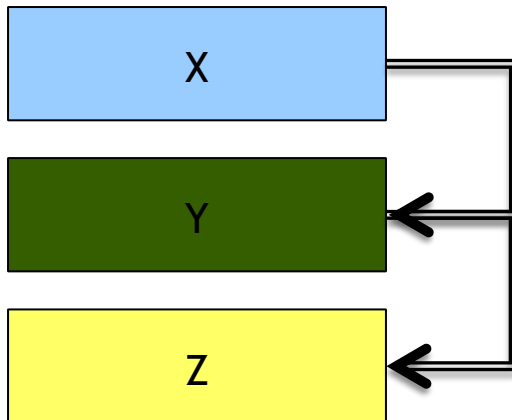


Pełna zależność funkcyjna



14

- Atrybut Y jest w **pełni funkcyjnie zależny** od zbioru atrybutów X, jeżeli jest zależny funkcyjnie i nie jest zależny od żadnego podzbioru właściwego dla zbioru X.



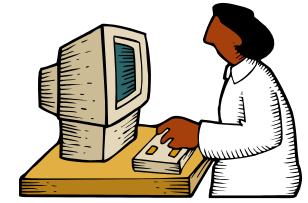
Przechodnia zależność funkcyjna (1 / 2)



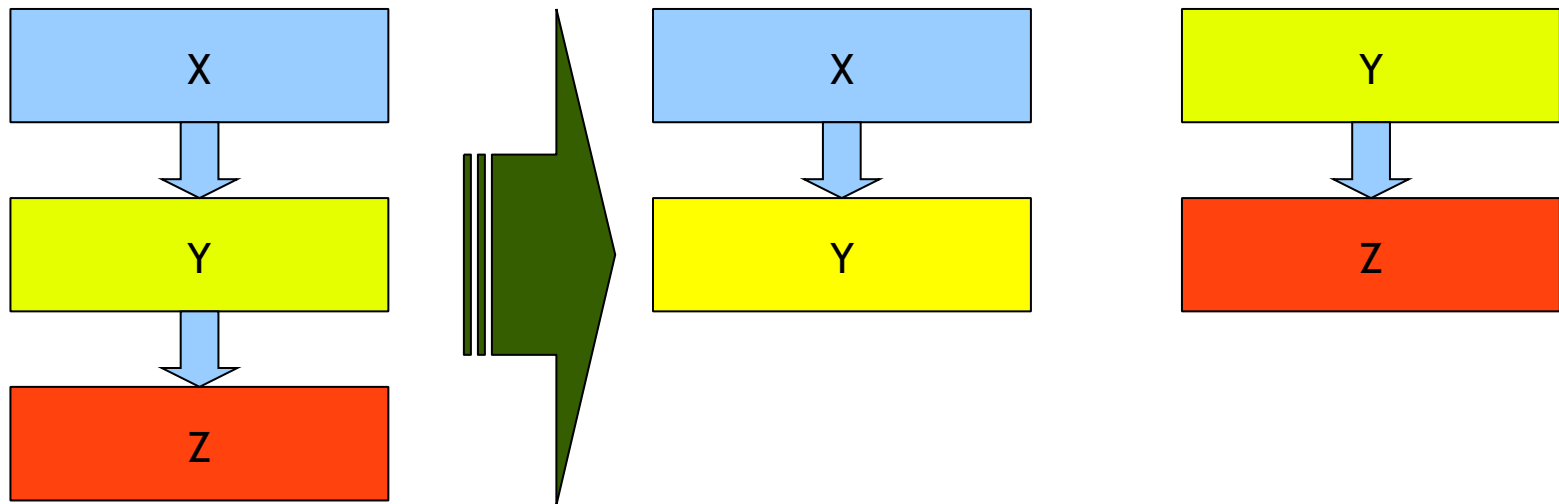
15

- Jeżeli X , Y i Z to rozłączne podzbiory atrybutów danej encji, to podzbiór atrybutów Z jest **przechodnio funkcyjnie zależny** od podzbioru atrybutów X , jeżeli Z funkcyjnie zależy od podzbioru Y , Y jest zależny od X , a X nie jest funkcjonalnie zależny od Y i Y nie jest funkcjonalnie zależny od Z .
- **Zależność przechodnia** występuje, gdy dany atrybut zależy nie tylko od klucza głównego, ale również innego atrybutu.

Przechodnia zależność funkcyjna (2/2)

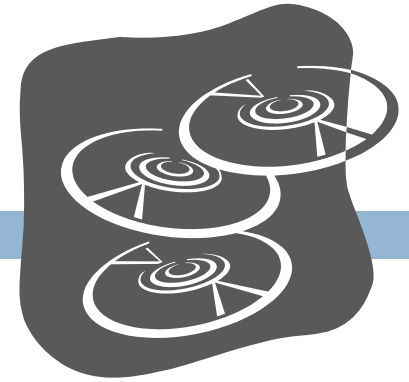


16



Jak rozpoznać postać normalną?

17



Normalizacja	Opis
Brak	Dane w jednym polu tabeli. Dane niejednoznaczne, różne wartości i typy danych.
1NF	Wartości atrybutów muszą być niepodzielne – jedna wartość w każdym polu tabeli.
2NF	Każdy atrybut tej tabeli, który nie należy do klucza potencjalnego, jest w pełni funkcyjnie zależny od wszystkich kluczy potencjalnych.
3NF	Każdy atrybut, który nie wchodzi w skład klucza potencjalnego, nie jest przechodnio funkcyjnie zależny od klucza potencjalnego tej tabeli.

Dane nieznormalizowane



18

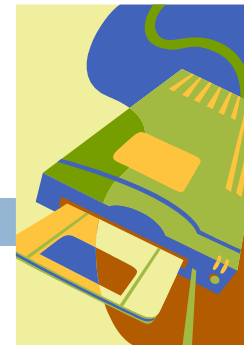
Ciąg wielu
atrybutów

Ciąg wielu
atrybutów

PESZC	Imię i nazwisko	Adres
89123013234	Jan Kowalski	80-400 Gdańsk, ul. Grunwaldzka 729A/3
92011517892	Anna Nowak	81-820 Sopot, al. Niepodległości 981B/8
91081213121	Ewa Malinowska	84-890 Gdynia, ul. Morska 613D/23

1NF – pierwsza postać normalna

19

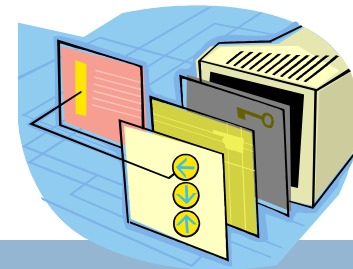


Dane
elementarne

PESEL	Imię	Nazwisko	Kod	Miast o	Ulica	Nr domu
89123013234	Jan	Kowalski	80-400	Gdańsk	ul. Grunwaldzka	729A/3
92011517892	Anna	Nowak	81-820	Sopot	al. Niepodległości	981B/8
91081213121	Ewa	Malinowska	84-890	Gdynia	ul. Morska	613D/23

Inny przykład – dane nieznormalizowane i 1NF

20



nieznormalizowane

Numer indeksu	Wykłady
178283	Technologie baz danych, Hurtownie danych
184382	Technologie baz danych, Systemy baz danych

1NF

Numer indeksu	Wykłady
178283	Technologie baz danych
178283	Hurtownie danych
184382	Technologie baz danych
184382	Systemy baz danych

2NF – druga postać normalna



21

- Mogą występować zależności przechodnie.
- Każdy atrybut zależny funkcyjnie od klucza.
- W poniższym przykładzie prowadzący laboratorium jest zależny przede wszystkim od numeru grupy. Dlatego nie jest to 3NF.

<u>Numer studenta</u>	Nazwisko studenta	Imię studenta	Numer grupy	Laboratorium BD
182382	Kowalski	Jan	351	Adam Nowak
192812	Zielińska	Ewa	341	Jan Woźniak

3NF – trzecia postać normalna

22



<u>Numer studenta</u>	Nazwisko studenta	Imię studenta	Numer grupy (klucz obcy)
182382	Kowalski	Jan	351
192812	Zielińska	Ewa	341

<u>Numer grupy</u>	Laboratorium BD
351	Adam Nowak
341	Jan Woźniak

Zadanie

23

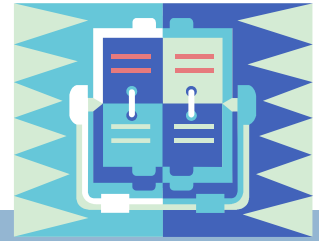


Numer faktury	REGON	Nazwa klienta	Adres klienta	Kod produktu	Nazwa produktu	Ilość	Jednostka miary	Cena	PKWiU	VAT	Cena łączna
10	00000123450000	ABC	Gdańsk, Grunwaldzka	5	Ołówek	23	sztuka	1 zł	1389	7%	23 zł
23	00000123430000	BCD	Gdynia, Morska	2	Pióro	54	sztuka	7 zł	1339	22%	378 zł
54	00000123540000	CDE	Sopot, Niepodległości	3	Długopis	290	sztuka	4 zł	1339	22%	1160 zł

Teoretyczne aspekty baz danych



Podstawowe pojęcia



25

Baza danych jest logicznie spójnym zasobem przechowywanych i udostępnianych danych, w aspekcie określonego wycinka rzeczywistości.

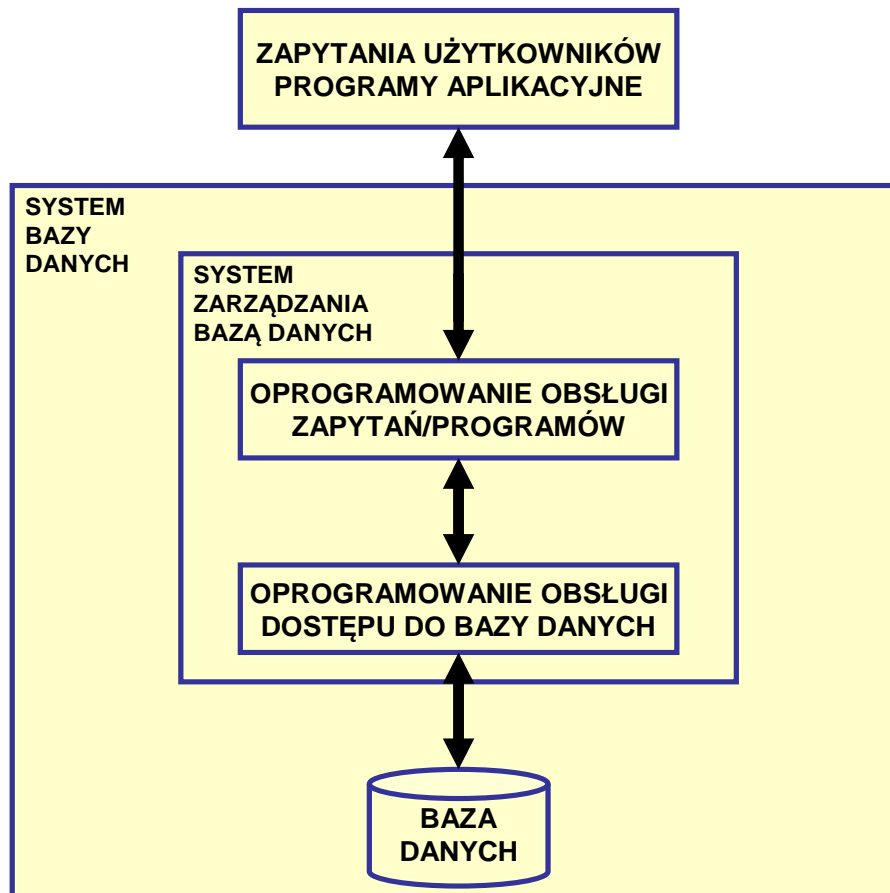
System zarządzania bazą danych to system oprogramowania, który umożliwia tworzenie i użytkowanie bazy danych.

System baz danych stanowi powiązanie bazy danych z obsługującym go systemem zarządzania bazą danych.

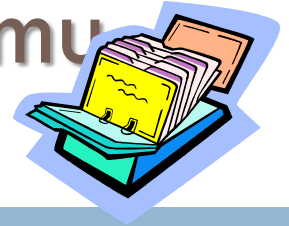
System baz danych



26



Trójwarstwowa architektura systemu bazy danych



27

POZIOM ZEWNĘTRZNY

Interfejs:

schemat zewnętrzny \Leftrightarrow konceptualny

POZIOM KONCEPTUALNY

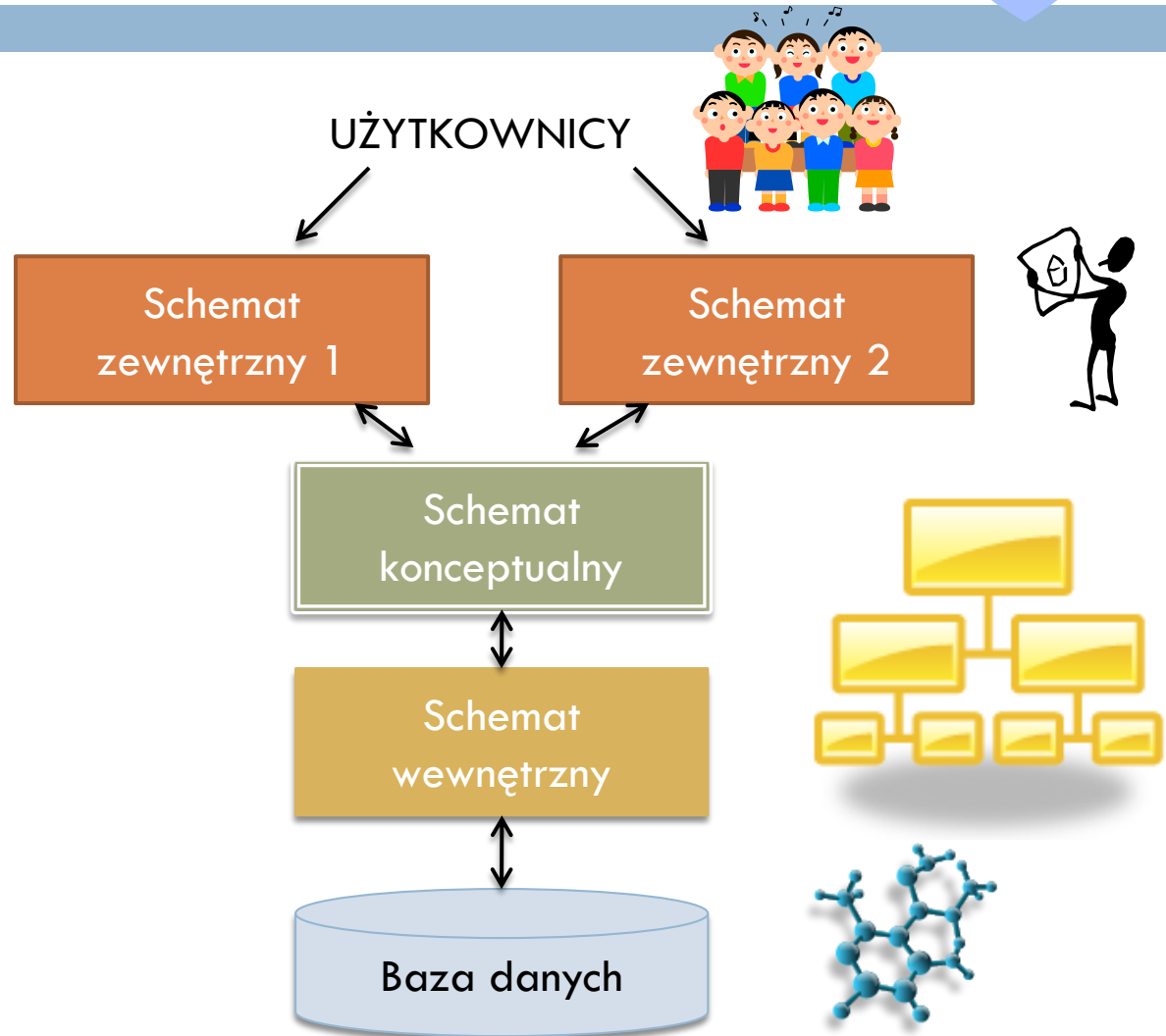
Interfejs:

schemat konceptualny \Leftrightarrow wewnętrzny

POZIOM WEWNĘTRZNY

Interfejs:

schemat wewnętrzny \Leftrightarrow baza danych



Podstawowe role użytkowników bazy danych



28

Administrator bazy danych

- Autoryzacja dostępu, koordynacja użytkowania, optymalizacja

Projektant bazy danych

- Opracowuje modele bazodanowe

Programista bazy danych

- Tworzy skrypty i oprogramowanie dla bazy danych

Użytkownicy

- Przypadkowi (niesystematycznie korzystający np. za pomocą SQL)
- Parametryczni (transakcje bazodanowe za pomocą programów przygotowanych przez programistów)
- Zaawansowani (całe spektrum SZBD)

Historia kolejnych generacji baz danych



29

pierwsza (lata pięćdziesiąte) – pliki o dostępie sekwencyjnym



druga (lata sześćdziesiąte) – pliki o dostępie bezpośrednim na dyskach magnetycznych



trzecia (lata siedemdziesiąte) – prerelacyjna



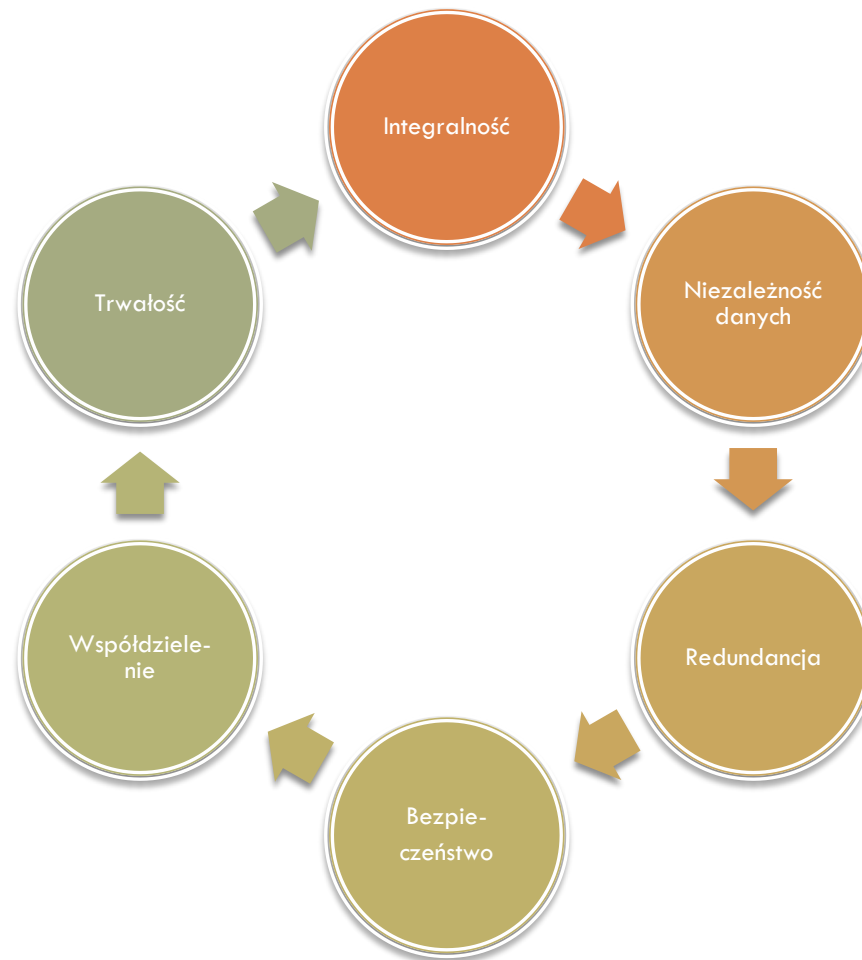
czwarta (lata osiemdziesiąte) – relacyjna



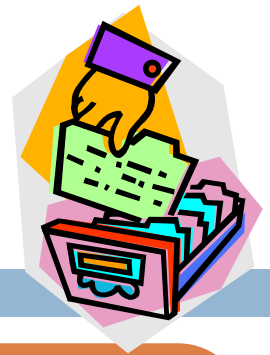
piąta (lata dziewięćdziesiąte) – postrelacyjna

Cechy baz danych

30

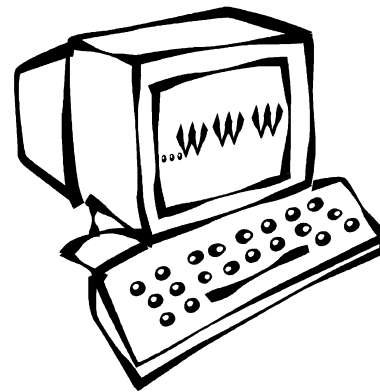
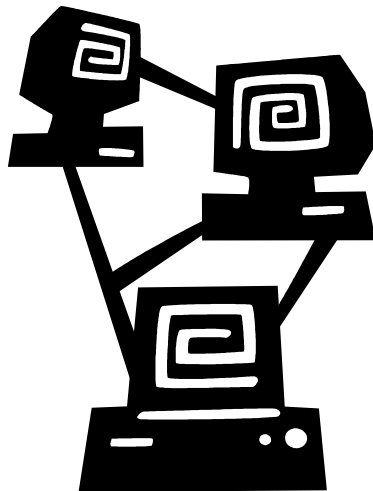


Rozproszona baza danych

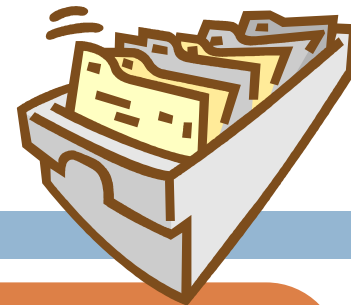


31

Rozproszona baza danych jest zasobem danych, które **logicznie** należą do tego samego systemu, natomiast **fizycznie** są przechowywane w różnych węzłach sieci komputerowej.



Zasada działania



32

System rozproszonych baz danych rezyduje na wielu komputerach zwanych **węzłami** oraz jest połączony **siecią komunikacyjną** w celu przesyłania danych między nimi.

Pewną odmianą rozproszonych baz danych są systemy **wielobazowe**. Każdy węzeł takiej bazy jest niezależną i autonomiczną bazą danych z własnym SZBD.

Aspekty rozproszonych baz danych



33



Pytanie?

34

- Sytuacja, w której każdy atrybut, który nie wchodzi w skład klucza potencjalnego, nie jest przechodnio funkcyjnie zależny od klucza potencjalnego tej tabeli, opisuję postać normalną:
 - ▣ 0NF
 - ▣ 1NF
 - ▣ 2NF
 - ▣ 3NF

