

Charakterystyka elementów bazy danych

Znajomość budowy bazy danych wymaga zwykle fachowego określenia jej elementów.

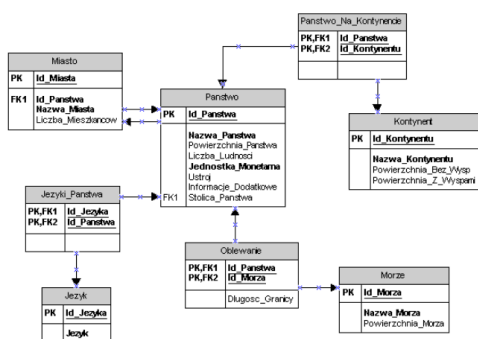
Twórca relacyjnego modelu danych - E.F. Codd - w pracy Relacyjny model danych dla dużych baz danych nie używa terminów tabela, kolumna, wiersz, lecz zamiast nich stosuje pojęcia: **relacja (zamiast tabela)**, **atrybut (zamiast kolumna)**, **krotka (zamiast wiersz)**.

Właściwe zrozumienie teorii baz danych jest kluczowe dla ich przyszłych projektantów i administratorów.

O **encji (entity)** mówimy wtedy, gdy potrzebujemy określić coś, co reprezentuje obiekt lub grupę obiektów. Pojęcia encji używamy, aby określić nie tylko obiekty fizyczne, lecz także niematerialne. Przykładami encji mogą być takie obiekty jak: OSOBA, której cechami mogą być wzrost, numer buty, waga. Cechy te nazywane są atrybutami encji.

Istnieje pogląd wskazujący na podobieństwo pomiędzy encją a obiektem (w programowaniu obiektowym). Porównanie to wskazuje na właściwości, atrybuty encji i ich podobieństwo do klas obiektu.

Dla graficznej reprezentacji encji atrybutów oraz związków używane są diagramy związków encji. Owale reprezentują atrybuty encji (autor, tytuł, ISBN), same encje zaś reprezentowane są przez prostokąty. Relacje pomiędzy encjami pokazane są za pomocą równoległoboku.



Krotka (tuple) może być zdefiniowana następująco: jeśli tabela spełnia wymogi relacji (jest relacją), a jej kolumny są atrybutami, to krotka jest jej wierszem (rekordem).

Krotka przechowuje stałe wartości o różnych typach danych, których to typów nie można zmodyfikować w kolejnej krotce. Dlatego typy, np. tytuł, ISBN, dla następnych krotek jednej tabeli będą stałe, a ich zawartości będą się różnić.

Atrybut definiowany jest jako **kolumna relacji** mająca identyfikator (nazwę). W relacyjnym modelu baz danych, gdy dwuwymiarową tabelę nazwiemy relacją (gdy spełnia warunki relacji), wówczas posiadające nazwę kolumny tej tabeli nazywamy atrybutami.

Ważną cechą tabeli - relacji jest to, że kolejność atrybutów nie powinna mieć znaczenia.

Przyjmując książkę w bibliotece jako encję, możemy wskazać jej następujące atrybuty: **ISBN**, **tytuł**, **autor**, **wydawnictwo**, **liczba_książek**, **rok_wydania**. Poniższy przykład pokazuje, że dla encji **Książka** mamy **dwie krotki**, mimo że teoretycznie mamy 3 książki (suma wartości w kolumnie **Liczba_książek**).

ISBN	Tytuł	Autor	Wydawnictwo	Liczba_książek	Rok_wydania
234-83-2623-741-8	Janko Panikant	Henryk Sienkiewicz	Greg	1	1979
779-83-2623-748-8	W pustyni i w puszczy	Henryk Sienkiewicz	Złota Era	2	2010
suma in set (0,000 row)					

Dziedzina jest zbiorem wartości, jakie może przyjąć atrybut krotki. Jeśli kolumna tabeli przechowywać będzie numery kuli używanych do losowania Lotto, dziedziną atrybutu będą numery od 1 do 49.

Gdy kolumna przechowywać będzie liczbę książek, wówczas zakładamy, że będą to wartości całkowite (ponieważ w bibliotece nie występują egzemplarze ułamkowe książek, np. 0,5 książki), zatem dziedziną kolumny będą wartości całkowite **integer**.

W bazie danych typ kolumny (**dziedzina**) może zostać zdefiniowany jako **int (11)**, co oznacza, że w kolumnie książki mogą znajdować się liczby całkowite o maksymalnej długości nieprzekraczającej jedenastu znaków.

```
! Liczba_książek ! int(11) !
```

Klucz główny

Dość często spotykamy problem na etapie projektowania bazy danych jest określenie, która kolumna (kolumny) będzie pełnić funkcję klucza głównego. Ponieważ każdy wiersz w tabeli musimy jednoznacznie zidentyfikować, zachodzi potrzeba wybrania atrybutów (kolumn), które spełnią to zadanie.

Klucz główny odgrywa bardzo ważną rolę w tabeli (relacji), dlatego jego wybór powinien zostać poprzedzony analizą typowanych przez nas kolumn pod kątem wymienionych poniżej własności.

Własności kolumn

Trwałość	Wartość kolumny powinna być stale obecna w wierszu, oznacza to, że kolumna taka (należąca do klucza głównego) nie może zawierać wartości NULL.
Unikatowość	Wartość klucza dla każdego z wierszy powinna być unikatowa, ponieważ w niepowtarzalny sposób powinien on identyfikować każdą krotkę (wiersz tabeli). Może się zdarzyć, że taki niepowtarzalny identyfikator otrzymamy, umieszczając w kluczu głównym więcej niż jedną kolumnę. Kombinacja wartości, np. trzech kolumn, które należą do klucza, będzie unikatowa i jednoznacznie zidentyfikuje każdą krotkę.
Stabilność	Wartości klucza nie powinny podlegać zmianom. Nie powinno się jako kluczy głównych używać kolumn przechowujących wartości nietrwałe, np. numer telefonu komórkowego, ponieważ mimo jego unikatowości każdy człowiek może go zmienić, np. gdy przejdzie do innego operatora.

Aby jednoznacznie zidentyfikować wiersz tabeli, stosuje się **atrybut (kolumnę)**, której poszczególne wartości dla kolejnych krotek (wierszy) będą **niepowtarzalne**.

Na rysunku na 4 slajdzie taka kolumna to numer ISBN. Ponieważ dla każdej książki jest on unikatowy. Oczywiście dany ISBN będzie identyfikował grupę książek o tym samym autorze, tytule oraz roku i wersji wydania, a także zawartości.

Ponieważ każda z tych książek będzie identyczna pod względem zawartości, nie ma potrzeby tworzyć w tabeli kolejnych wierszy, wystarczy wpisać, że mamy dwie książki pod tytułem W pustyni i w puszczy, a każda z nich jest wierną kopią poprzedniczki.

Atrybut będący kluczem głównym możemy stworzyć sztucznie, dla przykładu wprowadzając kolejne numerowanie wierszy 1, 2, 3, 4 itd.. pod warunkiem że każdy wiersz ma inny numer.

Możemy również posłużyć się określoną cechą (atrybutem) opisywanej rzeczywistości (encji), np. dokonując spisu ludności, możemy posłużyć się numerem PESEL.

Ponieważ każdy człowiek ma inny niepowtarzalny numer PESEL, nie zachodzi obawa, że może on się powtórzyć. Taką kolumnę (atrybut) nazywamy kluczem głównym (**primary key**).

Aby pomiędzy tabelami można było tworzyć związki, powinny one mieć przynajmniej jeden atrybut (kolumnę), który musi spełniać dwa podstawowe warunki. **Musi być unikatowy** (oznacza to, że jego wartości nie mogą się powtarzać). Drugim warunkiem jest jego **atomowość** (musi być minimalny, tzn. powinien zawierać elementarne, niepodzielne wartości).

Możemy rozróżnić następujące rodzaje kluczy:

- **klucz prosty** to taki, który jest jednoelementowy (składa się z jednej kolumny),

- **klucz złożony** to taki, który jest kilkuelementowy (składa się z więcej niż jednej kolumny). Kluczem może być zatem jedna lub kilka kolumn, które wspólnie będą w stanie jednoznacznie zidentyfikować pozostałe dane w tabeli (relacji).

Kolumny, które należą do kluczy (używa się ich do jednoznacznej identyfikacji wierszy tabeli), **nazywamy atrybutami podstawowymi**.

Kolumny nienależące do kluczy (zawierają dane, które w określonej relacji są przedmiotem opisu) **nazywamy atrybutami opisowymi**.

Do łączenia dwóch tabel (np. A i B) za pomocą związków używa się klucza. Klucz pochodzący z obcej tabeli B (**w której jest on kluczem głównym**), używany do łączenia tej tabeli z tabelą A, będzie dla tabeli A **kluczem obcym**.

Superklucz (superkey) to kolumna lub zestaw kolumn jednoznacznie identyfikujących każdą krotkę tabeli.

Superklucz może zawierać kolumny, które samodzielnie mogą nie identyfikować każdej z krotek. Unikatowa identyfikacja każdej krotki może odbywać się jedynie przez zestaw np. dwóch lub trzech atrybutów.

Przedmiotem zainteresowań projektantów baz danych jest taki superklucz, który zawiera minimalny zestaw atrybutów unikatowo identyfikujących krotkę.

Superkey

Name	Age	Gender	Y/N
Zhang Zheshen	21	M	Y
Peng Qijia	21	M	Y
Xu Chi	20	M	Y
Chen Yue	19	M	Y
Dai Yibo	20	F	Y
Yang Muyang	20	M	N
Zhang Yuxuan	18	F	N
Qian Dongliang	20	M	Y
Zhu Junlin	21	M	N

Klucz kandydujący (nadklucz, klucz potencjalny) to superklucz zawierający minimalną liczbę kolumn unikatowo identyfikujących krotki relacji. W praktyce to kolumna lub kolumny, których użycie w charakterze klucza głównego jest rozważane przez projektanta bazy danych.

To właśnie twórca bazy danych decyduje, której kolumnie (kolumnom) nada funkcję klucza głównego.

Candidate Keys

StudentId	firstName	lastName	courseId
1.0002345	Jim	Black	C002
1.0001254	Jamen	Harradine	A004
1.0002349	Amenda	Holland	C002
1.0001198	Simon	McCloud	S042
1.0023487	Peter	Murray	P001
1.0019453	Anne	Norris	S042

Klucz główny (primary key) to klucz, który został wybrany, aby unikatowo identyfikować krotki tabeli. Klucz główny jest podyktowany wyborem projektanta bazy danych.

Id	Imię	Nazwisko	PESEL
1	Maciej	Nowak	74092932157
2	Józef	Kowalski	65112164813

Ponieważ mogą występować osoby o tym samym imieniu i nazwisku, wybór atrybutów **Imię** lub **Nazwisko** jako klucza głównego mógłby powodować błędy w wypadku wystąpienia dwóch osób o tych samych imionach, tych samych nazwiskach lub tych samych imionach i nazwiskach. Ponieważ każdy człowiek ma niepowtarzalny, unikatowy numer PESEL, atrybut **PESEL** spełnia wymagania kolumny jednoznacznie identyfikującej każdą krotkę. Projektant baz danych jako klucz główny wybrał atrybut **Id**. Przykład ten pozwala zrozumieć różnicę pomiędzy kluczem głównym a kluczem kandydującym. W tym przypadku kluczami kandydującymi mogą być **Id** lub **PESEL**. Ponieważ projektant, definiując tabelę, mianował **Id** kluczem głównym, atrybut **PESEL** pozostał kluczem kandydującym.

Klucz obcy - kolumna lub zestaw kolumn w jednej tabeli, która pasuje do klucza kandydującego drugiej lub tej samej tabeli. Klucze obce wyznaczane są zwykle podczas tworzenia związków pomiędzy tabelami.

Dowolne dane w kolumnie (kolumnach) klucza głównego tabeli A muszą mieć swoje odpowiedniki w odpowiadającej kolumnie (kolumnach) tabeli B (**zgodność danych kolumn w dwóch tabelach**).

Zjawisko posiadania odpowiedników klucza głównego w obrębie różnych tabel nazywamy **integralnością referencyjną**.

Klucz obcy definiowany jest również jako **kombinacja jednego lub kilku atrybutów** (pełniących funkcję klucza głównego) wykorzystywanych do tworzenia związków pomiędzy tabelami lub w obrębie jednej tabeli.

Oprócz wymienionych powyżej definicji, istnieją również takie pojęcia jak:

- Klucz prosty - klucz zawierający tylko jeden atrybut.
- Klucz złożony - składa się z więcej niż jednego atrybutu.
- Klucz potencjalny, inaczej klucz kandydujący (candidate key).

Teoria relacyjna	Model ER	Relacyjne bazy	Aplikacje
Relacja	Encja	Tabela	Klasa
Krotka	Instancja	Wiersz	Rekord
Atrybut	Atrybut	Kolumna	Pole
Dziedzina	Dziedzina/Typ	Typ danych	-