

Teoretyczne podstawy informatyki

Wykład 8b Algebra relacyjna

Algebra relacyjna

- Algebra relacyjna (*ang. relational algebra*) to specjalny język opracowany w celu ułatwienia i sformalizowania zapytań realizowanych w bazach danych.
- Umożliwia przekształcanie wyrażeń realizujących zapytania za pomocą odpowiednich praw algebraicznych.

Operandy algebry relacyjnej

- W algebrze relacyjnej operandami są relacje.
- Operandy mogą być albo stałymi (konkretnymi relacjami) albo zmiennymi reprezentującymi nieznane relacje.
- Każdy operand jest zgodny ze specyficznym schematem – jest listą atrybutów będących nazwami kolumn relacji.

Schematem relacji jest $\{A, B, C\}$,
zaś należące do niej krotki to
 $(0, 1, 2)$, $(0, 3, 4)$ oraz $(5, 3, 2)$.

A	B	C
0	1	2
0	3	4
5	2	3

Operatory dla zbiorów w algebrze relacyjnej

Suma, przecięcie oraz różnica zbiorów

Dodatkowe założenie w stosunku do tego co już znamy z operacji na zbiorach to to, że schematy operandów muszą być takie same.

A	B
0	1
2	3

relacja R

A	B
0	1
4	5

relacja S

A	B
0	1
2	3
4	5

$R \cup S$

A	B
0	1

$R \cap S$

A	B
2	3

$R \setminus S$

- Relacje są zbiorami nie mogą więc zawierać dwóch lub więcej kopii tej samej krotki.

Operator selekcji (*ang. selection operator*)

- Operuje na pojedynczej relacji będącej jego operandem, ale zawiera także dodatkowe wyrażenia warunkowe stanowiące jego parametry.
- Operator selekcji zapisujemy w postaci:

$$\sigma_C(R)$$

R – relacja
C - warunek

- Warunek C może zawierać stałe, jak i operandy będące atrybutami ze schematu relacji R.
- Operatorami wykorzystywanymi w warunku C są typowe wyrażenia warunkowe z języka programowania C, czyli wyrażenia złożone z porównań arytmetycznych oraz logicznych łączników.
- Wynikiem operacji jest relacja której schemat jest identyczny ze schematem relacji R.
- W relacji tej umieszczamy wszystkie krotki t z relacji R, dla których warunek C jest prawdziwy po podstawieniu za każdy atrybut A właściwej dla niej składowej krotki t.

Relacja ZSO Zajęcia-StudentID-Ocena

zajęcia	student ID	ocena
CS101	12345	5.0
CS101	67890	4.0
EE200	12345	3.0
EE200	22222	4.5
CS101	33333	2.0
PH100	67890	3.5

Operator selekcji

$\sigma_{\text{Zajecia} = \text{„CS101”}}$ (ZSO)

Zajęcia	Student ID	Ocena
CS101	12345	5.0
CS101	67890	4.0
CS101	33333	2.0

Operator rzutowania (ang. projection operator)

- Operuje na pojedynczej relacji będącej jego operandem, ale zawiera także dodatkowe wyrażenia warunkowe stanowiące jego dodatkowe parametry.
- Operator rzutowania zapisujemy w postaci:

$$\pi_{B_1, B_2, \dots, B_n}(R)$$

R – relacja

B_1, B_2, \dots, B_n - atrybuty

- Jeśli R jest relacją ze zbiorem atrybutów $\{A_1, A_2, \dots, A_k\}$ oraz (B_1, B_2, \dots, B_n) jest listą pewnych atrybutów A, to $\pi_{B_1, B_2, \dots, B_n}(R)$, czyli **rzutowanie relacji R na atrybuty B_1, B_2, \dots, B_n** jest zbiorem krotek utworzonych przez wybranie z każdej krotki t tylko atrybutów B_1, B_2, \dots, B_n .
- Jedna lub więcej krotek może posiadać te same wartości atrybutów B_1, B_2, \dots, B_n .
Jako wynik operacji rzutowania pojawia się tylko jedna taka krotka.

Relacja ZSO Zajęcia-StudentID-Ocena

zajęcia	student ID	ocena
CS101	12345	5.0
CS101	67890	4.0
EE200	12345	3.0
EE200	22222	4.5
CS101	33333	2.0
PH100	67890	3.5

Operator rzutowania

$\pi_{\text{StudentID}} (\sigma_{\text{Zajecia} = \text{CS101}} (\text{ZSO}))$

Student ID
12345
67890
33333

Operator łączenia (*ang. join operator*)

- Umożliwia nam przechodzenie z jednej relacji do drugiej.
- Operator łączenia zapisujemy w postaci **&&**
- Przypuśćmy, że mamy dwie relacje R i S, których zbiory atrybutów (schematy) mają odpowiednio postać $\{A_1, A_2, \dots, A_n\}$ oraz $\{B_1, B_2, \dots, B_m\}$
- Z obu zbiorów wybieramy po jednym atrybucie – powiedzmy A_{ij} i B_j – i te atrybuty są parametrami naszej operacji złączenia, której argumentami są relacje R i S.
- Złączenie relacji R i S zapisujemy: $R \underset{A_i=B_j}{\&\&} S$
i jest utworzone w wyniku porównania każdej krotki r z relacji R z każdą krotką s z relacji S.
- Jeśli składowa r odpowiadająca atrybutowi A_i jest równa składowej s odpowiadającej atrybutowi B_j to tworzymy jedną krotkę.
- Schemat złączonej relacji jest $\{A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_{j-1}, B_{j+1}, \dots, B_m\}$
- A więc atrybut B_j się nie pojawia.
- Jeżeli atrybuty A_i i B_j mają tę samą nazwę to mówimy o złączeniu naturalnym.

ZDG

Zajęcia	Dzień	Godzina
CS101	Pn	9.15
CS101	S	9.15
EE200	Pt	8.30
EE200	W	13.00
CS101	Pt	9.15

Operacja łączenia

ZK & ZDG
Zajęcia=Zajęcia

ZK

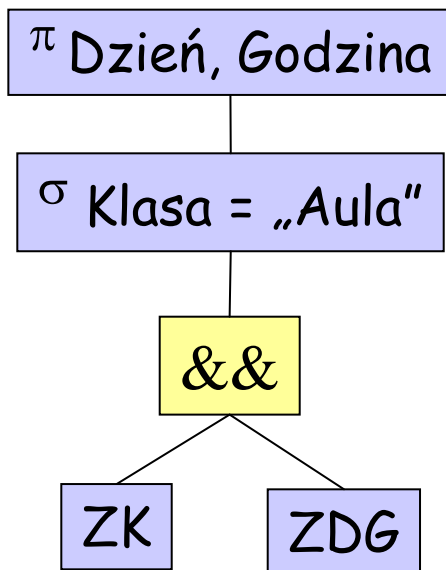
Zajęcia	Klasa
CS101	Aula
EE200	Hala
PH100	Laborat

Zajęcia	Klasa	Dzień	Godzina
CS101	Aula	Pn	9.15
CS101	Aula	S	9.15
EE200	Hala	Pt	8.30
EE200	Hala	W	13.00
CS101	Aula	Pt	9.15

Drzewa wyrażeń dla algebry relacyjnej

Drzewo wyrażenia algebry relacyjnej:

π Dzień, Godzina (σ Klasa=„Aula” (ZK && ZDG))



wynik po realizacji
drzewa wyrażenia

Dzień	Godzina
Pn	9.15
S	9.15
Pt	9.15

SQL - język oparty na algebrze relacyjnej

Wiele współczesnych systemów baz danych wykorzystuje język SQL (*ang. Structured Query Language* – strukturalny język zapytań) do wyrażania zapytań.

Operacja: $\pi_{\text{StudentID}} (\sigma_{\text{Zajęcia} = \text{„CS101”}} (\text{ZSO}))$

```
SELECT StudentID
FROM ZSO
WHERE Zajęcia=„CS101”
```

Operacja: $\pi_{\text{Dzień, Godzina}} (\sigma_{\text{Klasa} = \text{„Aula”}} (\text{ZK} \ \&\& \ \text{ZDG}))$

```
SELECT Dzień, Godzina
FROM ZK, ZDG
WHERE ZK.Zajęcia = ZGD.Zajęcia AND Klasa = „Aula”
```

Implementowanie operacji algebry relacyjnej

Suma, przecięcie i różnica:

- Można implementować analogicznie jak dla zbiorów.
- Należy przewidzieć sposób eliminacji duplikatów na pewnym etapie tych operacji
- Można wykorzystać indeks.

Rzutowanie:

- Operacja wymaga przejrzania wszystkich krotek i stworzenia kopii pozbawionych składowych odpowiadających atrybutom, których nie ma na liście rzutowania.
- Po wyznaczeniu rzutowania, np. $S = \pi_L(R)$, dla pewnej relacji R i liście atrybutów L , musimy wyeliminować duplikaty (stosujemy jedna z omówionych już metod).

Selekcja

- Wykonywanie operacji selekcji $S = s_C(R)$ na relacji R , dla której nie zdefiniowano żadnych indeksów
- Musimy przeanalizować wszystkie krotki w tej relacji w celu sprawdzenia warunku C .
- Jeżeli takie indeksy istnieją oraz można je wykorzystać do całkowitego lub częściowego sprawdzenia warunku C to bardzo przyspiesza to czas wykonania operacji.

Łączenie

- Istnieje szereg metod łączenia, różniących się czasem wykonania.
- Tylko wymieńmy: łączenie pętli zagnieżdżonej (*ang. nested loop join*), łączenie indeksowe (*ang. indeks-join*), łączenie przez sortowanie (*ang. sort-join*).
- Łączenie indeksowe wymaga istnienia indeksu na jednym z atrybutów wykorzystywanych do łączenia, łączenie przez sortowanie może być wykonywane na dowolnych relacjach.

Prawa algebraiczne dla relacji

- Podobnie jak w przypadku innych algebr, przekształcanie wyrażeń algebry zapytań umożliwia często „optymalizację” zapytań.
- Oznacza to, że możemy przekształcić kosztowne obliczeniowo wyrażenie w równoważne, którego obliczenie charakteryzuje się niższym kosztem.
- Podczas gdy przekształcanie wyrażeń arytmetycznych lub logicznych umożliwia niekiedy uzyskanie oszczędności rzędu kilku operacji, właściwe przekształcenia zastosowane do algebry relacyjnej może w znacznym stopniu skrócić czas potrzeby do wyznaczenia wartości wyrażenia.

Prawa dla łączenia

- Operator łączenia jest w pewnym sensie przemienne, w innym nie jest. Jeżeli istotna jest kolejność atrybutów to nie jest, $S \ \&\& \ R$ ma inne kolumny niż $R \ \&\& \ S$.
- Operator łączenia nie zawsze spełnia warunki prawa łączności.

Prawa dla selekcji

Najbardziej przydatne prawa algebry relacyjnej dotyczą operatora selekcji. Staramy się dokonywać selekcji na jak najwcześniejszym etapie.

- **Prawa przenoszenia selekcji** (*ang. selection pushing*)

$$(\sigma_C (R \ \&\& \ S)) == (\sigma_C(R) \ \&\& \ S)$$

$$(\sigma_C (R \ \&\& \ S)) == (R \ \&\& \ \sigma_C(S))$$

- **Prawo podziału selekcji** (*ang. selection splitting*)

$$\sigma_{C \ \text{AND} \ D} (R) == \sigma_C(\sigma_D(R))$$

- **Prawo przemienności selekcji**

$$\sigma_C(\sigma_D(R)) == \sigma_D(\sigma_C(R))$$

- Operacje selekcji możemy przenosić poniżej sumy, przecięcia i różnicy zbiorów.

„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

Wykonanie tego zapytania wymaga przechodzenia pomiędzy czterema relacjami:

1. ZSO (Zajęcia-StudentID-Ocena)
2. SNAT (StudentID-Nazwisko-Adres-Telefon)
3. ZDG (Zajęcia-Dzień-Godzina)
4. ZK (Zajęcia-Klasa)

Aby wykonać wyrażenie algebraiczne dla tego zapytania rozpocznijmy od złączenia „naturalnego” dla wszystkich czterech relacji.

1. ZSO && SNAT (porównując StudentID)
2. (ZSO && SNAT) && ZDG (porównując Zajęcia)
3. ((ZSO && SNAT) && ZDG) && ZK (porównując zajęcia)

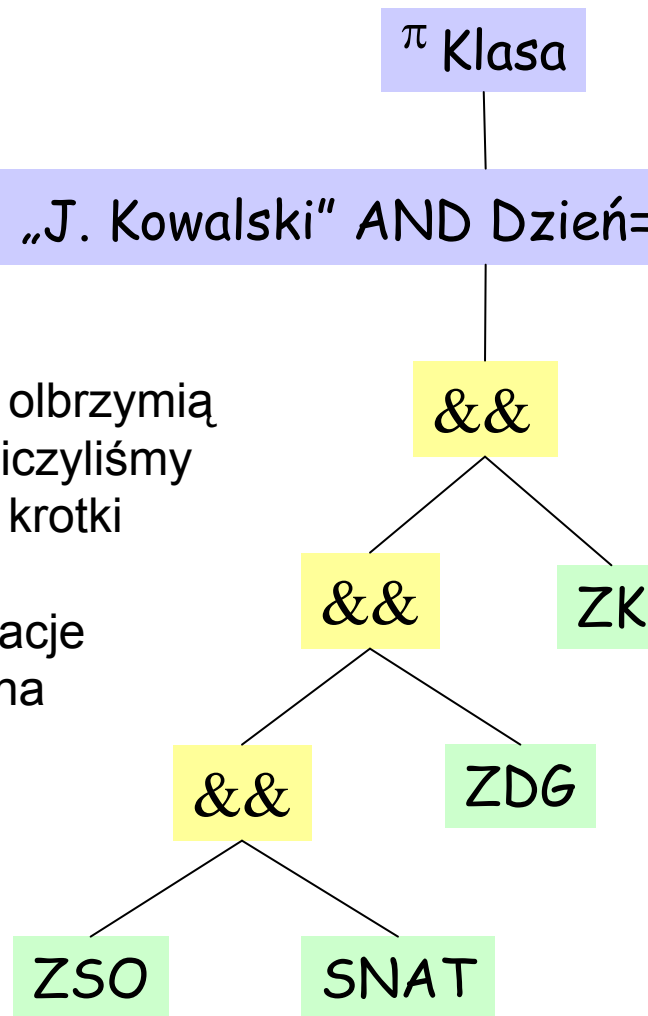
➤ Relacja wynikowa to:

{Zajęcia, StudentID, Ocena, Nazwisko, Adres, Telefon, Dzień, Godzina, Klasa}

„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?” π Klasa σ Nazwisko = „J. Kowalski” AND Dzień=„M” AND Godzina=„9.00”

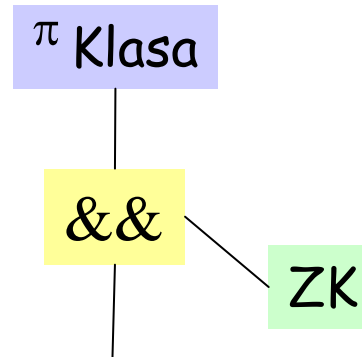
skonstruowaliśmy olbrzymią relację oraz ograniczyliśmy zbiór danych do 1 krotki

wykonaliliśmy operacje rzutowania na jedną składową.

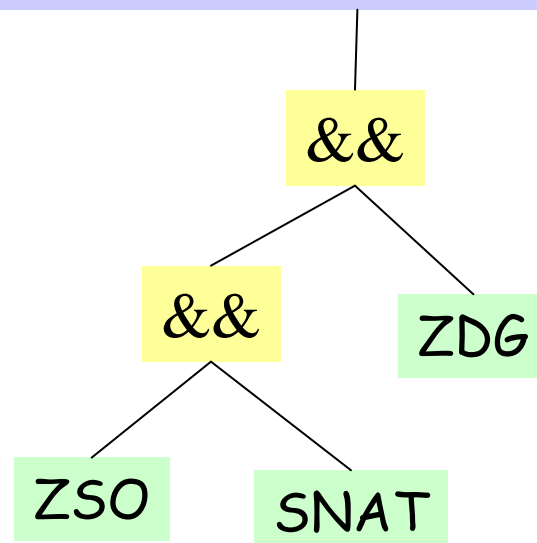


„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

przenieść operacje selekcji
poniżej najwyższej operacji
łączenia z relacją ZK

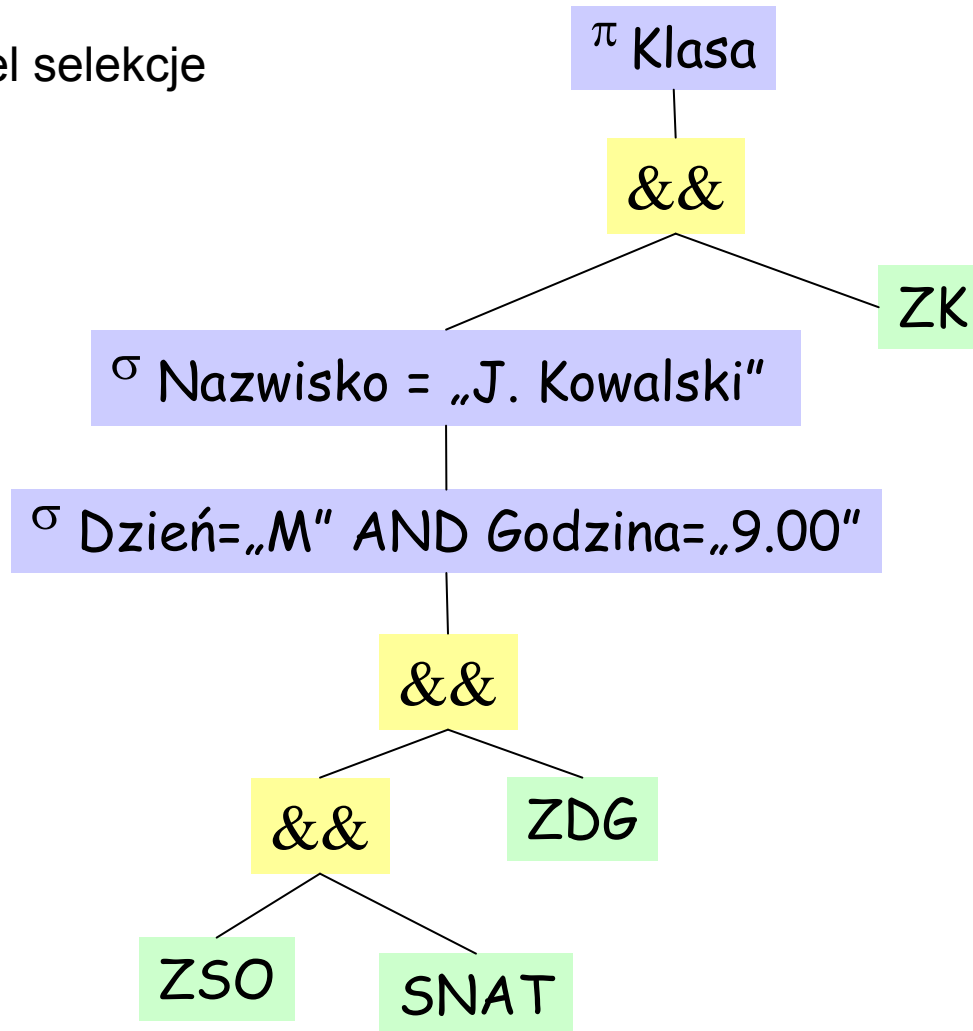


σ Nazwisko = „J.Kowalski” AND Dzień=„M” AND Godzina=„9.00”



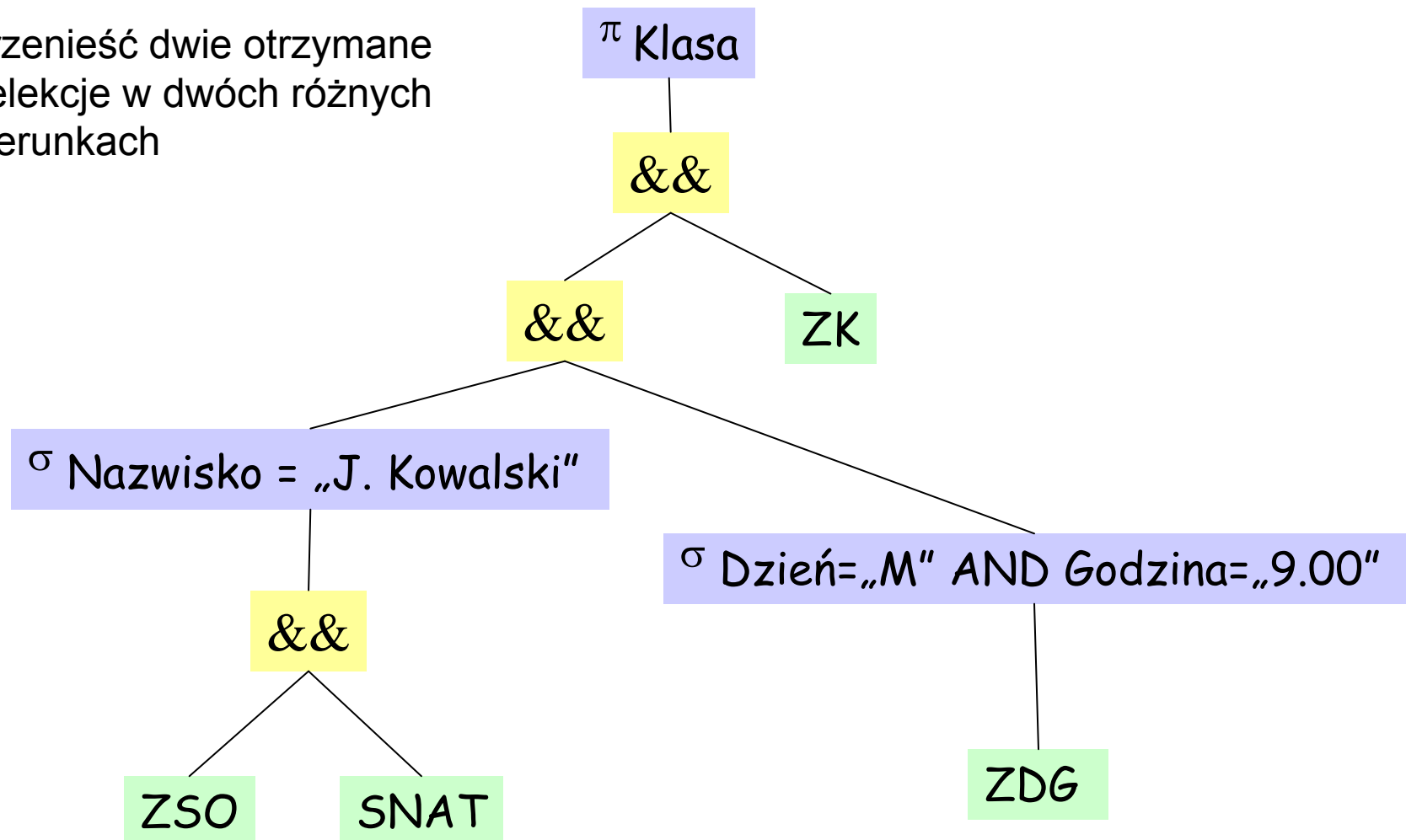
„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

rozdziel selekcje



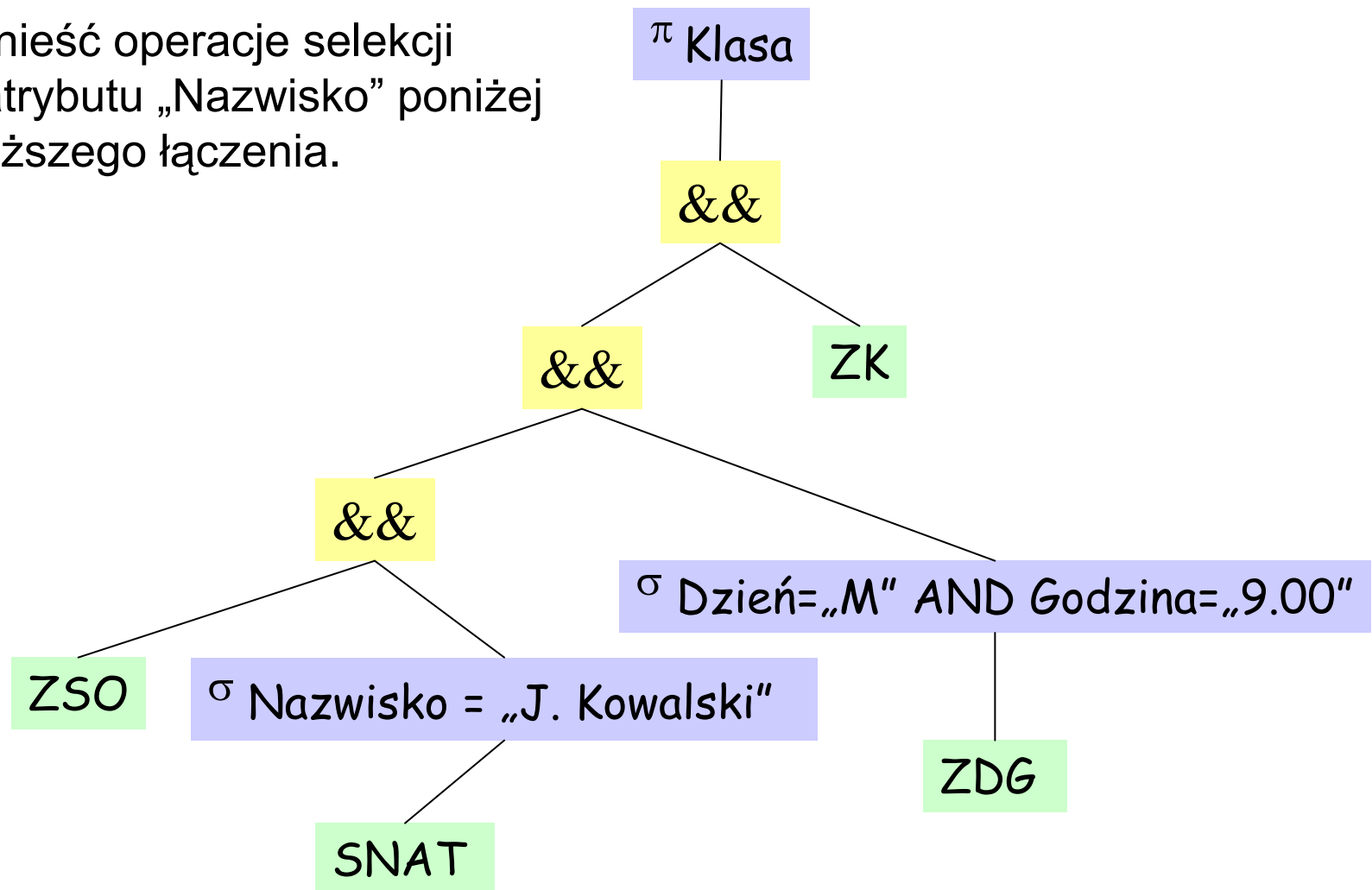
„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

przenieść dwie otrzymane selekcje w dwóch różnych kierunkach



„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

przenieść operacje selekcji
dla atrybutu „Nazwisko” poniżej
najniższego łączenia.



Prawa dla rzutowania

Rzutowanie możemy przenosić poniżej sum:

$$\pi_L(R \cup S) = (\pi_L(R) \cup \pi_L(S))$$

ale nie poniżej przecięć:

$$\pi_L(R \cap S) \neq (\pi_L(R) \cap \pi_L(S))$$

Prawo przenoszenia rzutowania:

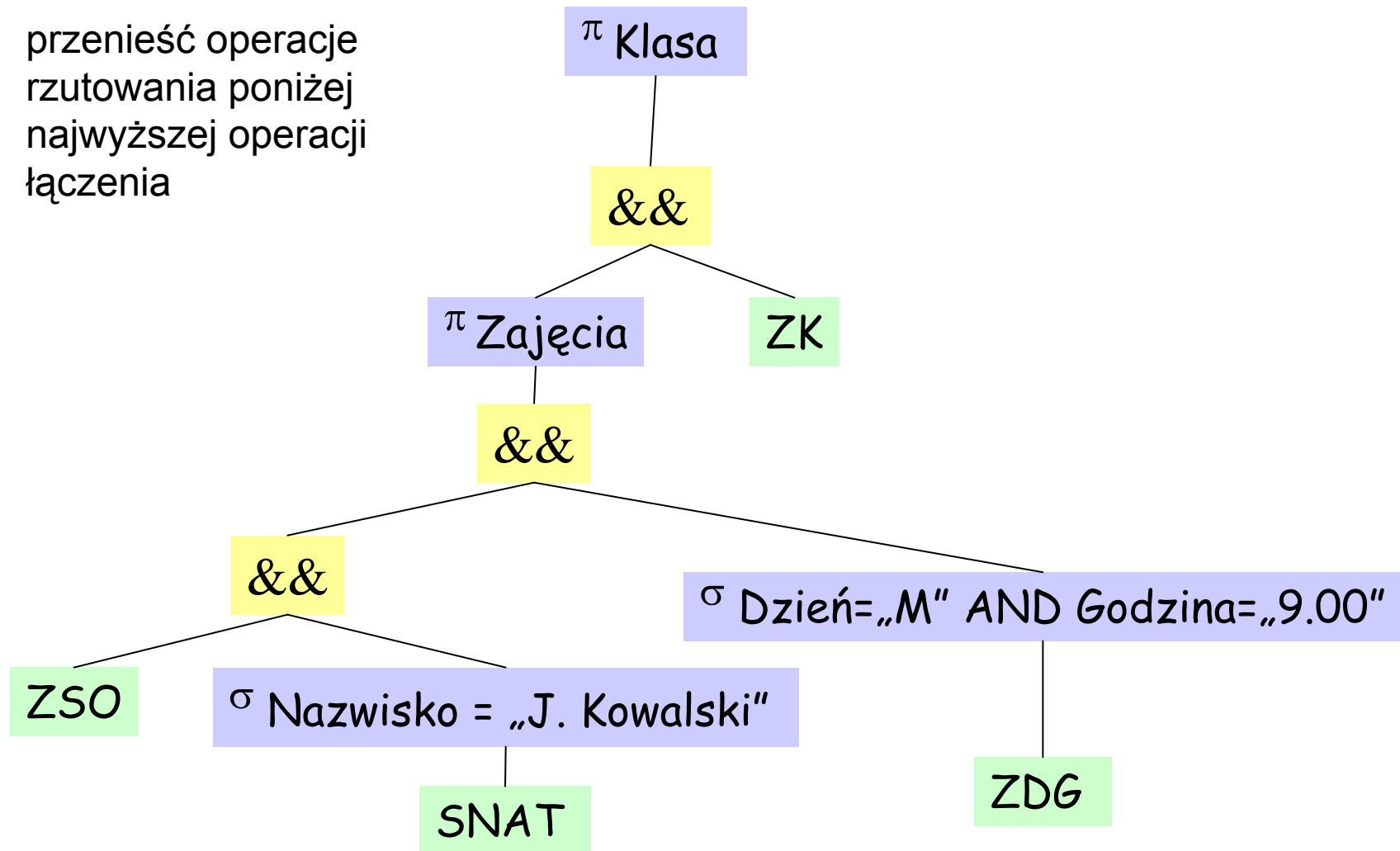
$$(\pi_L(R \underset{A=B}{\&\&} S)) = (\pi_L(\pi_M(R) \underset{A_i=B_j}{\&\&} \pi_N(S)))$$

M – lista atrybutów z listy L, które należą do schematu relacji R, plus atrybut A, jeśli nie ma go na liście L

N – lista atrybutów z listy L, które należą do schematu relacji R, plus atrybut B, jeśli nie ma go na liście L

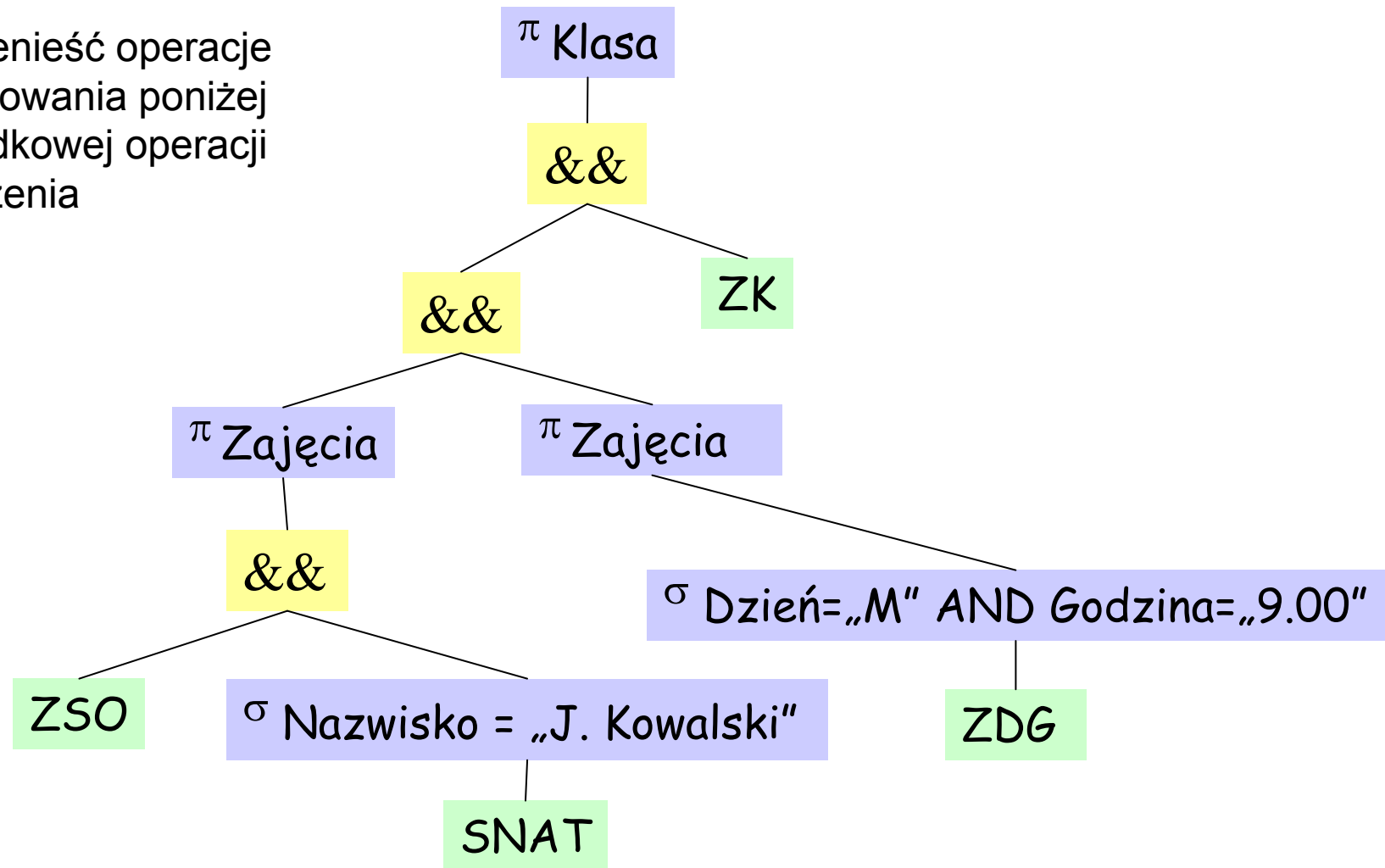
„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

przenieść operacje
rzutowania poniżej
najwyższej operacji
łączenia



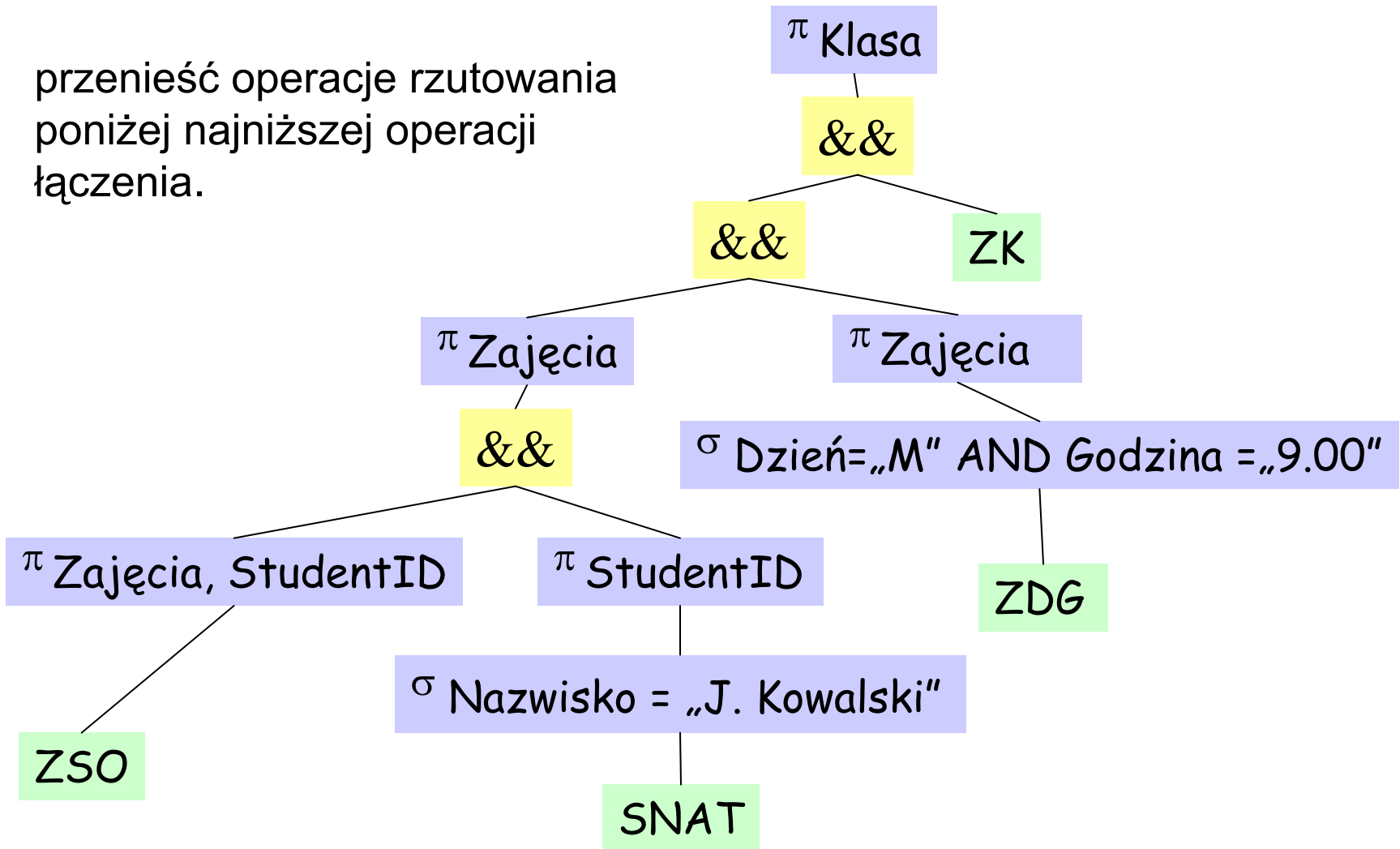
„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

przenieść operacje
rzutowania poniżej
środkowej operacji
łączenia



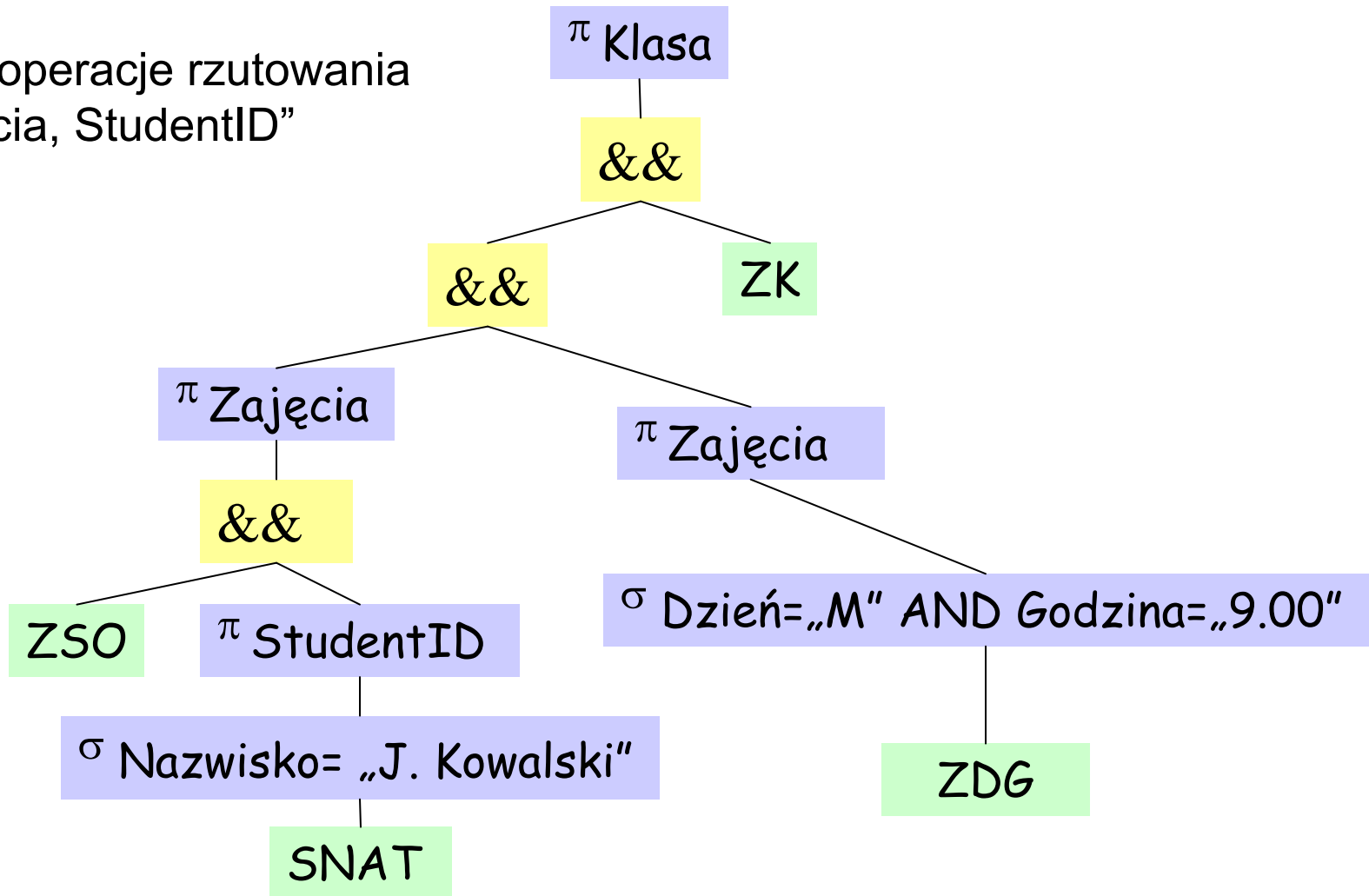
„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

przenieść operacje rzutowania
poniżej najniższej operacji
łączenia.



„Gdzie przebywa J. Kowalski w poniedziałek o 9-tej rano?”

usuń operacje rzutowania
„Zajęcia, StudentID”



Podsumowanie

- ⇒ Algebra relacyjna jest wysoko poziomą notacją definiowania operacji zapytań dotyczących jednej lub wielu relacji.
Głównymi operacjami tej algebry są: suma, przecięcie, różnica, selekcja, rzutowanie i złączenie.
Jest silną notacją wyrażania zapytań bez podawania szczegółów dotyczących planowanych operacji na otrzymanych danych
- ⇒ Istnieje wiele sposobów efektywnego implementowania operacji złączenia
- ⇒ Optymalizacja wyrażeń algebry relacyjnej może w znaczący sposób skrócić czas wyznaczania ich wartości, jest więc istotnym elementem wszystkich języków opartych na algebrze relacyjnej wykorzystywanych w praktyce do wyrażania zapytań;
- ⇒ Istnieje wiele sposobów skracania czasu obliczania danego wyrażenia. Najlepsze efekty przynosi przenoszenie operacji selekcji w dół drzewa wyrażenia.