

Wydział Zarządzania AGH

Katedra Informatyki Biznesowej i Inżynierii Zarządzania



Normalizacja relacji

Narzędzia informatyczne
w zarządzaniu

Program wykładu

- ❑ Normalizacja
- ❑ Pierwsza postać normalna
- ❑ Druga postać normalna
- ❑ Klucze
- ❑ Przykłady

Relacyjny model danych RMD

Oparty na relacjach; relacja jest tabelą spełniającą warunki:

- każda relacja w bazie ma jednoznaczną nazwę,
- każda kolumna w relacji ma jednoznaczną nazwę,
- wszystkie wartości w kolumnie są tego samego typu,
- porządek kolumn w relacji nie jest istotny,
- każdy wiersz w relacji musi być różny,
- porządek wierszy nie jest istotny,
- każde pole leżące na przecięciu kolumny i wiersza powinno mieć wartość atomową.

Relacyjny model danych RMD

- ❑ jeden-do-jednego: jeżeli pojedynczemu rekordowi z pierwszej tabeli przyporządkowany jest najwyżej jeden rekord z drugiej tabeli i na odwrót,
- ❑ jeden-do-wielu: jeżeli pojedynczemu rekordowi z pierwszej tabeli może odpowiadać jeden lub więcej rekordów z drugiej, ale pojedynczemu rekordowi z drugiej tabeli odpowiada najwyżej jeden rekord z tabeli pierwszej,
- ❑ wiele-do-wielu: jeżeli pojedynczemu rekordowi z pierwszej tabeli może odpowiadać jeden lub więcej rekordów z drugiej i na odwrót.

Relacyjny model danych - normalizacja

- ❑ **Normalizacja** (sprowadzanie do postaci normalnej) to proces projektowania relacyjnej bazy danych tak, aby utworzyć zbiór tabel o odpowiedniej strukturze; polega na kolejnych podziałach tabel na mniejsze tabele, które są z kolei łączone w zapytaniach.
- ❑ **Cele** normalizacji:
 1. minimalizacja redundancji,
 2. uniknięcie anomalii przy modyfikacji, wstawianiu i usuwaniu,
 3. uniknięcie niespójności.
- ❑ **Koszt**: zwiększenie liczby tabel, a tym samym wydłużenie czasu dostępu do danych.

RMD - normalizacja

- Jedynymi relacjami dozwolonymi w modelu relacyjnym są relacje spełniające następujący warunek:
 - każda wartość w relacji, tj. każda wartość atrybutu w każdej krotce, jest wartością atomową (wartością nierozkładalną),
 - 1PN oznacza, że tabela nie zawiera powtarzających się grup informacji, co znaczy, że każda kolumna jest wartością skalarną (atomową), a nie macierzą lub listą czy też czymkolwiek, co posiada własną strukturę.

Sklep	Towar
Żabka Brzezine	chleb 100szt; masło 50 kg
Lewiatan Ujazd	chleb 105 sztuk; masło 60 kg

RMD - normalizacja

Sklep	Zapas	
	Towar	Ilość
Żabka Brzezcie	chleb	100
	masło	50
Lewiatan Ujazd	chleb	105
	masło	60

- ❑ Relację "przed normalizacją" zdefiniowano na dwóch dziedzinach: **Sklep** i **Zapas**
- ❑ Elementami dziedziny **Zapas** są również relacje (zdefiniowane na dziedzinach **Towar** i **Ilość**)
- ❑ Relacja jest z punktu widzenia definicji relacją dwuczłonową, ale nie wszystkie jej dziedziny są proste (dziedzina prosta to taka, której wszystkie elementy są atomowe)

RMD - normalizacja

Tabela (relacja) znormalizowana

Relacja jest relacją trójczłonową, której wszystkie dziedziny są proste, jest więc znormalizowana.

Sklep	TowarwSklepie	IlośćwSklepie
Żabka Brzezcie	chleb	100
Żabka Brzezcie	masło	50
Lewiatan Ujazd	chleb	105
Lewiatan Ujazd	masło	60

Anomalie przy usuwaniu, wstawianiu i aktualizacji

Sklep	TowarwSklepie	Jm	IlośćwSklepie
Żabka Brzezcie	chleb	szt.	100
Żabka Brzezcie	masło	kg	50
Żabka Brzezcie	sok	litry	1000
Lewiatan Ujazd	chleb	szt.	105
Lewiatan Ujazd	masło	kg	60

- ❑ Usunięcie zaznaczonego rekordu spowoduje utratę informacji o tym, że sok sprzedaje się w litrach (anomalie przy usuwaniu).
- ❑ Nie można wprowadzić informacji, że mleko sprzedaje się w kartonach, póki któryś ze sklepów nie zacznie go sprzedawać (anomalie przy wstawianiu).
- ❑ Jeżeli zaczniemy sprzedawać sok w butelkach, to pojawią się zapisy dotyczące tego samego towaru z różnymi jednostkami miary (anomalie przy aktualizacji).

RMD – 1PN, nieporozumienia

Pracownik	Imię dziecka	Data ur. dziecka
Kowalski	Ania	01.01.2019
	Jaś	15.03.2011
Kot	Patrycja	20.10.2011
	Filemon	30.07.2013

Pracownik	Imię dziecka1	Data ur. dziecka1	Imię dziecka2	Data ur. dziecka2
Kowalski	Ania	01.01.2019	Jaś	15.03.2011
Kot	Patrycja	20.10.2011	Filemon	30.07.2013

Relacyjny model danych - klucz

- ❑ Minimalny zbiór identyfikujący.
- ❑ Taki zbiór atrybutów relacji, których kombinacje wartości jednoznacznie identyfikują każdą krotkę tej relacji, a żaden podzbiór tego zbioru nie posiada tej własności.
- ❑ W kluczu nie może zawierać się wartość *Null*.
- ❑ Klucz jest **kluczem prostym**, jeżeli powyżej opisany zbiór jest jednoelementowy; w przeciwnym razie mówimy o **kluczu złożonym**.
- ❑ W ogólności, w relacji można wyróżnić wiele kluczy, które nazywamy **kluczami potencjalnymi**. Wybrany klucz spośród kluczy potencjalnych nazywamy **kluczem głównym (Primary Key PK)**.
- ❑ Wiąż to powiązanie pomiędzy parą tabel. Istnieje ona wtedy, gdy dwie tabele są połączone przez klucz własny i klucz obcy.

Normalizacja - przykład

Wyniki zaliczeń i egzaminów

Rodzaj studiów	Rok studiów	Rok akadem.	Moduł	Prowadzący	Forma	Termin	Student		Ocena	Data
				Imię i nazwisko			Nr albumu	Nazwiskoi imię		
stacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	12345	Lis Maria	2,0	2016-06-22
stacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	12411	Kowal Jerzy	4,0	2016-06-22
stacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	12730	Nowak Ewa	2,0	2016-06-22
niestacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	12752	Bury Jan	4,5	2016-06-22
niestacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	12753	Kowal Jerzy	5,0	2016-06-22
stacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	13134	Mączka Karol	2,0	2016-06-22
stacj.	I	2015/16	ISZ	prof. Adam Kat	egzamin	I	13432	Pidek Maria	3,0	2016-06-22
stacj.	I	2015/16	ISZ	dr Ewa Nowak	zaliczenie	I	13653	Horosz Teresa	3,0	2016-06-19
stacj.	I	2015/16	ISZ	dr Ewa Nowak	zaliczenie	II	12345	Lis Maria	3,5	2016-06-19
niestacj.	I	2015/16	ISZ	dr Ewa Nowak	zaliczenie	II	14262	Krymski Olaf	4,0	2016-06-19
niestacj.	I	2015/16	ISZ	dr Ewa Nowak	zaliczenie	I	14661	Rzecka Kamila	4,5	2016-06-19
stacj.	I	2015/16	WF	mgr Jan Bury	zaliczenie	I	12411	Kowal Jerzy	2,0	2016-06-09
niestacj.	I	2015/16	WF	mgr Jan Bury	zaliczenie	I	14801	Nowak Janina	4,0	2016-06-09
niestacj.	I	2015/16	WF	mgr Jan Bury	zaliczenie	I	15038	Adamski Marek	4,0	2016-06-09
niestacj.	I	2015/16	WF	mgr Jan Bury	zaliczenie	I	15113	Zdeb Anna	4,5	2016-06-09
stacj.	I	2015/16	WF	mgr Jan Bury	zaliczenie	I	12345	Lis Ewa	3,5	2016-06-09
stacj.	II	2014/15	Informatyka	dr inż. Jan Bury	zaliczenie	III	15361	Kolasa Lidia	5,0	2016-09-15

RMD – zależności funkcyjne

- Podczas projektowania schematu bazy danych podstawowym pojęciem jest **zależność danych**.
- Przykładem zależności danych jest zależność atrybutu *student* od atrybutu *nr albumu*.

Należy podkreślić, że w celu wyznaczenia zależności występujących w danej bazie danych powinno się dokładnie rozważyć znaczenie atrybutów. Nie istnieją bowiem metody pozwalające na automatyczne określenie zależności. Wynikają one z wiedzy projektanta o naturze danych.

Student	
Nr albumu	Nazwiskoi imię
12345	Lis Maria
12411	Kowal Jerzy
12730	Nowak Ewa
12752	Bury Jan
12753	Kowal Jerzy
13134	Mączka Karol
13432	Pidek Maria
13653	Horosz Teresa
12345	Lis Maria
14262	Krymski Olaf
14661	Rzecka Kamila
12411	Kowal Jerzy
14801	Lis Maria
15038	Adamski Marek
15113	Zdeb Anna
12345	Lis Maria

RMD – druga postać normalna

- ❑ Relacja jest w **drugiej postaci normalnej**, jeśli każdy atrybut tej relacji niewchodzący w skład żadnego klucza potencjalnego jest w pełni funkcyjnie zależny wyłącznie od wszystkich podrelacji klucza głównego, czyli każdy atrybut, który nie jest kluczem, zależy funkcyjnie od klucza (a nie od podzbioru atrybutów stanowiących klucz).
- ❑ Jak łatwo zauważyć, nasza przykładowa relacja nie jest w drugiej postaci normalnej. **W tabeli jest to widoczne poprzez powtarzające się pary, trójki, itd. wartości atrybutów.**
nr albumu -> nazwisko i imię
Rodzaj studiów, Rok studiów, Rok akademicki, Moduł,
Forma -> Prowadzący
- ❑ Aby przekształcić naszą relację do drugiej postaci normalnej należy doprowadzić ją do 1PN oraz dokonać jej projekcji na trzy tabele.

RMD – przykład 2PN

Wyniki

Rodzaj studiów	Rok studiów	Rok akadem.	Moduł	Forma	Termin	Student	Ocena	Data
						Nr albumu		
stacj.	I	2015/16	ISZ	egzamin	I	12345	2,0	2016-06-22
stacj.	I	2015/16	ISZ	egzamin	I	12411	4,0	2016-06-22
stacj.	I	2015/16	ISZ	egzamin	I	12730	2,0	2016-06-22
niestacj.	I	2015/16	ISZ	egzamin	I	12752	4,5	2016-06-22
niestacj.	I	2015/16	ISZ	egzamin	I	12753	5,0	2016-06-22
stacj.	I	2015/16	ISZ	egzamin	I	13134	2,0	2016-06-22
stacj.	I	2015/16	ISZ	egzamin	I	13432	3,0	2016-06-22
stacj.	I	2015/16	ISZ	zaliczenie	I	13653	3,0	2016-06-19
stacj.	I	2015/16	ISZ	zaliczenie	II	12345	3,5	2016-06-19
niestacj.	I	2015/16	ISZ	zaliczenie	II	14262	4,0	2016-06-19
niestacj.	I	2015/16	ISZ	zaliczenie	I	14661	4,5	2016-06-19
stacj.	I	2015/16	WF	zaliczenie	I	12411	2,0	2016-06-09
niestacj.	I	2015/16	WF	zaliczenie	I	14801	4,0	2016-06-09
niestacj.	I	2015/16	WF	zaliczenie	I	15038	4,0	2016-06-09
niestacj.	I	2015/16	WF	zaliczenie	I	15113	4,5	2016-06-09
stacj.	I	2015/16	WF	zaliczenie	I	12345	3,5	2016-06-09
stacj.	II	2014/15	Informatyka	zaliczenie	III	15361	5,0	2016-09-15



Prowadzący

Rodzaj studiów	Rok studiów	Rok akadem.	Moduł	Forma	Prowadzący		
					Tytuł	Imię	Nazwisko
stacj.	I	2015/16	ISZ	egzamin	profesor	Adam	Kat
niestacj.	I	2015/16	ISZ	egzamin	profesor	Adam	Kat
stacj.	I	2015/16	ISZ	zaliczenie	doktor	Ewa	Nowak
niestacj.	I	2015/16	ISZ	zaliczenie	doktor	Ewa	Nowak
stacj.	I	2015/16	WF	zaliczenie	magister	Jan	Bury
niestacj.	I	2015/16	WF	zaliczenie	magister	Jan	Bury
stacj.	II	2014/15	Informatyka	zaliczenie	doktor inż.	Jan	Bury



Studenci

Student		
Nr albumu	Nazwisko	Imię
12345	Lis	Maria
12411	Kowal	Jerzy
12730	Nowak	Ewa
12752	Bury	Jan
12753	Kowal	Jerzy
13134	Mączka	Karol
13432	Pidek	Maria
13653	Horosz	Teresa
14262	Krymski	Olaf
14661	Rzecka	Kamila
14801	Lis	Maria
15038	Adamski	Marek
15113	Zdeb	Anna
15361	Kolasa	Lidia

RMD – przykład 2PN

- ❑ **Klucz sztuczny** - klucz stworzony wyłącznie dla potrzeb relacji w celu zastąpienia złożonego klucza głównego.
- ❑ Klucz w tabeli *Prowadzący*
{Rodzaj studiów, Rok studiów, Rok akademicki, Moduł, Forma}
może być zastąpiony sztucznym kluczem:
Moduły
- ❑ Uwzględniając powyższe dostajemy następujące trzy tabele.

RMD – przykład 2PN

Wyniki

Moduły	Termin	Student	Ocena	Data
		Nr albumu		
1	I	12345	2,0	2016-06-22
1	I	12411	4,0	2016-06-22
1	I	12730	2,0	2016-06-22
2	I	12752	4,5	2016-06-22
2	I	12753	5,0	2016-06-22
1	I	13134	2,0	2016-06-22
1	I	13432	3,0	2016-06-22
3	I	13653	3,0	2016-06-19
3	II	12345	3,5	2016-06-19
4	II	14262	4,0	2016-06-19
4	I	14661	4,5	2016-06-19
5	I	12411	2,0	2016-06-09
6	I	14801	4,0	2016-06-09
6	I	15038	4,0	2016-06-09
6	I	15113	4,5	2016-06-09
5	I	12345	3,5	2016-06-09
7	III	15361	5,0	2016-09-15

Studenci

Student		
Nr albumu	Nazwisko	Imię
12345	Lis	Maria
12411	Kowal	Jerzy
12730	Nowak	Ewa
12752	Bury	Jan
12753	Kowal	Jerzy
13134	Mączka	Karol
13432	Pidek	Maria
13653	Horosz	Teresa
14262	Krymski	Olaf
14661	Rzecka	Kamila
14801	Lis	Maria
15038	Adamski	Marek
15113	Zdeb	Anna
15361	Kolasa	Lidia

Prowadzący

Moduły	Rodzaj studiów	Rok studiów	Rok akadem.	Moduł	Forma	Prowadzący		
						Tytuł	Imię	Nazwisko
1	stacj.	I	2015/16	ISZ	egzamin	profesor	Adam	Kat
2	niestacj.	I	2015/16	ISZ	egzamin	profesor	Adam	Kat
3	stacj.	I	2015/16	ISZ	zaliczenie	doktor	Ewa	Nowak
4	niestacj.	I	2015/16	ISZ	zaliczenie	doktor	Ewa	Nowak
5	stacj.	I	2015/16	WF	zaliczenie	magister	Jan	Bury
6	niestacj.	I	2015/16	WF	zaliczenie	magister	Jan	Bury
7	stacj.	II	2014/15	Informatyka	zaliczenie	doktor inż.	Jan	Bury

RMD – klucz sztuczny

- ❑ **Klucz sztuczny** może być wykorzystany do kodowania atrybutów tekstowych (w niektórych przypadkach także liczbowych; **z tym – ostrożnie!**) o powtarzających się wartościach, dla których można utworzyć listę.
- ❑ I tak:
 - Rodzaj studiów* można zastąpić kluczem *Id_rodzaju*
 - Moduł* można zastąpić kluczem *Id_modułu*
 - Forma* można zastąpić kluczem *Id_forma*
 - Tytuł* można zastąpić kluczem *Id_tytułu*
- ❑ Użycie klucza sztucznego wymaga stworzenia dodatkowej tabeli (słownika) pozwalającego na „rozkodowanie” klucza.

RMD – klucz sztuczny

Prowadzący

Moduły	Id_rodzaju	Rok studiów	Rok akadem.	Id_modułu	Id_formy	Prowadzący		
						Id_tytułu	Imię	Nazwisko
1	1	I	2015/16	1	1	1	Adam	Kat
2	2	I	2015/16	1	1	1	Adam	Kat
3	1	I	2015/16	1	2	3	Ewa	Nowak
4	2	I	2015/16	1	2	3	Ewa	Nowak
5	1	I	2015/16	2	2	4	Jan	Bury
6	2	I	2015/16	2	2	4	Jan	Bury
7	1	II	2014/15	3	2	2	Jan	Bury

Rodzaje studiów

Id_rodzaju	Rodzaj
1	stacj.
2	niestacj.

Moduł

Id_modułu	Nazwa
1	ISZ
2	WF
3	Informatyka

Formy

Id_formy	Nazwa
1	egzamin
2	zaliczenie

Tytuły

Id_tytułu	Nazwa
1	profesor
2	doktor inż.
3	doktor
4	magister

RMD – powtarzająca się grupa

- ❑ Powtarzająca się grupa danych to podzbiór relacji zawierający co najmniej dwa atrybuty, posiadająca własny klucz prosty, w którym istnieją powtarzające się krotki.
- ❑ Powtarzanie się takich samych krotek wymuszone jest faktem, że mamy do czynienia z grupą, dla której część atrybutów jest strukturą, a nie wartością skalarną.
- ❑ W tabeli **Prowadzący** wielokrotnie powtarza się grupa atrybutów:
Id_rodzaju, Rok studiów, Rok akademicki
oraz
Id_tytułu, Imię, Nazwisko (prowadzącego)
- ❑ Grupy te można zastąpić kluczem sztucznym i stworzyć dodatkowe tabele grupujące powtarzające się atrybuty.

RMD – powtarzająca się grupa

Prowadzący

Moduły	Id studiów	Id modułu	Id formy	Id prac
1	1	1	1	1
2	2	1	1	1
3	1	1	2	3
4	2	1	2	3
5	1	2	2	4
6	2	2	2	4
7	3	3	2	2

Pracownicy

Id prac	Id tytułu	Imię	Nazwisko
1	1	Adam	Kat
2	3	Ewa	Nowak
3	4	Jan	Bury
4	2	Jan	Bury

Tytuły

Id tytułu	Nazwa
1	profesor
2	doktor inż.
3	doktor
4	magister

Lata studiów

Id studiów	Id rodzaju	Rok studiów	Rok akad.
1	1	I	2015/16
2	2	I	2015/16
3	1	II	2014/15

Moduł

Id modułu	Nazwa
1	ISZ
2	WF
3	Informatyka

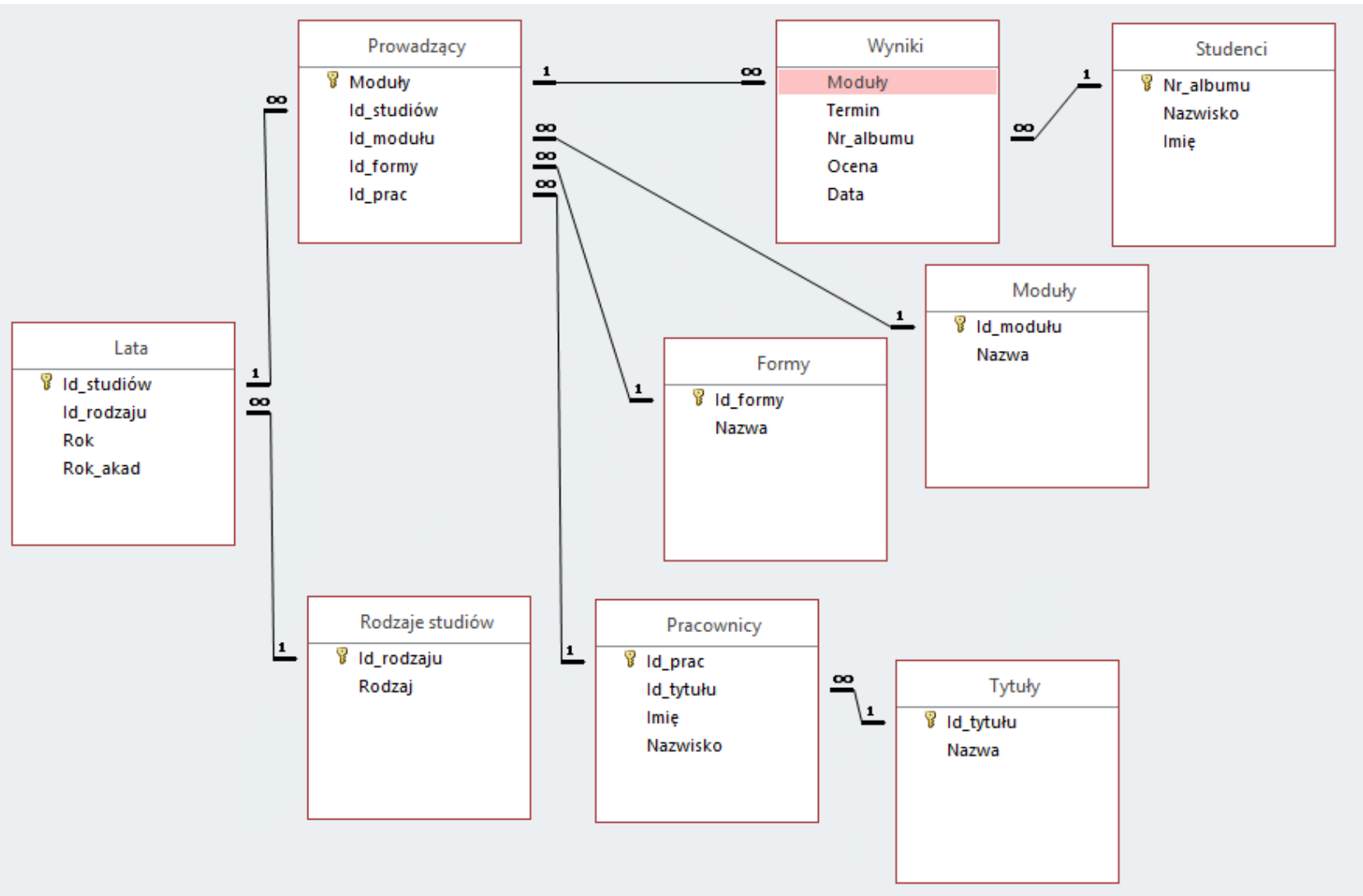
Rodzaje studiów

Id rodzaju	Rodzaj
1	stacj.
2	niestacj.

Formy

Id formy	Nazwa
1	egzamin
2	zaliczenie

RMD – rozwiązanie

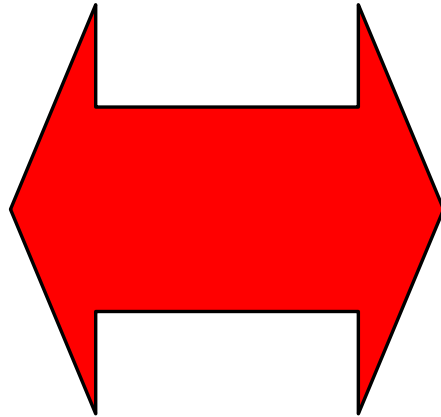


RMD – przykłady

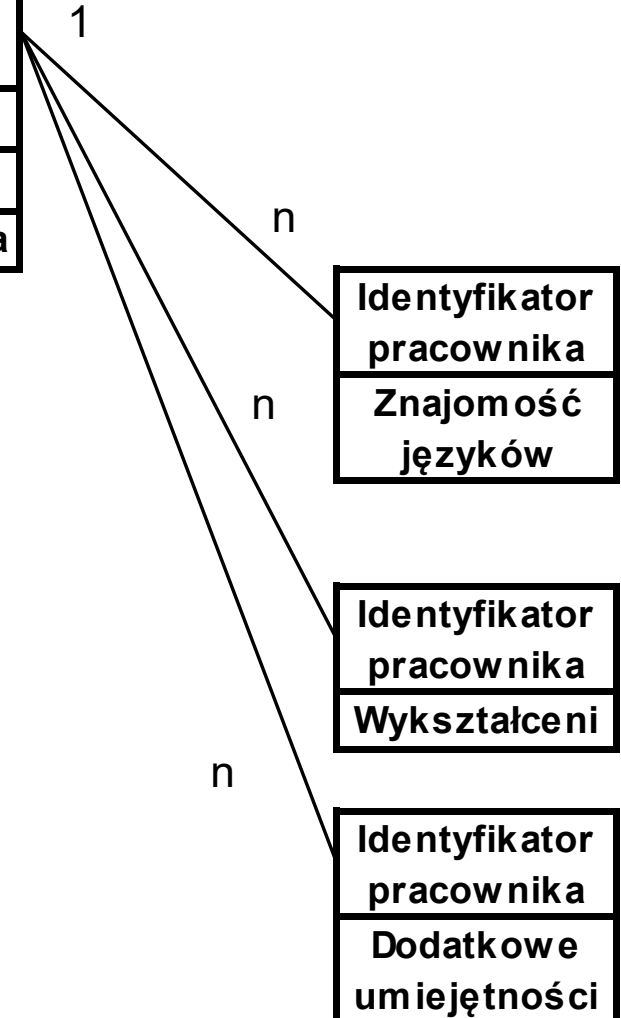
Nr Pesel	Identyfikator pracownika	Imię	Nazwisko	Data urodzeni	Znajomość języków	Wykształcenie	Dodatkowe umiejętności
56031608123	121/97	Anna	Lis	16.03.56	angielski	wyższe	prawo jazdy
56031608123	121/97	Anna	Lis	16.03.56	niemiecki		
56031608123	121/97	Anna	Lis	16.03.56	francuski		
77101012345	134/21	Anna	Kozłowska	10.10.77	niemiecki	średnie techn.	biegły księgowy
77101012345	134/21	Anna	Kozłowska	10.10.77	francuski	wyższe ekon.	
77101012345	134/21	Anna	Kozłowska	10.10.77	rosyjski		
65021467843	98/85	Adam	Lewiński	14.02.65	niemiecki	średnie	prawo jazdy
65021467843	98/85	Adam	Lewiński	14.02.65	niemiecki		biegły księgowy
66010181321	112/90	Adam	Lewiński	01.01.66	rosyjski	wyższe	prawo jazdy
66010181321	112/90	Adam	Lewiński	01.01.66	rosyjski		biegły księgowy
60121286541	115/90	Adam	Osowski	15.12.60	rosyjski	wyższe	prawo jazdy
60121286541	115/90	Adam	Osowski	15.12.60	rosyjski	dr ekon.	

RMD – przykłady

Nr Pesel
Identyfikator pracownika
Imię
Nazwisko
Data
Znajomość języków
Wykształcenie
Dodatkowe umiejętności



Nr Pesel
Identyfikator pracownika
Imię
Nazwisko
Data urodzenia



RMD – przykłady

Nr Pesel	Identyfikator pracownika	Imię	Nazwisko	Data urodzenia
56031608123	121/97	Anna	Lis	16.03.56
77101012345	134/21	Anna	Kozłowska	10.10.77
65021467843	98/85	Adam	Lewiński	14.02.65
66010181321	112/90	Adam	Lewiński	01.01.66
60121286541	115/90	Adam	Osowski	15.12.60

Identyfikator pracownika	Znajomość języków
121/97	angielski
121/97	niemiecki
121/97	francuski
134/21	niemiecki
134/22	francuski
134/23	rosyjski
98/85	niemiecki
112/90	rosyjski
115/90	rosyjski

Identyfikator pracownika	Wykształcenie
121/97	wyższe
134/21	średnie techn.
134/21	wyższe ekon.
98/85	średnie
112/90	wyższe
115/90	wyższe
115/90	dr ekon.

Identyfikator pracownika	Dodatkowe umiejętności
121/97	prawo jazdy
134/21	biegły księgowy
98/85	prawo jazdy
98/85	biegły księgowy
112/90	prawo jazdy
112/90	biegły księgowy
115/90	prawo jazdy
98/85	kurs wózków