

Wykład 3. Projektowanie baz danych. Diagramy związków encji.

W wykładzie 3 przedstawiona jest podstawowa metoda projektowania (tworzenia modelu) relacyjnej bazy danych. Jest ona dwustopniowa. W pierwszej fazie projektujemy model danych dla rozważanej dziedziny zastosowań, nazywany diagramem związków encji. W drugiej fazie przekształcamy otrzymany model danych w schemat bazy danych. Specjalna grupa programów, nazywana ogólnie narzędziami CASE (Computer Aided Software Engineering), dostarcza narzędzi graficznych do projektowania i rysowania diagramów na ekranie komputera. Co więcej, dostarcza narzędzi do automatycznego generowania schematu bazy danych w konkretnym systemie baz danych.

Wprowadzane w wykładzie pojęcia ilustrowane są za pomocą zdjęć ekranów pochodzących z programu Microsoft Visio (wersja 2003), który służy do rysowania różnego rodzaju diagramów, w tym także diagramów związków encji.

Celem procesu projektowania schematu bazy danych jest:

1. Wyprecyzowanie wymagań użytkowników przyszłej bazy danych oraz dokonanie analizy tych wymagań.
2. Utworzenie schematu (modelu) przyszłej bazy danych, spełniającego wymagania użytkowników i jednocześnie gwarantującego poprawne funkcjonowanie bazy danych w ramach całego systemu informacyjnego.

Na ogół, zanim utworzy się bazę danych, dokonuje się analizy wymagań informacyjnych i przedstawia się je, w postaci modelu danych nazywanego diagramem związków encji. W diagramie tym, abstrahujemy od szczegółów technicznych związanych z implementacją danych w konkretnym systemie baz danych.

Diagram związków encji – pośredni model projektowy.

Diagram związków encji stanowi z jednej strony model stanowiący model obiektów oraz łączących je zależności, z drugiej jest modelem przyszłej bazy danych. Powinien w sposób jednoznaczny określać wymagania użytkowników, umożliwiając im sprawdzenie, czy analityk systemu dobrze zrozumiał ich intencje i specyfikę działania firmy. Ponadto powinien być istotnie prostszy od schematu bazy danych, ponieważ abstrahuje od szczegółów implementacyjnych, które muszą być później opracowane przez projektanta bazy danych tak, aby baza danych mogła powstać i spełniać postawione przed nią zadania.

Składniki diagramu związków encji.

Encja

Entia non sunt multiplicanda praeter necessitatem (Novacula Occami)

Encja (łacińskie entia - byt, istota, obiekt) to coś, co istnieje, co jest odróżnialne od innych, o czym informację trzeba znać lub przechowywać. Encje o tych samych własnościach tworzą typy (zbiory) encji. Reprezentacją graficzną encji na diagramach jest ramka (prostokąt).

Należy odróżniać **typ encji**, jako opis pewnej klasy obiektów, od jej **instancji** (egzemplarza), czyli pojedynczego przedstawiciela danej klasy. Na przykład **OSOBA**, jako **typ** (klasa obiektów) i jako konkretny **obiekt** (instancja, egzemplarz, czyli fizyczna osoba).

Encja stanowi model tabeli. Atrybuty modelują kolumny tabeli. Nazwy atrybutów to nazwy kolumn tabeli, dziedziny atrybutów to typy danych akceptowane przez odpowiednie kolumny tabeli.

Student	
PK	<u>Student Id</u>
I1	Nazwisko Imie Data_rekrutacji Data_obrony Nr_indeksu

Atrybut

Atrybut jest to właściwość encji danego typu, opisywana pewną wartością np. liczbą całkowitą, liczbą rzeczywistą, napisem, datą. Inaczej mówiąc – atrybut to cecha (właściwość) obiektu, którego modelem jest encja.

Zbiór **atrybutów** opisuje encję, a zbiór konkretnych wartości atrybutów opisuje instancję, czyli konkretny egzemplarz encji.

Atrybut powinien opisywać encję, przy której się go umieszcza, a nie związki z innymi encjami(!). Na przykład *NumerMiejsca_w_Samolocie* jest atrybutem encji **Samolot** lub encji **Miejsce_w_samolocie**, a nie atrybutem encji **Bilet**, choć na nim się także pojawia. Podobnie, miejsce zajęte przez zawodnika w zawodach sportowych nie jest atrybutem encji **Zawody**, ani encji **Zawodnik**. Jest ono atrybutem encji **Wyniki_zawodów**.

Pierwsza postać normalna

Dla każdego egzemplarza encji, każdy jej atrybut powinien przyjmować pojedynczą, atomową (niepodzielną) wartość. Wymóg atomowości wobec wartości atrybutu instancji encji jest jednym z postulatów Codd'a. Z tego powodu, atrybut *DzieciPracownika* nie jest dobrym kandydatem na atrybut encji **Pracownik**.

Przy doborze atrybutów należy pomijać atrybuty wyliczane, czyli takie, których wartości dadzą się wyliczyć z innych, już istniejących w bazie wartości. Np. wiek osoby jest wyliczany z daty urodzenia, cena brutto wynika z ceny netto i stawki podatku VAT, wartość zakupu z liczby zakupionych przedmiotów i ich cen. Zapisywanie tego typu wartości wprowadza redundancję (nadmiarową informację) i może być przyczyną przechowywania w bazie niespójnych danych.

Klucz - jednoznaczny identyfikator.

Klucz jest to zbiór (może być jednoelementowy) atrybutów danej encji, których wartości jednoznacznie identyfikują każdą instancję tej encji. Dla każdej encji jeden klucz jest określany, jako główny. Pozostałe, (jeśli istnieją) to klucze alternatywne (kandydujące). Jedna encja może mieć wiele kluczy, ale tylko jeden z nich jest wybierany do roli klucza głównego. Można powiedzieć, że klucz to właściwość, a klucz główny to rola, która może być do zbioru atrybutów przypisana, jeżeli zbiór spełnienia tą właściwość.

Atrybuty klucza głównego wyróżnia się etykietą PK i/lub podkreśleniem.

Encje zależne (slabe) i niezależne.

Niektóre encje, w celu jednoznacznego ich zidentyfikowania, wymagają powiązania z innymi encjami, czego efektem jest umieszczenie klucza obcego wśród atrybutów klucza głównego. Encje takie nazywają się słabe albo zależne. Pozostałe encje to encje niezależne. Analizę zaczynamy od zidentyfikowania encji niezależnych.

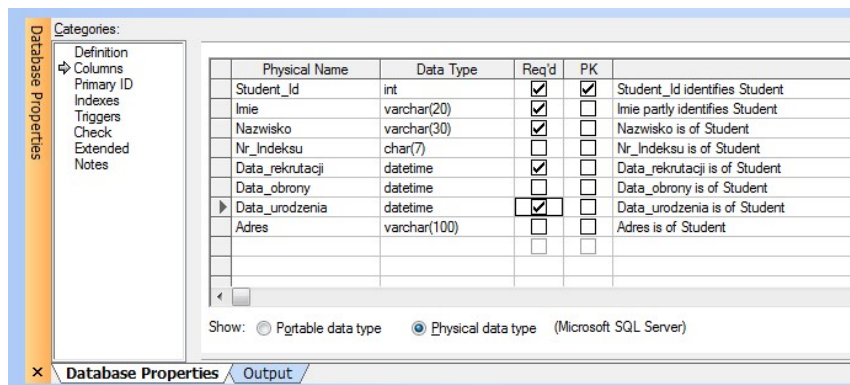
Dziedziny atrybutów.

Typy danych, czyli dziedziny atrybutów, to zbiory wartości, które mogą być zapisywane w kolumnach tabel. Każda kolumna w tabeli musi mieć zdefiniowany typ danych, zatem każdy atrybut encji musi mieć określoną dziedzinę.

W MS Visio atrybuty encji określa się w zakładce "Columns".

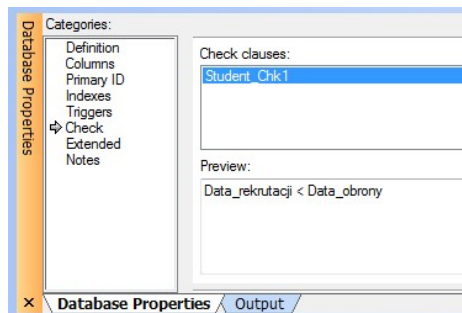
Odróżnia się dwie specyfikacje typów danych: niezależną od systemu baz danych (Portable Data Type), oraz zorientowaną na generowanie bazy danych do konkretnego systemu baz danych (Physical Data Type).

Poniżej widok okna MS Visio, pozwalającego zdefiniować dziedziny atrybutów (pokazane ustawienie dla MS SQL Server), więzy klucza głównego (PK) oraz więzy NOT NULL (Req'd – Required).



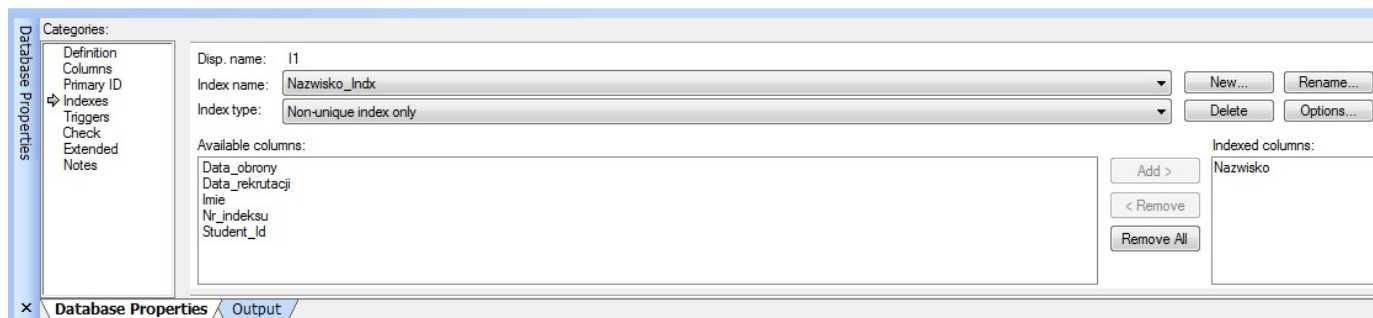
Więzy spójności.

Więzy spójności dla encji (inne niż więzy klucza głównego, jednoznacznego, więzy Not Null oraz referencyjne) w MS Visio określa się zakładce Check. Definiowane są przez wyrażenia. Poniżej przykład zdefiniowanych w MS Visio więzów Check dla encji student, eliminujących przestawienie dat.



Indeksy

Wśród atrybutów encji wyróżniamy atrybuty, bądź grupy atrybutów, których wartości są używane przy wyszukiwaniu egzemplarzy encji. Dla takich atrybutów specyfikujemy indeksy. Indeks na kluczu głównym jest zakładany automatycznie i nie trzeba dodatkowo go specyfikować. Na diagramie atrybuty posiadające indeksy są etykietowane literą I ze wskaźnikiem. W MS Visio indeksy specyfikuje się w zakładce "Indexes". Przykład pokazany na slajdzie to specyfikacja indeksu do wyszukiwania osób według ich nazwisk.



Uwaga: nie wszystkie serwery bazy danych są w stanie implementować więzy w ten sposób zdefiniowane w MS Visio.

Związki encji.

Związek definiujemy, jako uporządkowaną listę encji. Poszczególne encje mogą na niej występować wielokrotnie. Każdy związek określa pewną zależność między zbiorami egzemplarzy (instancji) encji wchodzącymi w skład związku, czyli instancję związku.

Związek można formalnie zapisać przy użyciu notacji relacyjnej:

Z(E1 ,...,En)

co oznacza: encje E1 ,...,En wchodzą w skład związku Z

Można także opisać związek werbalnie, np.:

Pracownik pracuje w dziale (związek pomiędzy encjami **Dział** i **Pracownik**)

Pracownik w projekcie pełni rolę (związek pomiędzy encjami **Pracownik** i **Projekt**)

Firma produkuje towar (związek pomiędzy encjami **Firma** i **Towar**)

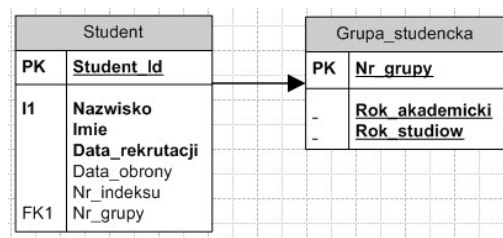
W praktyce, podczas projektowania bazy danych, związki identyfikujemy przez ich werbalny opis.

Jako przykład rozważmy zależność zachodzącą pomiędzy studentem, a grupą studencką. Opisowo (werbalnie), możemy ją opisać następująco: student zapisany jest do jednej grupy studenckiej, ale do tej samej grupy może być zapisanych kilku (wielu) studentów.

Model tego związku zostanie przedstawiony, jako związek dwóch encji: **Student** i **Grupa_studencka**.

Związek binarny (dwuargumentowy – w jego skład wchodzi dwie encje) jest reprezentowany graficznie, jako linia łącząca dwie ramki (encje). MS Visio automatycznie tworzy w encji **Student** atrybut *Nr_Grupy* (z etykietą FK1 - klucz obcy) - określający powiązania instancji encji **Student** z instancjami encji **Grupa**.

Instancja związku binarnego jest dwuargumentową relacją na iloczynie kartezyjskim zbiorów instancji encji. W naszym przykładzie - zbioru studentów, ze zbiorem grup studenckich.



Wszystkie narzędzia CASE, niezależnie od używanej notacji, umożliwiają łatwe modelowanie związków binarnych pomiędzy encjami. W MS Visio symbolem związku jest strzałka, której grot wskazuje na stronę jeden związku.

Związek jest uporządkowaną (def) listą encji, czyli nie jest strukturą symetryczną. W związku wyróżniamy *stronę jeden* i *stronę wiele*. Strona *jeden* związku to encja, której klucz główny definiuje związek. Strona *wiele* związku to encja, której klucz obcy definiuje związek. Jak już wcześniej wspomnieliśmy, w praktyce związek definiowany jest przez opis werbalny, który pozwala także zidentyfikować stronę *jeden* związku (encję nadrzędną) i stronę *wiele* związku (encję podrzędną).

W tworzonym modelu, po połączeniu dwóch encji związkiem, w encji podrzędnej (po stronie wiele) umieszczany jest atrybut (lub atrybuty) klucza głównego encji nadrzędnej (po stronie jeden), z przypisaną odpowiednią dziedziną (typem danych) i oznaczone symbolem FK (klucz obcy). Narzędzia CASE automatycznie, po utworzeniu związku, w encji podrzędnej tworzą atrybut klucza obcego, o nazwie i typie danych zgodnymi z kluczem głównym encji po stronie jeden związku.

Związek jednoznaczny.

Gdy instancja związku binarnego jest funkcją (częściową), związek nazywa się jednoznaczny.

Instancja związku jednoznacznego między encjami Grupa i Student jest funkcją częściową, ze zbioru studentów, w zbiór grup studenckich, przy założeniu, że każdy student jest zapisany, do co najwyżej jednej grupy, lub funkcją zupełną, przy założeniu, że każdy student jest przypisany do dokładnie jednej grupy.

Część związku odpowiadająca dziedzinie funkcji jest nazywana stroną wiele związku (lub encją podrzędną), a część odpowiadająca przeciwdziedzinie funkcji stroną jeden związku (lub encją nadrzędną) – i jest oznaczana strzałką w notacji przyjętej w MS Visio.

Grupa jest encją po stronie jeden (nadrzędną), a **Student** encją po stronie wiele (podrzędną).

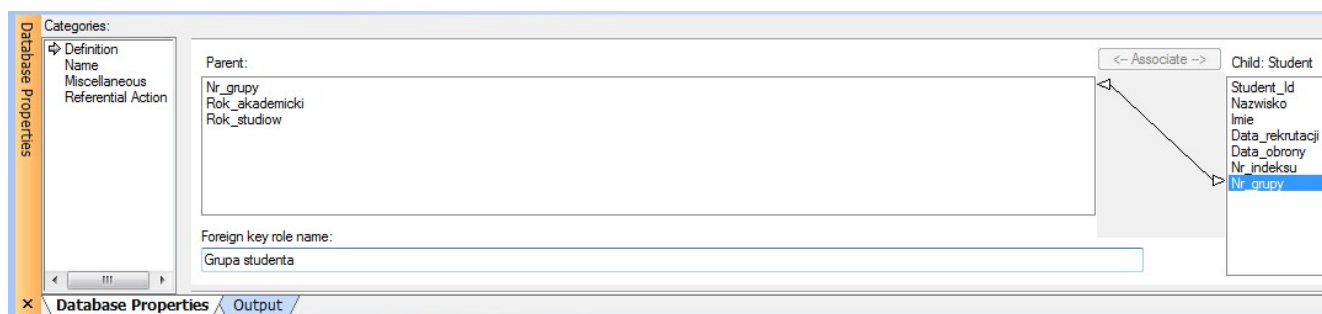
Implementacja związku jednoznacznego

Dwie encje, połączone związkiem jednoznacznym, są implementowane odpowiednio przez dwie tabele. W encji po stronie wiele dodawany jest klucz obcy, określający powiązanie z instancją encji po stronie jeden. W naszym przykładzie, do encji **Student** jest dodany klucz obcy *Nr_grupy* etykietowany przez FK1, określający powiązanie z instancją encji **Grupa** - przez wartość klucza głównego *Nr_grupy*.

Definiowanie związku jednoznacznego w MS Visio

Dostęp do okna właściwości związku w MS Visio uzyskujemy przez kliknięcie w strzałkę symbolizującą związek. W zakładce Definition związek definiowany jest, jako powiązanie dwóch zbiorów atrybutów: tworzonego automatycznie klucza obcego w encji po stronie "wiele" i klucza głównego w encji po stronie "jeden". Pole „Foreign key role name” jest miejscem, gdzie można opisać rolę, jaką w encji podrzędnej pełni klucz obcy (może być to przekształcone w etykietę kolumny przyszłej tabeli).

Poniżej sposób zdefiniowania związku encji Student z encją Grupa w oknie edycji związku MS Visio.

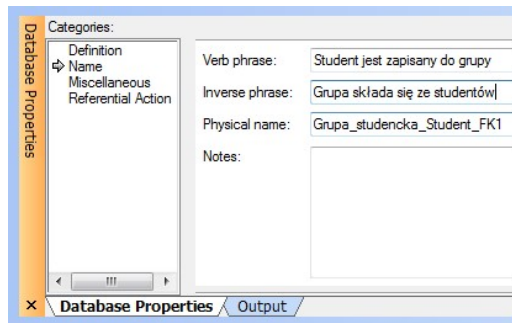


Nazwa związku

Związek encji zostanie „materializowany”, jako referencyjne więzy integralności (obiekt bazy danych) pomiędzy tabelami, w które zostaną przekształcone encje tworzące

związek. Jako obiekt bazy, otrzyma nazwę wygenerowaną automatycznie przez SZBD, lub nadaną na etapie projektowania.

W MS Visio w zakładce Name, możemy określić sposób odczytywania istoty związku, oraz nazwę dla więzów klucza obcego: Grupa_Student_FK1. Pola Verb phrase i Inverse phrase pozwalają opisać werbalnie związek, ułatwiając zrozumienie jego roli. Opis może zostać wykonany w ujęciu „od strony wiele do strony jeden” i/lub odwrotnie. MS Visio pozwala wyświetlić opisy związków na diagramie (opcja).



Liczność związku (cardinality)

Na etapie projektowania bazy możemy określić także tzw. licznosc związku, określającą ile egzemplarzy encji po stronie wiele może być połączone z jednym egzemplarzem encji po stronie jeden. Może to być konkretna liczba np. 2 albo określenie typu "zero lub wiele" (domyślne), "jeden lub wiele", "zero lub jeden".

Typ związku

Typ związku decyduje o tym, jaką rolę pełni w encji podrzędnej (encji po stronie „wiele”) klucz obcy.

Związek identyfikujący

Jeżeli klucz obcy wchodzi w skład klucza głównego, związek nazywamy identyfikującym, a encję po stronie „wiele” związku encją słabą (zależną). Zależność encji wynika z faktu, iż do identyfikacji egzemplarza encji po stronie wiele potrzebny jest odpowiadający mu egzemplarz encji po stronie jeden. Taka encja po stronie „wiele” nazywa się encją słabą (zależną). Inaczej mówiąc, atrybut klucza obcego wchodzi w skład zbioru atrybutów klucza głównego encji po stronie wiele.

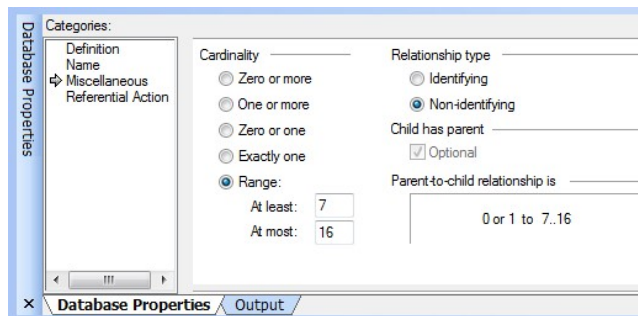
Związek nieidentyfikujący

W związku nieidentyfikującym klucz obcy nie wchodzi w skład klucza głównego encji po stronie „wiele”. Encja, w której żaden atrybut klucza obcego nie znajduje się w zbiorze atrybutów klucza głównego, nazywa się encją niezależną.

Opcjonalność związku

Związek jest nazywany opcjonalnym, gdy atrybut klucza obcego dopuszcza wartość NULL. Związek jest wymagany, gdy wartość klucza obcego musi być określona, tzn. nie może być NULL, czyli dla każdej instancji encji po stronie wiele, istnieje odpowiadająca jej instancja encji po stronie jeden. W bazie danych wymagalność związku realizowana jest przez wymuszenie NOT NULL dla więzów referencyjnych realizujących związek.

W MS Visio ustawienie typu związku, jego licznosci i opcjonalności można wykonać na zakładce **Miscellaneous**.



Akcje referencyjne przy usuwaniu i aktualizacji rekordów

Usuwanie rekordów z tabeli nadrzędnej może prowadzić do niespójności danych w bazie, jeżeli usuwane są rekordy, do których odwołują się rekordy z tabel podrzędnych. Dopuszczenie do takiej sytuacji prowadziłoby do powstania tzw. „osieroconych rekordów”, czyli rekordów odwołujących się do już nieistniejących rekordów w tabeli nadrzędnej. Identyczny problem może wystąpić przy aktualizacji rekordów tabeli nadrzędnej. W obu przypadkach rozwiązaniem jest ustanowienie jednej z możliwych akcji referencyjnych.

No action, Restricted

Odmawiaj wykonania akcji - nie wykonuj zmiany (wycofaj operację), jeśli naruszyłaby ona więzy spójności referencyjnej. System zarządzania bazą danych „odmówi” wykonania polecenia, informując o tym użytkownika i/lub zapisując informację o tym w logu. Jest to na ogół domyślne ustawienie SZBD.

Cascade

Propaguj zmiany do encji podrzędnej - przy aktualizacji instancji encji nadrzędnej, uaktualnij wartość klucza obcego w encji podrzędnej, a przy usuwaniu, razem z egzemplarzem encji nadrzędnej, usuń wszystkie powiązane przez wartość klucza obcego egzemplarze encji podrzędnej. Ta akcja jest w szczególności naturalną alternatywą wobec „no action” dla wszystkich związków identyfikujących.

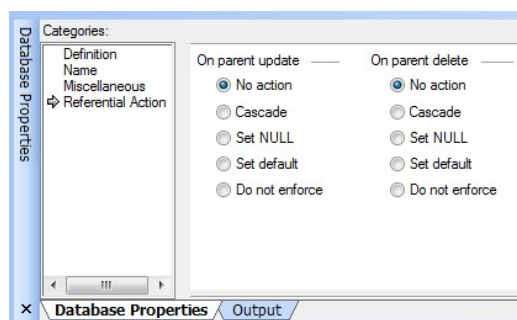
Set NULL / Set Default

Wstaw NULL (Set Null) - w przypadku aktualizacji lub usuwania instancji encji nadrzędnej, w miejsce wartość klucza obcego w odpowiadających jej instancjach encji podrzędnej wstaw NULL. W sposób oczywisty niedopuszczalne dla związków identyfikujących.

Niektóre systemy dopuszczają jeszcze jedną opcję - wstaw wartość domyślną (Set Default). W przypadku aktualizacji, lub usuwania instancji encji nadrzędnej, zamiast wartości klucza obcego w odpowiadających jej instancjach encji podrzędnej wstaw wartość domyślną.

Do not enforce

Wyłącz więzy spójności referencyjnej i wykonaj operację. System zarządzania bazą danych przestaje odpowiadać za spójność danych. Akcja przeznaczona niemal wyłącznie do zadań administracyjnych.



Akcje referencyjne, zdefiniowane na etapie projektu diagramu związków encji, w bazie danych są realizowane przez odpowiednie zdefiniowanie więzów referencyjnych.