

Bazy danych i systemy zarządzania

Wykład VI

Operacje algebraiczne w relacyjnych bazach danych

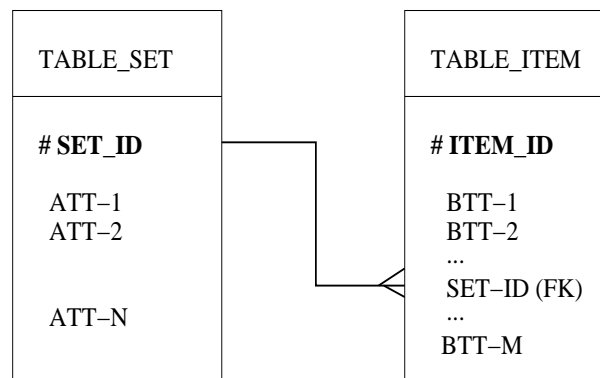
Część II

Operacje łączenia tablic

Łączenie tablic - idea

Tablice relacyjnej bazy danych mogą być łączone. Łączenie tablic wykonuje się w celu uzyskania dostępu do informacji zbiorczej znajdującej się w dwóch lub więcej tablicach. Łączenie tablic definiuje się poprzez wskazanie które rekordy jednej tablicy będą łączone z którymi rekordami drugiej tablicy.

Podstawowy i najczęściej spotykany schemat połączenia to połączenie wynikające z relacji *klucz – klucz obcy*. W tabeli nadrzędnej klucz identyfikuje obiekty nadrzędne, takie jak wydział, zbiór, grupa. W tabeli podrzędnej encje takie jak pracownik, element, osoba mają swój identyfikator, a ich przynależność do obiektów nadrzędnych wskazana jest poprzez wartość klucza obcego (klucza z tablicy nadrzędnej pełniącego w tablicy podrzędnej rolę atrybutu).



Rysunek 1: Schemat podstawowy łączenia tablic.

Operacje łączenia tablic (join)

Operacja θ złączenia (θ -join)

Technologia relacyjnych baz danych umożliwia łączenie informacji zawartej w wielu tabelach; połączenia definiuje się pomiędzy rekordami dwóch wybranych tabel. Najbardziej ogólną operacją łączenia tabel jest operacja θ -złączenia, gdzie θ oznacza warunek dopuszczający połączenie rekordów.

Niech $R(A_1, A_2, \dots, A_n)$ oraz $S(B_1, B_2, \dots, B_m)$ będą dowolnymi relacjami, przy czym $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$. Niech θ będzie warunkiem logicznym określonym na atrybutach zbioru $\mathbf{A} \cup \mathbf{B}$ (niekoniecznie wszystkich); będziemy pisać $\theta(\mathbf{A}, \mathbf{B})$. Wynikiem operacji θ -złączenia relacji R i S jest relacja $T = R \bowtie_{\theta} S$ zdefiniowana w następujący sposób:

$$R \bowtie_{\theta(\mathbf{A}, \mathbf{B})} S = \sigma_{\theta(\mathbf{A}, \mathbf{B})}(R \times S)$$

Tak więc, operacja θ -złączenia pozwala połączyć te rekordy tabeli R oraz S , które spełniają warunek θ dla wartości wybranych atrybutów (selekcja przy warunku $\theta(\mathbf{A}, \mathbf{B})$). Jeżeli warunek θ jest zawsze prawdziwy (lub nie został określony), to operacja θ -złączenia daje jako wynik iloczyn kartezyjański relacji R oraz S .

Jeżeli warunek θ jest określony na pojedynczych atrybutach, np. A_i oraz B_j , to operację θ -złączenia relacji R oraz S można zapisać w postaci takiej jak $R \bowtie_{A_i \theta B_j} S$ lub $R \bowtie_{i \theta j} S$.

Logiczna definicja operacji θ -złączenia ma postać:

$$\begin{aligned} & \| (R \bowtie_{\theta(\mathbf{A}, \mathbf{B})} S)(d_1, \dots, d_n, d_{n+1}, \dots, d_{n+m}) \| = \\ & \| R(d_1, \dots, d_n) \| \wedge \| S(d_{n+1}, \dots, d_{n+m}) \| \wedge \theta(\mathbf{A}, \mathbf{B}). \end{aligned}$$

Przykład operacji θ -złączenia

Rozważmy relację R zadaną tablicą Pracownik:

Nazwisko	Imię	Data ur	Stawka
Abacki	Adam	61-01-01	550,00 zł
Abakowski	Alojzy	61-01-02	574, 00 zł
Adamski	Antoni	61-01-03	1275,00 zł
Adamski	Arnold	61-01-03	1280,00 zł
Adamski	Arnold	61-01-03	1295,00 zł
Aron	Antonina	61-01-03	575,00 zł
Batman	Bogusław	67-02-13	1224,00 zł
Celińska	Mirosława	69-03-08	975,00 zł
Dioniziak	Dariusz	71-10-17	3000,00 zł

oraz tablicę Nagroda

Data ur	Stawka1	Stawka2	Nagroda
61-01-03	300,- zł	1300,- zł	300,- zł
72-01-01	2999,- zł	3001,-zł	3000,- zł

przy czym θ niech będzie warunkiem postaci Pracownik.[Data ur] <= Nagroda.[Data ur] AND Pracownik.Stawka > Nagroda.Stawka1 AND Pracownik.Stawka < Nagroda.Stawka2. Wynikiem operacji θ -złączenia będzie tabela

Nazwisko	Imię	Data ur	Stawka	Data ur	Stawka1	Stawka2	Nagroda
Abacki	Adam	61-01-01	550,00 zł	61-01-03	300,- zł	1300,- zł	300,- zł
Abakowski	Alojzy	61-01-02	574, 00 zł	61-01-03	300,- zł	1300,- zł	300,- zł
Aron	Antonina	61-01-03	575,00 zł	61-01-03	300,- zł	1300,- zł	300,- zł
Dioniziak	Dariusz	71-10-17	3000,00 zł	72-01-01	2999,- zł	3001,- zł	3000,- zł

Równozłączenie (equijoin)

Jednym z najczęstszych przypadków operacji θ -złączenia jest *operacja równozłączenia*. W przypadku równozłączenia warunek θ ma postać równości wybranych atrybutów obu tabel.

Niech $R(A_1, A_2, \dots, A_n)$ oraz $S(B_1, B_2, \dots, B_m)$ będą dowolnymi relacjami, przy czym $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$. Wynikiem operacji równozłączenia relacji R i S po atrybutach $A^1, A^2, \dots, A^k \in \mathbf{A}$ ze schematu relacji R oraz atrybutach $B^1, B^2, \dots, B^k \in \mathbf{B}$ ze schematu relacji S jest relacja

$$R \bowtie_{A^1=B^1 \wedge A^2=B^2 \wedge \dots \wedge A^k=B^k} S = \sigma_{A^1=B^1 \wedge A^2=B^2 \wedge \dots \wedge A^k=B^k}(R \times S)$$

Tak więc, operacja równozłączenia pozwala połączyć te rekordy tabeli R oraz S dla których wartości wybranych atrybutów są równe. Zauważmy, że dla realizacji operacji równozłączenia nie musi istnieć wspólny atrybut dla obu relacji (może być $\mathbf{A} \cap \mathbf{B} = \emptyset$).

Jeżeli warunek θ jest określony na pojedynczych atrybutach, np. A_i oraz B_j , to operację θ -złączenia relacji R oraz S można zapisać w postaci takiej jak $R \bowtie_{A_i=B_j} S$ lub $R \bowtie_{i=j} S$.

Operacja równozłączenia jest szczególnym (ale najczęściej spotykanym) przypadkiem operacji θ -złączenia.

Maksymalna liczba elementów w relacji wynikowej dla operacji θ -złączenia wynosi $\text{card}(R) \cdot \text{card}(S)$. Jeżeli atrybuty definiujące połączenie z relacji R tworzą klucz (lub indeks jednoznaczny), to maksymalna liczba elementów relacji wynikowej wynosi $\text{card}(S)$; jeżeli oba zbiory atrybutów definiujących połączenie stanowią klucze, to maksymalna liczba elementów relacji wynikowej wynosi $\min(\text{card}(R), \text{card}(S))$.

Złączenie naturalne (natural join)

Operacja złączenia naturalnego stanowi z kolei szczególny przypadek operacji równozłączenia. Polega ona na połączeniu dwóch relacji według wartości wspólnych atrybutów.

Niech $R(A_1, A_2, \dots, A_n)$ oraz $S(B_1, B_2, \dots, B_m)$ będą dowolnymi relacjami, przy czym $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$. Istotnym założeniem pozwalającym na realizację złączenia naturalnego jest istnienie wspólnych atrybutów w łączonych relacjach, tzn.:

$$\mathbf{A} \cap \mathbf{B} \neq \emptyset.$$

Niech zatem $\mathbf{A} \cap \mathbf{B} = \{A^1, A^2, \dots, A^k\}$. Wynikiem operacji równozłączenia relacji R i S po atrybutach A^1, A^2, \dots, A^k jest relacja

$$R \bowtie_{R.A^1=S.A^1 \wedge \dots \wedge R.A^k=S.A^k} S = \pi_{\mathbf{A} \cup \mathbf{B}} \sigma_{R.A^1=S.A^1 \wedge \dots \wedge R.A^k=S.A^k} (R \times S)$$

Tak więc, operacja złączenia naturalnego pozwala połączyć te rekordy tabeli R oraz S dla których wartości *wszystkich wspólnych* atrybutów są równe. W tabeli wynikowej powtarzające się kolumny są pomijane (projekcja na atrybuty zbioru $\mathbf{A} \cup \mathbf{B}$). Zauważmy, że dla realizacji operacji złączenia naturalnego musi istnieć przynajmniej jeden wspólny atrybut dla obu relacji; ponadto, jeżeli znane są schematy obu relacji, to zbiór wszystkich wspólnych atrybutów jest zadany jednoznacznie i nie trzeba go podawać jawnie. Stosowane są wtedy zapisy:

$$R \bowtie S$$

$$R * S.$$

Operacja złączenia naturalnego jest szczególnym (ale najczęściej spotykanym) przypadkiem operacji równozłączenia. Maksymalna liczba elementów w relacji wynikowej jest ograniczana identycznie jak w przypadku równozłączenia.

Przykład operacji złączenia naturalnego

Rozważmy ponownie tablice dostawców i odbiorców materiałów budowlanych.

<i>Dostawca</i>	<i>Produkt</i>
<i>Budex</i>	<i>cegła</i>
<i>Budex</i>	<i>pustak</i>
<i>Budex</i>	<i>cement</i>
<i>Budex</i>	<i>piasek</i>
<i>Matbud</i>	<i>pustak</i>
<i>Matbud</i>	<i>cement</i>
<i>Matbud</i>	<i>gips</i>
<i>Matbud</i>	<i>piasek</i>

oraz

<i>Produkt</i>	<i>Odbiorca</i>
<i>cegła</i>	<i>Wykbud</i>
<i>pustak</i>	<i>Wykbud</i>
<i>cement</i>	<i>Rembud</i>
<i>cement</i>	<i>Wykbud</i>
<i>piasek</i>	<i>Rembud</i>
<i>piasek</i>	<i>Wykbud</i>
<i>gips</i>	<i>Rembud</i>

W wyniku operacji złączenia naturalnego dostaniemy tabelę postaci:

<i>Dostawca</i>	<i>Produkt</i>	<i>Odbiorca</i>
<i>Budex</i>	<i>cegła</i>	<i>Wykbud</i>
<i>Budex</i>	<i>pustak</i>	<i>Wykbud</i>
<i>Budex</i>	<i>cement</i>	<i>Rembud</i>
<i>Budex</i>	<i>cement</i>	<i>Wykbud</i>
<i>Budex</i>	<i>piasek</i>	<i>Rembud</i>
<i>Budex</i>	<i>piasek</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>pustak</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>cement</i>	<i>Rembud</i>
<i>Matbud</i>	<i>cement</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>gips</i>	<i>Rembud</i>
<i>Matbud</i>	<i>piasek</i>	<i>Rembud</i>
<i>Matbud</i>	<i>piasek</i>	<i>Wykbud</i>

Typy złączenia

W rezultacie złączenia (θ -złączenia, równozłączenia, złączenia naturalnego) relacji R i relacji S powstaje relacja wynikowa zawierająca tylko te konkatencje rekordów obu tabel, dla których spełnione jest kryterium złączenia. Jest to tzw. *złączenie wewnętrzne* (inner join) i jest ono przyjmowane jako domyślne. W praktycznych przypadkach, może zaistnieć potrzeba rozszerzenia relacji wynikowej o rekordy jednej lub drugiej relacji nie pasujące do żadnego rekordu drugiej tabeli. Mamy wówczas możliwość utworzenia czterech innych typów złączeń:

- *złączenie lewostronne zewnętrzne* (left outer join), tzn. złączenie zawierające wszystkie rekordy tabeli R uzupełnione tymi rekordami z tabeli S , które spełniają warunek połączenia,
- *złączenie prawostronne zewnętrzne* (right outer join), tzn. złączenie zawierające wszystkie rekordy tabeli S uzupełnione tymi rekordami z tabeli R , które spełniają warunek połączenia,
- *złączenie zewnętrzne pełne* (full outer join), tzn. złączenie zawierające wszystkie rekordy tabel R oraz S uzupełnione wartościami typu *NULL* w przypadku gdy do danego rekordu nie pasuje żaden rekord z drugiej tabeli; w istocie jest to kombinacja złączenia lewo- i prawostronnego,
- *złączenie zewnętrzne typu union*, tzn. złączenie zawierające wszystkie rekordy tabeli R nie pasujące do żadnego rekordu S uzupełnione tymi rekordami z tabeli S , które nie pasują do żadnego rekordu tabeli R ; ten typ złączenia stanowi przeciwieństwo złączenia wewnętrznego (dokładniej: uzupełnienie do iloczynu kartezyjańskiego).

We wszystkich przypadkach pola puste wypełniane są wartościami *NULL*.

Przykłady złączeń

Rozważmy następujące tabele:

<i>Produkt</i>	<i>Dostawca</i>		<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	oraz	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Matbud</i>		<i>Budex</i>	222 – 222
<i>cement</i>	<i>Budex</i>		<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Matbud</i>		<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>		<i>Neobud</i>	666 – 666
<i>gips</i>	<i>Unibud</i>		<i>Probud</i>	777 – 777
	<i>Zerobud</i>	<i>Startbud</i>		

Wynikiem złączenia wewnętrznego (naturalnego) będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666

Wynikiem złączenia typu UNION będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
	<i>Probud</i>	777 – 777
	<i>Startbud</i>	
<i>gips</i>	<i>Unibud</i>	
	<i>Zerobud</i>	

Przykłady złączeń

Wynikiem operacji lewostronnego złączenia zewnętrznego będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666
<i>gips</i>	<i>Unibud</i>	
	<i>Zerobud</i>	

Wynikiem operacji prawostronnego złączenia zewnętrznego będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666
	<i>Probud</i>	777 – 777
	<i>Startbud</i>	

Przykłady złączeń

Wynikiem operacji złączenia zewnętrznego pełnego będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666
<i>gips</i>	<i>Unibud</i>	
	<i>Zerobud</i>	
	<i>Probud</i>	777 – 777
	<i>Startbud</i>	

W wyniku poszukiwania wszystkich materiałów i ich dostawców nie posiadających telefonów otrzymamy tabelę:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>gips</i>	<i>Unibud</i>	

W wyniku poszukiwania wszystkich dostawców dostarczających określone materiały, np. *cement* oraz *piasek* posiadających telefony dostaniemy tabelę postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666

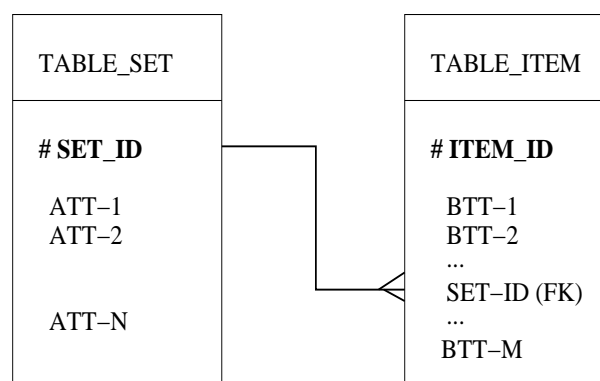
Łączenie statyczne i dynamiczne tablic

Definiowanie połączeń pomiędzy tablicami może odbywać się na dwa sposoby:

- **statycznie** – poprzez zdefiniowanie więzów integralności w definicjach odpowiednich tabel wynikających z projektu diagramu ERD (za pomocą DDL; okienko “Relacje” w ACCESS-ie),
- **dynamicznie** – poprzez zdefiniowanie na kilku tablicach operacji realizującej zapotrzebowanie na dane pochodzące z tych tablic (za pomocą DQL, DML; w systemie ACCESS w kwerendach – szablon QBE).

Statyczne więzy integralności zdefiniowane na etapie projektowania pozwalają kontrolować i utrzymywać integralność danych – dane w polu klucza obcego muszą być wcześniej zdefiniowane w polu klucza tablicy nadrzędnej.

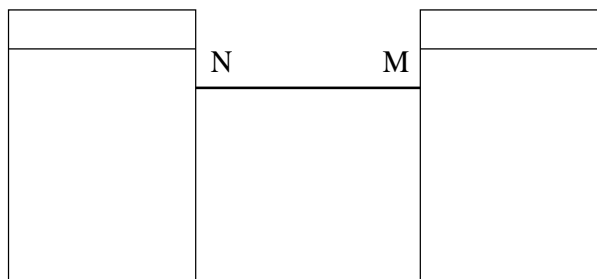
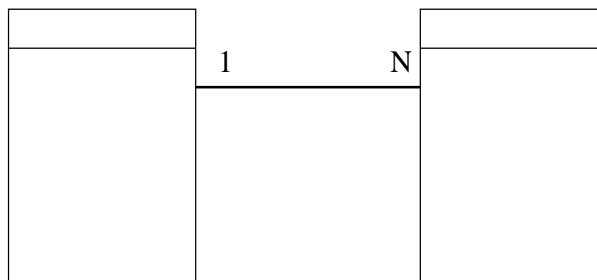
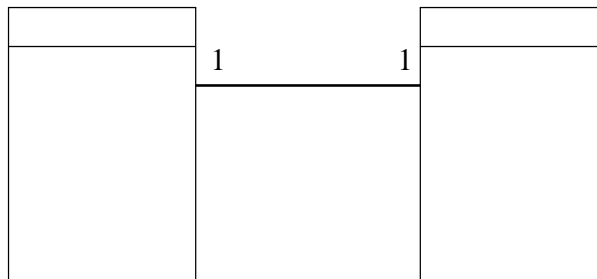
Dynamiczne łączenie tablic wykonuje się w celu uzyskania dostępu do informacji zbiorczej znajdującej się w dwóch lub więcej tablicach. Łączenie tablic definiuje się za pomocą SQL lub w kwerendach, poprzez wskazanie które rekordy jednej tablicy będą łączone z którymi rekordami drugiej tablicy.



Rysunek 2: Przykład definiowania łączenia statycznego 1–N.

Rodzaje połączeń pomiędzy tablicami

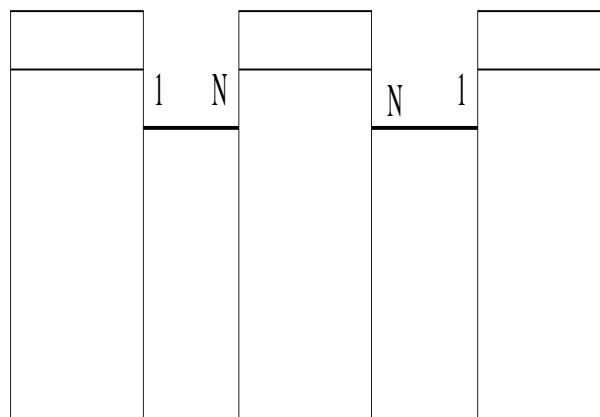
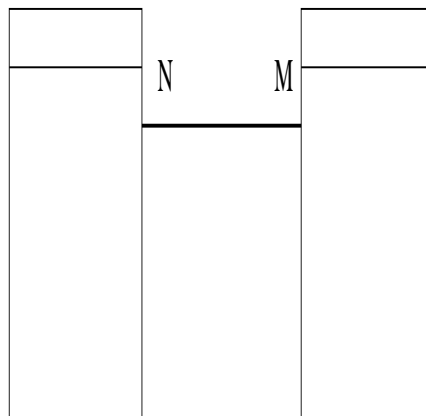
Połączenia tablic – jeden-do-jeden, jeden-do-wielu, wiele-do-wielu



Rysunek 3: Połączenia typu 1-1, 1-N, N-M.

Redukcja połączeń N–M do 1–N

Zamiana połączenia N–M na dwa połączenia jeden-do-wielu



Rysunek 4: Redukcja połączeń wiele-do-wielu

Wymuszanie więzów integralności

Więzy integralności

W praktyce definiuje się jedynie dwa typy połączeń między tabelami, są to:

- połączenia typu *jeden-do-jeden*,
- połączenia typu *jeden-do-wielu*.

Połączenia typu *wiele-do-wielu* realizowane są poprzez zdefiniowanie tablicy pośredniej (zamiana połączenia typu N–N na dwa połączenia 1–N, N–1).

Istnieje możliwość wymuszenia więzów integralności (ACCESS) nakładanych na związek i dane w łączonych tabelach – realizowane jest wtedy sprawdzanie warunków zgodności danych. Warunki te są następujące:

- zdefiniowany związek musi dotyczyć dwóch tabel w tej samej bazie danych,
- pole (pola) łączące tabeli głównej stanowią klucz lub indeks jednoznaczny,
- w odpowiadających im polach tabeli dołączanej (polach klucza obcego) nie mogą pojawiać się dane, nie występujące w tabeli głównej,
- odpowiadające sobie pola muszą być tego samego typu (Autonumer == Liczba całkowita długa).

Po wymuszeniu więzów integralności możliwe jest uaktywnienie następujących opcji:

- kaskadowa aktualizacja powiązanych pól,
- kaskadowe usuwanie powiązanych rekordów.

Opcje te pozwalają automatycznie zachowywać spójność danych.

Operacje agregujące

W przypadku dużych tablic bazodanowych często stosuje się operacje łączenia danych z wielu rekordów z wykorzystaniem funkcji agregujących, takich jak suma, średnia, czy zliczanie. Pozwala to uzyskać informację sumaryczną, ilościową, na wysokim poziomie abstrakcji.

Zasady realizacji operacji agregujących:

- grupy rekordów o identycznej wartości wskazanego atrybutu wg którego następuje grupowanie zastępowane są pojedynczym rekordem,
- dla tak powstałych rekordów sumarycznych wyliczane są dane zagregowane,
- dane indywidualne rekordów wyjściowych są pomijane (nie mogą być uwzględnione),
- grupy można formować hierarchicznie, wg kilku atrybutów grupujących.

Funkcje agregujące:

- `sum(.)` – oblicza sumę,
- `avg(.)` – oblicza średnią,
- `count(.)` – zlicza rekordy w grupie,
- `min(.)`, `max(.)` – wyznaczają wartość minimalną i maksymalną w grupie,
- `variance(.)` – wyznacza wariancję,
- `stddev(.)` – wyznacza odchylenie standardowe.

Argumentem powyższych funkcji może być atrybut tablicy (kolumna) lub wyrażenie. Atrybut(y) wg którego należy dokonać grupowania wskazywany jest w opcji `GROUP BY`.

Własności operacji algebry relacji

Operacje sumy uogólnionej $+$ (algebraicznej dla relacji zgodnych) i złączenia naturalnego $*$ (iloczynu kartezjańskiego dla relacji o rozłącznych zbiorach atrybutów) mają następujące własności:

1. Idempotentność: $R + R = R, R * R = R,$
2. Łączność: $R + (S + T) = (R + S) + T, R * (S * T),$
3. Przemienność: $R + S = S + R, R * S = S * R,$
4. Prawa rozdzielności:
 - $((R + S) * T)(X, Y, Z) = ((R * T)(X, Z) + (S * T)(Y, Z)),$
 - $(R + (S * T))(X, Y, Z) = ((R + S)(X, Y) * (R + T)(X, Z))$
5. Prawa deMorgana:
 - $\overline{(R + S)}(X, Y) = (\overline{R} * \overline{S})(X, Y),$
 - $\overline{(R * S)}(X, Y) = (\overline{R} + \overline{S})(X, Y),$

Znane operacje algebraiczne w relacyjnych bazach danych obejmują 12 operacji: $R + S, R \cup S, R \setminus S, R \cap S, R * S, R \bowtie_{\theta} S, R \times S, R \div S, R[A^1, A^2, \dots, A^k], R]B[, \overline{R}, \sigma_F(R).$ Można je wyrazić za pomocą pięciu operacji elementarnych:

- sumy,
- różnicy,
- projekcji,
- iloczynu kartezjańskiego,
- selekcji.

Np. dopełnienie $\overline{R}(A_1, A_2, \dots, A_n) = R[A_1] \times R[A_2] \times \dots \times R[A_n] \setminus R.$

Operację iloczynu można zdefiniować jako $R \cap S = \pi_R(\sigma_{=}(R \times S)).$ Inna możliwość to $R \cap S = R \setminus (R \setminus S).$ Operację różnicy można zdefiniować jako z wykorzystaniem lewostronnego złączenia zewnętrznego lub za pomocą selekcji gdy do budowy kryterium można wykorzystać negację i warunek przynależności rekordu do zbioru.

Problemy realizacji zapytań i optymalizacji ich wykonania

Ogólne zasady optymalizacji zapytań

Dla optymalizacji zapytań należy stosować odpowiednie zasady konstrukcji:

- należy unikać mnożenia krotek (iloczynu kartezjańskiego, nieprecyzyjnego θ -złączenia, niepotrzebnych operacji złączenia),
- operacje selekcji i projekcji wykonywać na jak najniższym poziomie (jak najwcześniej),
- dla częstych operacji wyszukiwania lub zawierających ukryte wyszukiwanie posługiwać się indeksami,
- operację selekcji oraz operację złączenia i następujące po niej operacje projekcji należy wykonywać łącznie (jako jedną operację),
- unikać operacji wymagających (wielokrotnego) dostępu do dysku.

Wybrane zasady szczegółowe

- uszeregować kolejne warunki (kryteria) selekcji od najsilniejszych do najsłabszych (pierwsze operacje eliminują maksimum rekordów),
- dokonywać możliwie wcześnie projekcji na możliwie mały zestaw atrybutów (z opcją DISTINCT),
- operacje selekcji przesuwac "do wewnątrz" innych operacji (np. iloczynu kartezjańskiego, sumy, różnicy),
- operacje projekcji na wybrane atrybuty przesuwac "do wewnątrz" innych operacji (np. iloczynu kartezjańskiego, sumy).