

Przypomnienie najważniejszych pojęć z baz danych.

1. Co to jest baza danych (Database)?

Definicja 1:

Baza danych jest zbiorem danych z określonej dziedziny tematycznej zorganizowany w sposób ułatwiający do nich dostęp oraz narzędzi systemu DBMS (System zarządzania bazą danych, SZBD (ang. Database Management System, DBMS) przeznaczonego do zarządzania bazą danych oraz gromadzenia, przekształcania i wyszukiwania danych.

Definicja 2:

Baza danych to zbiór danych, który dotyczy rzeczywistości – a konkretnie określonego jej fragmentu, który reprezentuje.

Definicja 3:

„Baza danych jest modelem pewnego aspektu rzeczywistości danej organizacji. Rzeczywistość tę nazywamy obszarem analizy (OA). (...) Baza danych jest magazynem danych z nałożoną na niego wewnętrzną strukturą. Ogólnym celem takiego magazynu jest przechowywanie danych związanych z pewnym zbiorem zadań organizacyjnych.*”

2. Jakie są podstawowe właściwości baz danych?

Własności bazy danych:

- współdzielenie danych – dane są dostępne dla więcej niż jednego użytkownika
- integracja danych – bez zbędnych i powtarzających się niepotrzebnie danych
- integralność danych – właściwe odzwierciedlenie obszaru analizy przez model
- bezpieczeństwo danych – ograniczenia dostępu do danych
- abstrakcja danych – odwzorowywanie istotnych szczegółów obszaru analizy
- niezależność danych – oddzielenie danych od procesów, które ich używają

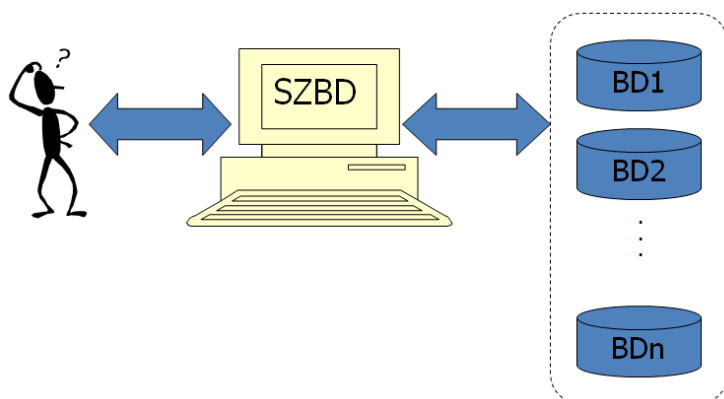
3. Co to jest system zarządzania bazą danych (DBMS)?

System zarządzania bazą danych SZBD (DBMS – Database Management System) to program zarządzający danymi w bazie i ułatwiający ich przetwarzanie. SZBD obsługuje użytkowników bazy danych, umożliwiając im eksploatację oraz tworzenie baz danych.

System bazy danych to baza danych i system zarządzania bazą danych.

Przez program zarządzający danymi rozumie się: dodawanie (wstawianie rekordów), usuwanie, modyfikowanie, nadawanie, odbieranie uprawnień.

System zarządzania bazami danych (SZBD, ang. DBMS) jest zbiorem narzędzi stanowiących warstwę pośredniczącą pomiędzy bazą danych a użytkownikiem i umożliwiającą dostęp do danych oraz zarządzanie bazami danych.



Podstawowe funkcje SZBD:

- zarządzanie plikami,
- przeszukiwanie danych,
- zarządzanie bazą danych.

4. Co oznacza relacyjna baza danych?

Fundamentem relacyjnych baz danych jest teoria relacyjna, przedstawiona przez Franka Edgara Codd'a w 1970 roku. Projekt każdej relacyjnej bazy danych, rozpoczyna etap konceptualny (abstrakcyjny), opierając się o model E-R (Entity-Relationship Model), którego autorem jest dr. Peter Chen. Jest to opis czysto teoretyczny, wymagający przełożenia na język praktyki. Ma on na celu, opisanie fragmentu rzeczywistości za pomocą związków encji. W tym modelu używamy definicji, które mają swoje odzwierciedlenie na późniejszym etapie – wdrożenia projektu w życie. W terminologii dotyczącej relacyjnych baz danych, pojawia się często wiele pojęć z różnych płaszczyzn, modeli.

Krótki rzut oka na nomenklaturę podstawowych obiektów bazodanowe w różnych terminologiach. Od czystej teorii – do wdrożenia.

Teoria relacyjna	Model ER	Relacyjne bazy	Aplikacje
Relacja	Encja	Tabela	Klasa
Krotka	Instancja	Wiersz	Instancja klasy (obiekt)
Atrybut	Atrybut	Kolumna	Właściwość, atrybut
Dziedzina	Dziedzina/Typ	Typ danych	Typ danych

Relacyjny model danych wykorzystuje tylko jedną strukturę danych – relację. **Relacja jest dwuwymiarową tabelą** spełniającą następujące zasady:

- każda relacja ma jednoznaczną nazwę;
- każda kolumna w relacji ma jednoznaczną nazwę w ramach relacji;
- wszystkie wartości w kolumnie muszą być tego samego typu;
- porządek kolumn w relacji jest nieistotny;
- każdy wiersz relacji musi być różny (powtarzanie wierszy jest niedozwolone);
- porządek wierszy nie jest istotny;
- każde pole leżące na przecięciu kolumny/wiersza w relacji powinno zawierać wartość atomową.

Twórcą założeń relacyjnego modelu danych jest dr E. F. Codd (z wykształcenia matematyk), który opublikował je w roku 1970.

Każda relacja posiada **nagłówek**, określający jej **atrybuty**.

Nazwa relacji wraz ze zbiorem jej atrybutów określana jest mianem **schematu relacji**, przedstawianego w sposób następujący:

NazwaRelacji(atrybut-1, atrybut-2,...,atrybut-n).

Kolejne wiersze relacji, poza wierszem nagłówkowym nazywane są **krotkami**.

Dla każdej relacji musi być określony jednoznaczny identyfikator określany mianem **klucz główny**.

Atrybuty			
ID	Nazwisko	Imię	Telefon
1	Kowalski	Jan	12345
2	Kowalski	Jan	67890

Podstawowe pojęcia:

- **Tabela**- pojemnik na dane, przedstawiana zwykle jako układ poziomych wierszy i pionowych kolumn. W jednej bazie danych może być wiele tabel, mogą wówczas występować między nimi powiązania logiczne związki(czasami w niektórych podręcznikach nazywane relacjami 1:1, 1:n,n:m)
- **Kwerenda(ang. Query)**- inaczej zapytanie do bazy danych. Istnieje wiele rodzajów zapytań np. kwerendy wybierające, aktualizujące, dołączające, krzyżowe, tworzące tabele
- **Rekord (krotka)**- to pojedynczy wiersz w tabeli np. zestaw cech danego człowieka

id_ucznia	imie	nazwisko	klasa
1	Adam	Mickiewicz	IV TI
2	Henryk	Sienkiewicz	I TI

- **Pole(atrybut)**- część tabeli przechowująca jednostkowe dane np. nazwisko konkretnej osoby

2	Henryk	Sienkiewicz	I TI
3	Aleksander	Fredro	III TI

- **Dana (ang.data)** □ najmniejsza, elementarna jednostka informacji o obiekcie będąca przedmiotem przetwarzania komputerowego.
- **Typ pola (ang. Data type)** rodzaj danych przechowywanych w polu;
- **Klucz podstawowy (primary key)**- zwany też kluczem głównym to jedno lub więcej pól, których wartość jednoznacznie identyfikuje każdy rekord w tabeli. Taka cecha klucza nazywana jest unikalnością. Klucz podstawowy służy do powiązania rekordów w jednej tabeli z rekordami z innej tabeli. Klucz podstawowy jest nazywany kluczem obcym, jeśli odwołuje się do innej tabeli.

Cechy klucza głównego:

1. NOT NULL – musi istnieć
2. ID - UNIKALNY
3. Krótki
4. Atrybut niezmienny w czasie

Uwaga! Pesel – nie jest dobrym przykładem na klucz główny ponieważ jest długi, a po drugie obcokrajowcy nie mają PESEL i wtedy nie można zostawić wartości pustej. Najlepiej klucz główny tworzy się przez nazwę ID_nazwatabeli, np. ID_uczen, ID_pracownik, ID_towar.

Zdarzało się na egzaminach, głównie na maturze z informatyki, że z „góry”, czyli ktoś kto układał pytania na egzamin w OKE założył że pesel jest kluczem głównym, wówczas należy postępować z myślą układającego pytania.

- **Klucz podstawowy jednopolowy**- Jeśli istnieje pole zawierające dane unikatowe, np. liczba porządkowa czy numer identyfikacyjny, można je zadeklarować jako klucz podstawowy. Jeśli jednak w polu tym powtarzają się wartości, klucz podstawowy nie zostanie ustawiony. Należy w takim przypadku usunąć rekordy o powtarzających się wartościach bądź zdefiniować klucz wielopolowy.
- **Klucz podstawowy wielopolowy (klucz złożony)**- W sytuacji, gdy żadne z pól nie gwarantuje unikatowości wartości w nim zawartych, należy rozważyć możliwość utworzenia klucza podstawowego złożonego z kilku pól.
- **Klucz obcy (foreign key)**- klucz główny z innej tabeli
- **Relacja (ang. Relation)** to sposób łączenia powiązanych danych z różnych tabel w logiczną całość.
Relacja może być jednego z trzech typów:

- **Relacja jeden-do-jednego**

W relacji jeden-do-jednego każdy rekord w tabeli **A** może mieć tylko jeden odpowiadający mu rekord z tabeli **B**, i na odwrót: każdy rekord w tabeli **B** może mieć tylko jeden odpowiadający mu rekord z tabeli **A**.

Ten typ relacji spotyka się rzadko, ponieważ większość informacji powiązanych w ten sposób byłoby zawartych w jednej tabeli. Relacji jeden-do-jednego można używać do podziału tabeli z wieloma polami, do odizolowania części tabeli ze względów bezpieczeństwa, albo do przechowania informacji odnoszącej się tylko do podzbioru tabeli głównej, np. w tabelach słownikowych.

- **Relacja jeden-do-wielu**

Relacja jeden-do-wielu jest najbardziej powszechnym typem relacji. W relacji jeden-do-wielu rekord w tabeli **A** może mieć wiele odpowiadających mu rekordów z tabeli **B**, ale rekord w tabeli **B** ma tylko jeden odpowiadający mu rekord w tabeli **A**.

- **Relacja wiele-do-wielu**

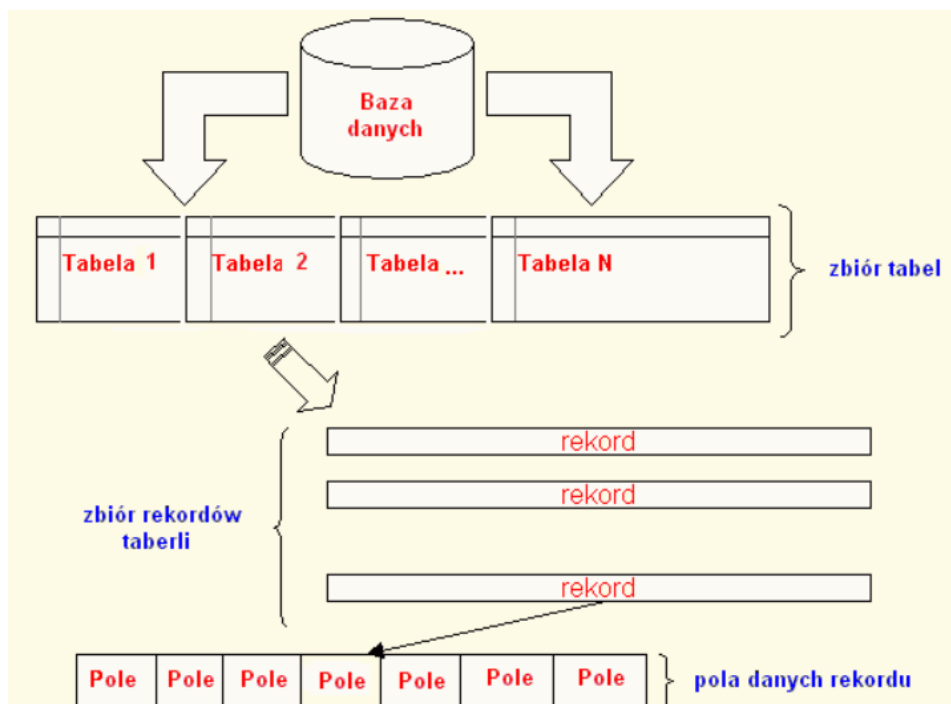
W relacji wiele-do-wielu, rekord w tabeli **A** może mieć wiele odpowiadających mu rekordów z tabeli **B** i odwrotnie: rekord w tabeli **B** może mieć wiele odpowiadających mu rekordów z tabeli **A**.

Jest to możliwe tylko przez zdefiniowanie trzeciej tabeli (nazywanej *tabelą łącznikową*), której klucz podstawowy składa się z dwóch pól: z kluczy obcych z tabel **A** i **B**.

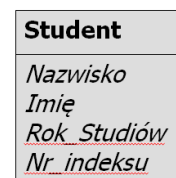
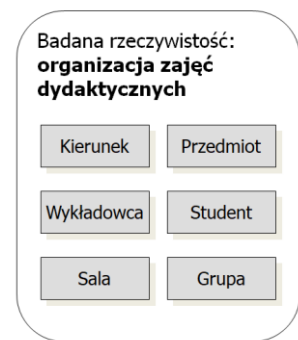
Relacja wiele-do-wielu jest w istocie podwójną relacją jeden-do-wielu z trzecią tabelą. Np.: tabele "Zamówienia" i "Produkty" mogą być powiązane relacją wiele-do-wielu zdefiniowaną przez utworzenie dwóch relacji jeden-do-wielu z tabelą "Opisy zamówień".

- **Tabele słownikowe** to tabele pomocnicze przechowujące wartości poszczególnych atrybutów opisujących jeden z typów obiektów. Na tabele słownikowe nadają się te atrybuty, których zakres dopuszczalnych wartości jest ściśle określony. Na przykład atrybut może przyjąć jedną z następujących wartości: od A do Z

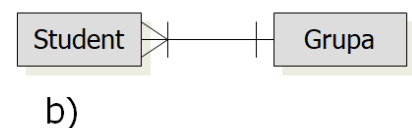
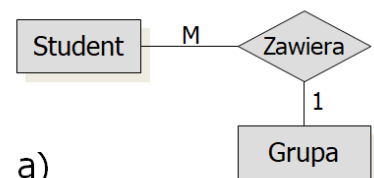
Ogólny schemat bazy danych przedstawia rysunek poniżej:



- **Encja** to pewien wyodrębniony logicznie i jednoznacznie określony byt (obiekt), rozpoznawalny w badanej rzeczywistości i pełniący w niej określoną rolę. Encja może być zarówno obiektem fizycznym (takim jak np. samochód, drzewo, książka itp.) jak również zdarzeniem (np. sprzedaż samochodu, zasadzenie drzewa, zakup książki itp.). Każda encja jest jednoznacznie identyfikowana na podstawie swojej nazwy. Przyjęło się, że nazwy encji są *rzeczownikami w liczbie pojedynczej*. Graficznie każda encja jest reprezentowana przez prostokąt.
- **Atrybuty encji** to cechy (własności) charakteryzujące daną encję w badanej rzeczywistości, którym przypisywane są określone wartości. Wartości poszczególnych atrybutów pozwalają odróżniać encje od siebie.

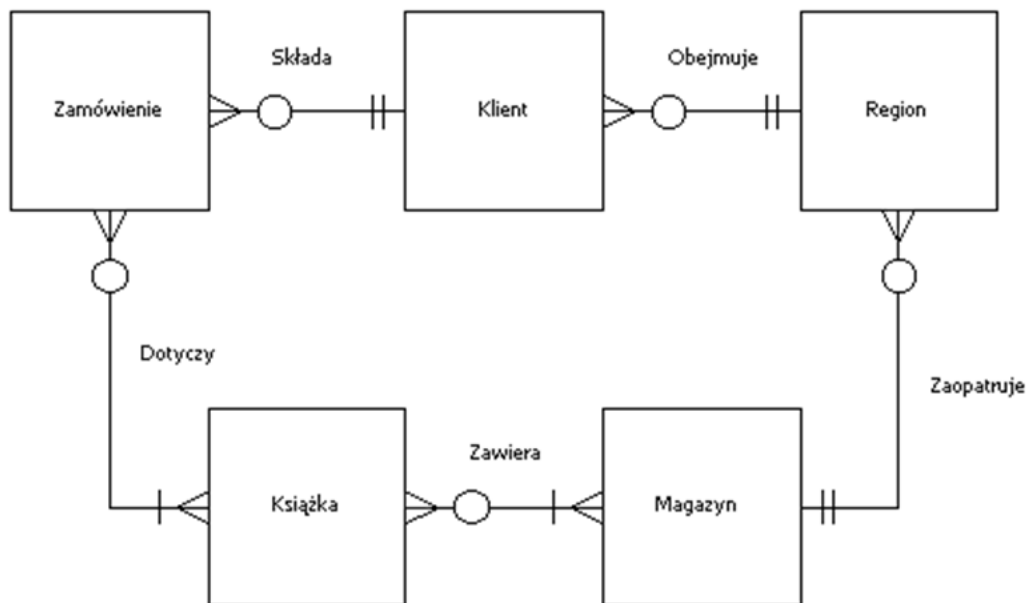


- **Związki** określają powiązania pomiędzy poszczególnymi encjami. W najprostszych modelach uwzględniane są związki występujące pomiędzy dwoma encjami. Pomiedzy dwoma różnymi encjami może zachodzić wiele związków, ale pomiędzy dwoma tymi samymi encjami może zachodzić tylko jeden związek. Każdy związek posiada swoją nazwę. Przyjęło się, że nazwy związków są czasownikami. Graficznie związek jest zwykle reprezentowany przez romb połączony liniami z encjami, pomiędzy którymi zachodzi wraz z oznaczeniem jego liczebności (a) lub też przez samą linię zakończoną symbolami określającymi jego liczebność (b). W niektórych notacjach nazwy związków się pomija.
- **Liczebność związku ENCJI**. Liczebność (stopień) związku encji określa liczbę instancji biorących udział w danym związku. Rozróżnia się związki:



- **jednojednoznaczne (jeden-do-jeden, 1:1)** – każdej instancji pierwszej encji odpowiada dokładnie jedna instancja drugiej encji i odwrotnie;
 - **jednoznaczne (jeden-do-wiele, 1:M)** – każdej instancji pierwszej encji odpowiada M instancji drugiej encji, ale każdej instancji drugiej encji odpowiada tylko jedna instancja pierwszej encji;
 - **wieloznaczne (wiele-do-wiele, M:N)** – każdej instancji pierwszej encji odpowiada M instancji drugiej encji, a każdej instancji drugiej encji odpowiada N instancji pierwszej encji.
- **Diagram związków encji (ERD)** jest graficznym przedstawieniem encji występujących w modelu wraz ze związkami między nimi. Stanowi podstawowe narzędzie koncepcyjnego etapu projektowania bazy danych.





Rysunek 1. Diagram ERD

Modele baz danych.

MODELE BAZ DANYCH

Bazy danych mogą różnić się sposobem, w jaki dane są zorganizowane i powiązane ze sobą. Przez **model danych** rozumiemy zbiór pojęć stosowanych do opisu struktury bazy danych.

Struktura BD obejmuje:

- typy danych, związki między nimi i ograniczenia nałożone na dane
- zbiór operacji do definiowania, wyszukiwania i uaktualniania BD

Wśród modeli danych można wyróżnić następujące kategorie:

- **Koncepcyjne modele danych.** Są to modele najbardziej zbliżone poziomem abstrakcji do wymagań projektantów BD, stosowane w pierwszych etapach projektów, w celu weryfikacji wyróżnionych w nim obiektów i związków między nimi
- **Implementacyjne modele danych,** stosowane do transformacji wcześniej przygotowanego modelu koncepcyjnego do konkretnego modelu BD, a więc do postaci, która jest zgodna z wymaganiami określonego SZBD.

Wśród modeli implementacyjnych wyróżniamy modele:

- ***hierarchiczny***
- ***sieciowy***
- ***relacyjny***
- ***obiektowy***
- ***obiekto-relacyjny***
- **Fizyczne modele danych,** określające sposoby organizacji danych w pamięci zewnętrznej komputerów. Operuje się tu pojęciami takimi jak np. rekord, plik, adres.