



Bazy danych

Program wykładu

- ❑ Wprowadzenie
- ❑ Podstawowe pojęcia
- ❑ Systemy zarządzania bazami danych
- ❑ Cele bazy danych
- ❑ Modele baz danych
- ❑ Relacyjny model baz danych
- ❑ Projektowanie baz danych
- ❑ Hurtownie danych
- ❑ Bazy danych – czynniki sukcesu

Wprowadzenie

Stare założenia:

- ❑ Każdy oddział najlepiej zna swoje potrzeby. Dane muszą być przechowywane w specjalny sposób, odzwierciedlający potrzeby poszczególnych oddziałów.
- ❑ Otrzymanie informacji wymaga czasu. Każdy oddział posiada swoje własne dane i podejmuje własne decyzje; jeśli potrzebujesz informacji z danego wydziału - poproś o nie.
- ❑ **Tak było:** systemy zorientowane na aplikacje: każdy oddział instytucji miał własne aplikacje obsługujące jego dane.
- ❑ **Konsekwencje:** nadmiarowość (redundancja) i trudności w zdobyciu przekrojowych informacji.

Wprowadzenie

Zbiór kredytów budowlanych

Nr rachunku	Nazwisko	Imię	Kod	Miasto	Ulica	Dane kredytowe
-------------	----------	------	-----	--------	-------	-------	----------------

Zbiór pożyczek samochodowych

Nr rachunku	Nazwisko	Imię	Kod	Miasto	Ulica	Dane o pożyczce
-------------	----------	------	-----	--------	-------	-------	-----------------

Zbiór rachunków rozliczeniowych

Nr rachunku	Nazwisko	Imię	Kod	Miasto	Ulica	Dane o rachunku
-------------	----------	------	-----	--------	-------	-------	-----------------

Wprowadzenie

Nowe założenia:

- ❑ Dane są wspólnymi zasobami; są własnością organizacji - nie poszczególnych oddziałów.
- ❑ Dowolna informacja jest dostępna przez sieć dla wszystkich innych oddziałów.

Nie chodzi tutaj o sprawy czysto techniczne, lecz raczej o nowy sposób widzenia organizacji.

BD – podstawowe pojęcia

- ❑ **Baza danych** - zintegrowana grupa ogólnie dostępnych zbiorów danych. Ogólną funkcją bazy danych jest przechowywanie danych związanych z codzienną działalnością organizacji.
- ❑ „**Baza danych** jest to taka organizacja zintegrowanych zbiorów danych z pewnej dziedziny informacji, która pozwala na zaspokojenie potrzeb jednego lub wielu użytkowników bez uprzedniego sortowania w różne pożądane struktury potrzebne do przetwarzania lub bezpośredniego udzielania informacji. Dane są tak zorganizowane, że korzystanie z nich jest niezależne od oprogramowania jak i od sprzętu”

Źródło: Kisielnicki J., Sroka H., Systemy informacyjne biznesu. Informatyka dla zarządzania, Placet, Warszawa 2005

BD – podstawowe pojęcia

Baza danych

Zintegrowana grupa powiązanych zbiorów



Zbiór

Grupa powiązanych rekordów

Nazwisko	Imię	Miejscowość	Nr telefonu	Nr Pesel
Kowalska	Maria	Warszawa	504-555-123	56070703222
Nowak	Adam	Kraków	606-555-678	73120405333

Rekord

Nowak	Adam	Kraków	606-555-678	73120405333
-------	------	--------	-------------	-------------

Pole

Nowak (nazwisko)
Adam (imię)
Kraków (miejscowość)
606-555-678 (nr telefonu)
73120405333 (nr Pesel)

Bajt



Bit

0 lub 1

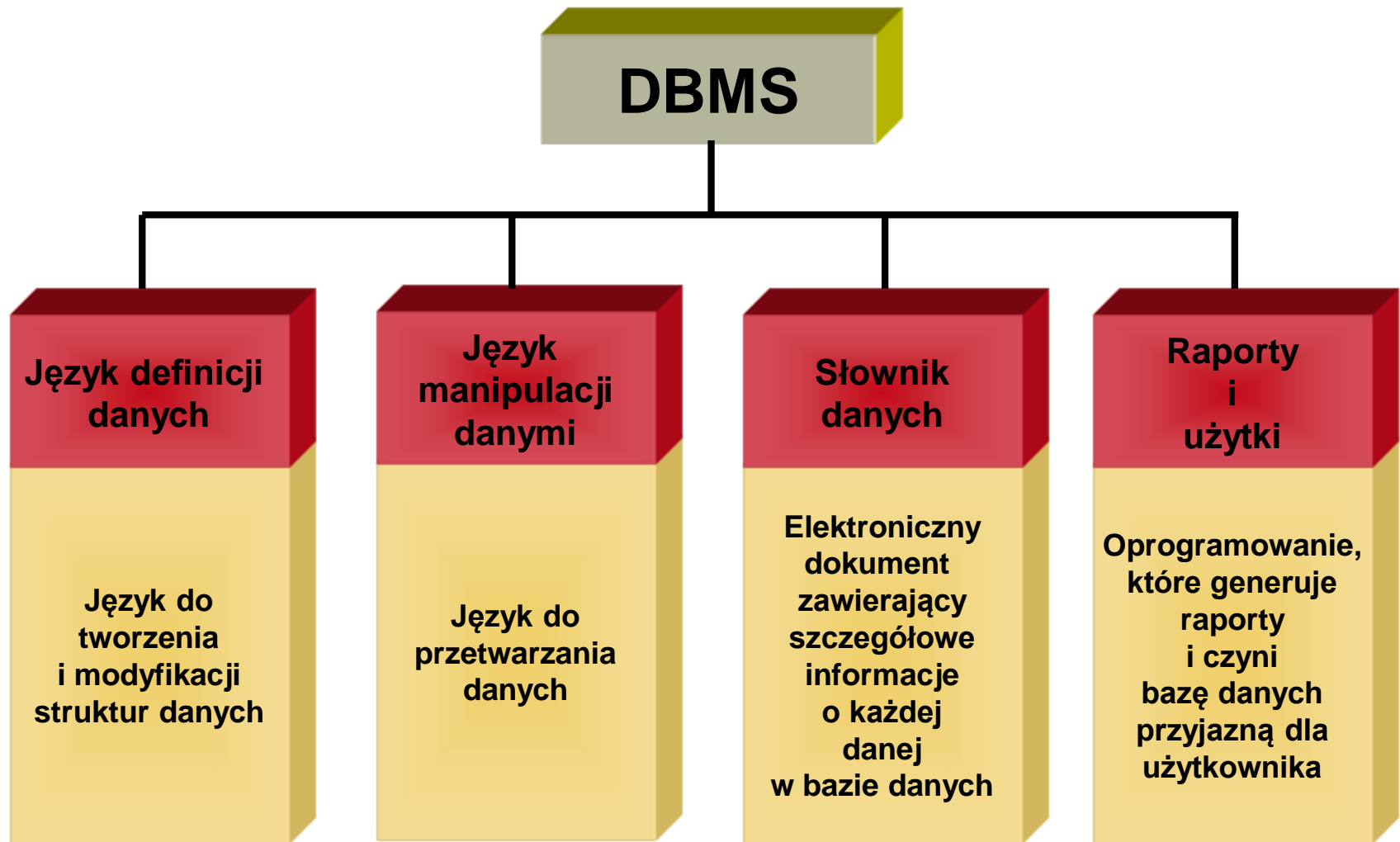
BD – podstawowe pojęcia

- **SZBD (Systemy Zarządzania Bazami Danych)**
DBMS (Database Management Systems)

to zestaw narzędzi umożliwiających dostęp i zarządzanie bazą danych. Podstawowymi funkcjami tych systemów są:

- zarządzanie plikami,
- wyszukiwanie informacji,
- zarządzanie bazą danych.

BD – podstawowe pojęcia



BD – podstawowe pojęcia

- ❑ **Język definicji danych
(DDL - Data Definition Language)**

Służy do tworzenia i aktualizacji struktury danych.

- ❑ **Język manipulacji danymi
(DML - Data Manipulation Language)**

Służy do przetwarzania i aktualizacji danych.

- ❑ **Strukturalny język zapytań
(SQL – Structured Query Language)**

Język (interpreter) do komunikacji z serwerem baz danych; odpowiada za: integralność danych, manipulację danymi, dostęp do danych, wyprowadzanie danych i bezpieczeństwo danych.

BD – podstawowe pojęcia

Przykładowe zapytania:

- ❑ `CREATE TABLE pracownicy (imie varchar(50), nazwisko varchar(50), placa float, staz int);`
Tworzy tabelę „pracownicy” zawierającą pola tekstowe zmiennej długości (varchar) o nazwach „imie” (imię) i „nazwisko”, o maksymalnej długości 50 znaków, pole „placa” zapisane za pomocą liczby rzeczywistej (float od ang. floating point) oraz „staz” zapisany za pomocą liczby całkowitej (int od ang. integer).
- ❑ `DROP TABLE pracownicy;`
Usuwa z bazy tabelę „pracownicy”.
- ❑ `ALTER TABLE pracownicy ADD dzial varchar(255);`
Dodaje do struktury tabeli „pracownicy” kolumnę „dzial” (dział), jako pole tekstowe o długości maks. 255 znaków.

Źródło: <http://pl.wikipedia.org/wiki/SQL>

BD – podstawowe pojęcia

Przykładowe zapytania:

- ❑ `SELECT * FROM pracownicy WHERE placa > 2000 ORDER BY staz DESC;`
Zwraca tabelę (listę) utworzoną ze wszystkich kolumn (*) tabeli „pracownicy” (FROM pracownicy) zawierającą pracowników, których placa jest większa niż 2000 (WHERE placa > 2000) i sortuje wynik malejąco według parametru staz (ORDER BY staz DESC).
- ❑ `INSERT INTO pracownicy (imie, nazwisko, placa, staz) VALUES ('Jan', 'Kowalski', 5500, 1);`
Dodaje do tabeli „pracownicy” (INTO pracownicy) wiersz (rekord) zawierający dane pojedynczego pracownika.
- ❑ `UPDATE pracownicy SET placa = placa * 1.1 WHERE staz > 2;`
Podwyższa o 10% placa (SET placa = placa * 1.1) pracownikom, których staż jest większy niż 2 (np. lata).
- ❑ `DELETE FROM pracownicy WHERE imie = 'Jan' AND nazwisko = 'Kowalski';`
Usuwa z tabeli „pracownicy” wszystkie wiersze (rekordy) dotyczące pracownika o imieniu „Jan” i nazwisku „Kowalski”

BD – podstawowe pojęcia

- ❑ **Słownik danych** zawiera szczegółowy opis wszystkich danych w bazie:
 - Lokalizacja danych (w jakim zbiorze)
 - Typ danych (liczba, znak, audio, etc.)
 - Rozmiar danych (ile bajtów)
 - Dopuszczalny zakres wartości dla każdego pola
 - Źródło danych (gdzie dana powstaje)
 - Wykorzystanie (kto używa danych)
 - Własność (kto ma prawa do modyfikacji danych)
 - Metody dostępu i ochrony danych

Cele bazy danych (po co tworzyć)

□ **Cele systemów z bazą danych:**

- elastyczny dostęp do danych,
- integralność danych,
- bezpieczeństwo danych,
- obniżenie redundancji,
- niezależność danych od aplikacji,
- współdzielenie danych,
- standaryzacja opisu.

Cele bazy danych

❑ **Łatwe selekcjonowanie i prezentacja danych**

Właściwie chodzi tu o komunikację z użytkownikiem (czyli interfejs). W SZBD rolę tę spełniają SQL oraz generatory raportów.

- Pokaż wszystkich klientów mieszkających przy ul. Akacjowej
- Podnieś wynagrodzenie o 15% wszystkim specjalistom pracującym w firmie ponad 5 lat

Cele bazy danych

Integralność danych - dane są prawidłowe, spójne i aktualne

- ❑ Informacja powinna być wprowadzana natychmiast po powstaniu.
- ❑ Wprowadzanie danych powinno być kontrolowane „w locie” (np. student nie może mieć oceny cząstkowej ani średniej większej niż 6.0)

Cele bazy danych

Bezpieczeństwo danych - ochrona danych przed nieupoważnionym dostępem i przypadkowym uszkodzeniem

- ❑ Hasła i prawa dostępu na różnym poziomie, dobrze zabezpieczone miejsce (łatwiejsze do przeprowadzenia, gdy wszystko jest w jednym miejscu)
- ❑ Odporność na działanie czynników zewnętrznych (kopie bezpieczeństwa)

Cele bazy danych

Obniżenie redundancji – dana powinna być przechowywana tylko w jednym miejscu w bazie

- ❑ Dzięki temu zmiany mogą być dokonywane raz, istnienie niespójnych danych jest niemożliwe, a przestrzeń na dysku jest wykorzystana optymalnie.
- ❑ Jeśli dana jest przechowywana w jednym miejscu, łatwiejsze staje się ich wzajemne powiązanie, przez co zyskujemy silne narzędzie do wspomagania decyzji.

Cele bazy danych

Niezależność danych i aplikacji - fizyczna i logiczna organizacja danych jest oddzielona od aplikacji

- ❑ Baza danych jest jednym z głównych zasobów instytucji; więc nie do przyjęcia jest, by aplikacja jednego użytkownika wpływała na sposób widzenia danych przez innych użytkowników.

Dawniej: COBOL czy Pascal.

Cele bazy danych

Współdzielenie danych - umożliwienie różnym użytkownikom korzystania z tych samych (nie nadmiarowych) danych

- ❑ Jeśli decydujemy się na obniżenie redundancji do minimum, jedynym wyjściem jest współdzielenie danych z innymi użytkownikami. Nie zawsze jest to mile widziane przez użytkowników.

Cele bazy danych

Standaryzacja opisu - jednolite definicje danych dotyczące ich nazw i formatu

- ❑ Wszystkie SZBD dostarczają język opisu danych umożliwiający stworzenie tzw. **słownika danych**. SD to uporządkowany wykaz wszystkich elementów danych mających związek z systemem, wraz z ich precyzyjnym określeniem, aby użytkownik i analityk jednakowo rozumieli wszystkie dane.

Cele bazy danych

Czyżby bez wad?

- ❑ kosztowne na początku - nawet kilkaset tysięcy USD za SZBD,
- ❑ duże wymagania pamięciowe,
- ❑ bardzo skomplikowane dla przeciętnego użytkownika (szczególnie SQL),
- ❑ trudności w zrozumieniu i tworzeniu rozbudowanych relacji (w modelu relacyjnym).

Architektura baz danych

Ze względu na **sposób dostępu** do danych oraz ich przetwarzania wyróżniamy:

- systemy scentralizowane,
- systemy zdecentralizowane,
- systemy rozproszone,
- systemy klient/serwer.

Modele baz danych

Ze względu na **logiczne powiązania** danych wyróżniamy:

- model hierarchiczny,
- model sieciowy,
- model relacyjny,
- model obiektowy.

Modele baz danych

- **Model hierarchiczny**
 - Przypomina schemat organizacyjny
 - Rekord może mieć tylko jednego rodzica
 - Idealnie reprezentuje zależności jeden-do-wielu
- **Model sieciowy**
 - Reprezentuje zależności wiele-do-wielu
 - Rekord może mieć wielu rodziców

Modele baz danych

- **Model relacyjny**
 - najpopularniejszy,
 - oparty na relacjach (powiązanych tabelach).
- Relacja jest tabelą spełniającą warunki:
 - każda relacja w bazie ma jednoznaczną nazwę,
 - każda kolumna w relacji ma jednoznaczną nazwę,
 - wszystkie wartości w kolumnie są tego samego typu,
 - porządek kolumn w relacji nie jest istotny,
 - każdy wiersz w relacji musi być różny,
 - porządek wierszy nie jest istotny,
 - każde pole leżące na przecięciu kolumny i wiersza powinno mieć wartość atomową.

Modele baz danych

- ❑ W latach 60 i 70 dominował model hierarchiczny (IMS).
- ❑ W roku 1970 naukowiec z firmy IBM E. Codd opublikował pracę ***A relational model for large shared data banks***, w której po raz pierwszy przedstawił założenia modelu relacyjnego.
- ❑ W 1983 IBM przedstawił DB2 - pierwszą relacyjną bazę danych dla dużych komputerów.
- ❑ W tym czasie istniały już relacyjne bazy danych dla komputerów osobistych dBaseII (Ashton Tate).

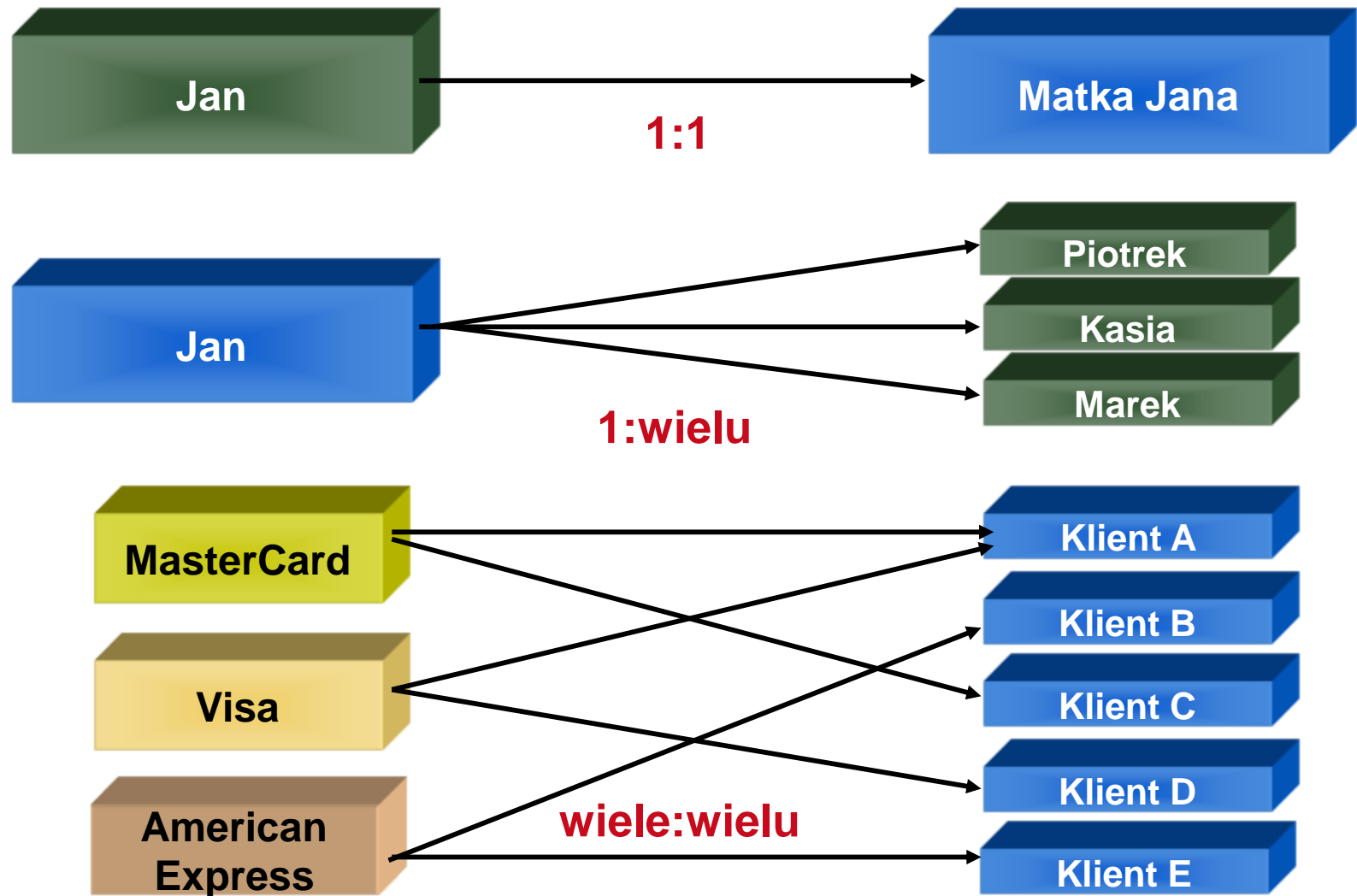
Relacyjny model bazy danych

- ❑ W modelu relacyjnym **zbiór** nazywamy **tabelą**, **rekord** - **wierszem** (lub **krotką**), **kolumnę** - **atrybutem**, a **pole** - **elementem danych**. Całość przypomina arkusz kalkulacyjny.
- ❑ **Relacjami** w relacyjnej bazie danych nazywamy wszystkie możliwe rodzaje powiązań między rekordami (wierszami). Z matematycznego punktu widzenia są one tabelą równą iloczynowi kartezyjańskiemu zbiorów.

Relacyjny model bazy danych

- **Relacja** to logiczne połączenie tabel wg zadanego klucza (klucz własny i obcy). Efektem utworzenia relacji jest natychmiastowy dostęp do danych zawartych w połączonych tabelach; innymi słowy: przechodząc do innego rekordu w tabeli A mamy natychmiastowy dostęp do powiązanych z nim danych w tabeli B.
- Rozważmy relację, której atrybutami są nazwisko, imię, wiek. Relację tę można zapisać następująco:
PRAC <nazwisko, imię, wiek>, gdzie PRAC jest nazwą danej relacji.
- A oto trzy krotki relacji PRAC:
 - <Kowalski, Jan, 36>
 - <Tomaszewski, Wojciech, 40>
 - <Wiśniewski, Marek, 50>.

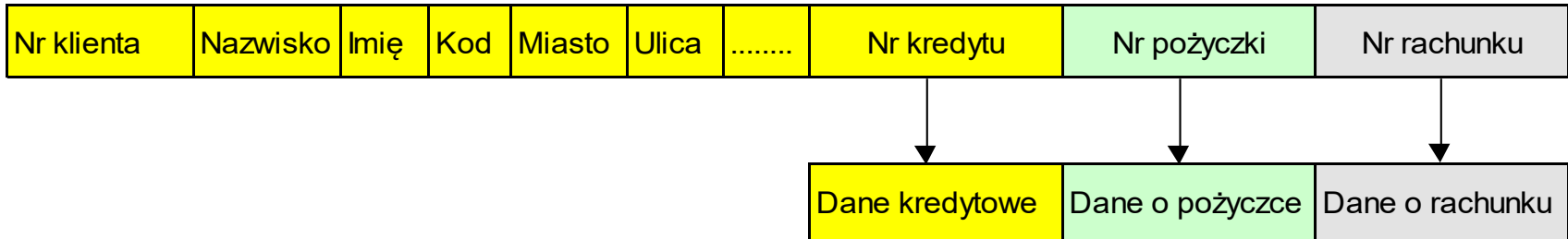
Relacyjny model bazy danych



Relacyjny model bazy danych

Ograniczenie redundancji

Zbiór klientów



Zbiór klientów

Nr klienta	Nazwisko	Imię	Kod	Miasto	Ulica
------------	----------	------	-----	--------	-------	-------

Zbiór kredytów budowlanych

Nr klienta	Nr kredytu	Dane kredytowe
------------	------------	----------------

Zbiór pożyczek samochodowych

Nr klienta	Nr pożyczki	Dane o pożyczce
------------	-------------	-----------------

Zbiór rachunków rozliczeniowych

Nr klienta	Nr rachunku	Dane o rachunku
------------	-------------	-----------------

Relacyjny model bazy danych

- ❑ Mimo że model relacyjny jest koncepcyjnie prosty, nie jest łatwo prawidłowo zaprojektować tabele.
- ❑ W celu wyznaczenia zależności występujących w danej bazie danych powinno się dokładnie rozważyć znaczenie atrybutów. Nie istnieją bowiem metody pozwalające na automatyczne określenie zależności. Wynikają one z wiedzy i doświadczenia projektanta.
- ❑ **Normalizacja** umożliwia pogrupowanie elementów w tabele, w celu zapewnienia spójności, obniżenia redundancji, minimalizacji błędów przy wprowadzaniu danych oraz zajętości pamięci.

Relacyjny model bazy danych - normalizacja



Klienci : Tabela						
	Nr klienta	Nazwa klienta	Adres	Prowadzący	Tel prowadzącego	Rejon
	12345	ABC sp. z o.o.	ul. Spokojna 5	Kowalski	234-567	Południe
	21342	Kolonel	ul. Koniewa 12/4	Nowak	123-654	Północ
	32765	Kaskader s.c.	ul. Markotna 2	Sendorek	567-547	Północ
	37651	PPHU Boryna	al. Mickiewicza 12	Nowak	122-645	Północ
	98342	Na Fali S.A.	Os. Młodych 34	Kowalski	234-567	Południe
▶						



Projektowanie bazy danych

- ❑ Analiza potrzeb informacyjnych
- ❑ Projektowanie logicznej struktury bazy
- ❑ Projektowanie fizycznej struktury bazy

Projektowanie bazy danych

Analiza potrzeb informacyjnych

- ❑ **Główne czynności:** identyfikacja potrzeb użytkowników
- ❑ **Cel:** określenie zjawisk, dostępności i użyteczności danych, ich formatu i sposobów obliczeń.

Projektowanie bazy danych

Projektowanie logicznej struktury bazy

- **Główne czynności 1:** określenie rekordów i zdefiniowanie atrybutów opisujących rekordy
- **Cel 1:**
 - przyporządkowanie rekordów do zjawisk
 - standaryzacja nazw i formatów
 - identyfikacja źródeł danych
- **Główne czynności 2:** określenie relacji między rekordami
- **Cel 2:** identyfikacja typu relacji (1-1, 1-M, N-M)
- **Główne czynności 3:** normalizacja relacji
- **Cel 3:** obniżenie redundancji i wyeliminowanie anomalii (usuwania, wstawiania i aktualizacji)

Projektowanie bazy danych

Projektowanie fizycznej struktury bazy

- ❑ **Główne czynności:** wdrożenie systemu z bazą danych
- ❑ **Cel:** nałożenie struktury logicznej na fizyczne urządzenia

Hurtownie danych

- ❑ Niejednokrotnie im firma jest lepiej z informatyzowana - tym analizowanie informacji może być bardziej skomplikowane (ze względu na ogrom przechowywanych danych).
- ❑ Sytuacja jest jeszcze gorsza kiedy informacje przechowywane są w wielu różnych systemach (system obsługi sprzedaży, system F-K, system kadrowo-płacowy itd.). W takim przypadku dodatkowo zachodzi potrzeba odpowiedniego konwertowania danych i dopasowywania ich do siebie.

Hurtownie danych

- **Hurtownia danych**

Baza danych będąca kolekcją mniejszych baz, zawierająca użyteczne dane zebrane w celu wspomaganie procesu decyzyjnego.

- Hurtownia danych gromadzi dane ze wszystkich systemów funkcjonujących w obrębie organizacji, stanowi więc źródło analiz dotyczących całej działalności. Dane zorganizowane są pod kątem dostarczania w jak najkrótszym czasie rzetelnych i spójnych informacji o aktualnym stanie organizacji.

Hurtownie danych

- Systemy te przeznaczone są dla firm borykających się z co najmniej jednym problemem typu:
 - duża ilość informacji magazynowana w systemie komputerowych - wydajność systemu nie pozwala na prowadzenie analiz,
 - w firmie współdziała wiele systemów komputerowych pochodzących od różnych producentów, systemy te nie są dobrze zintegrowane,
 - wyciąganie wniosków na podstawie zawartości baz danych jest utrudnione - danych jest tak dużo, że można je analizować jedynie ze statystycznego punktu widzenia.

Hurtownie danych a bazy danych

- ❑ W większości przypadków hurtownie danych są znacznie większe niż bazy danych.
- ❑ Bazy danych są aktualizowane bardzo często (on-line), hurtownie danych – nie.
- ❑ Zarówno bazy jak i hurtownie wspomagają szybkie tworzenie on-line zapytań i raportów dla menadżerów.
- ❑ Obsługa hurtowni jest dużo łatwiejsza niż bazy: nie jest wymagane instalowanie specjalnego oprogramowania na komputerach użytkowników systemu, wystarczy arkusz lub przeglądarka internetowa.
- ❑ Bazy danych są zwykle zorganizowane wokół jednostek organizacyjnych (np. kadry) lub wokół funkcji (np. marketing). Hurtownie danych zawierają obraz całej organizacji.

BD – czynniki sukcesu

- **Używaj bazy danych do usprawnienia podejmowania decyzji**
 - Wiele organizacji kolekcjonuje ogrom informacji, których nie jest w stanie wykorzystać
 - Staranna analiza danych oraz podejmowanie odpowiednich decyzji są podstawą sukcesu organizacji i zadowolenia klientów
- **Weź pod uwagę, że bazy danych są narzędziem walki konkurencyjnej**
 - Wiele organizacji przeszukuje publiczne bazy danych w celu wydobycia wartościowej informacji dotyczącej konkurencji

BD – czynniki sukcesu

- **Baza danych służy użytkownikowi a nie projektantowi**
 - Informacje zawarte w bazie muszą być zrozumiałe i dostępne dla wszystkich użytkowników uczestniczących w podejmowaniu decyzji
- **Pokaż menadżerom jak ich decyzje wpływają na organizację**
 - Bazy danych umożliwiają menadżerom na wszystkich poziomach zarządzania zbadanie jak ich decyzje przyczyniają się do działania organizacji

BD – czynniki sukcesu

- ❑ **Używaj bazy danych jako narzędzia wspomagającego zadowolenie twoich klientów**
 - Integracja systemu informacyjnego i celów organizacji jest istotnym wyzwaniem biznesowym
 - Organizacje muszą rozwijać swe bazy danych wspierające ich strategię biznesową w celu poprawy obsługi klientów
- ❑ **Planuj odpowiednią ochronę**
 - Baza danych zawierająca wartościowe informacje musi być chroniona
 - Zarządy muszą planować i finansować odpowiedni stopień bezpieczeństwa oraz wprowadzać nowe wersje zabezpieczeń w miarę rozwoju technologii

BD – czynniki sukcesu

- **Planuj utrzymanie bazy danych**
 - Utrzymanie bazy danych jest najważniejszym etapem cyklu życia systemu informacyjnego
 - O utrzymaniu bazy danych należy myśleć już na początku procesu projektowania bazy