**Tydzień 4. 24.09.2019 (45 min.)**

Temat: Normalizowanie baz danych

Normalizacja polega na rozmieszczaniu danych w mniejszych tablicach w celu wyeliminowania nadmiarowości danych (redundancji) i uniknięcia problemów związanych z przetwarzaniem danych.

Przykład nieunormowanej tablicy

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **bajkowo** | | | | | | |
| **nr** | **uczen** | **klasa** | **specjalizacja** | **adres** | **nauczyciel** | **email\_naucz** |
| 1 | Grafi | 1a | sok z gumijagód | Gumisiowa Dolina | Bunia | bunia@gmail.com |
| 2 | Tami | 1b | sok z gumijagód | Gumisiowa Dolina | Zami | zami@gmail.com |
| 3 | Sani | 1b | sok z gumijagód | Gumisiowa Dolina | Zami | zami@gmail.com |
| 4 | Kabi | 1a | sok z gumijagód | Gumisiowa Dolina | Bunia | bunia@gmail.com |
| 5 | Marshall | 2a | ratownictwo | Adventure Bay | Ryder | ryder@gmail.com |
| 6 | Rubble | 2a | ratownictwo | Adventure Bay | Ryder | ryder@gmail.com |
| 7 | Skye | 2b | ratownictwo | Adventure Bay | Pani Goodway | panig@gmail.com |
| 8 | Zuma | 2b | ratownictwo | Adventure Bay | Pani Goodway | panig@gmail.com |
| 10 | Muminek | 3a | wędrówki | Dolina Muminków | Włóczykij | wloczykij@gmail.com |
| 11 | Mała Mi | 3a | wędrówki | Dolina Muminków | Włóczykij | wloczykij@gmail.com |
| 12 | Ryjek | 3a | wędrówki | Dolina Muminków | Włóczykij | wloczykij@gmail.com |

Problemy mogące wynikać z nieunormowanych tablic:

* Zwiększenie rozmiaru bazy danych
* Dodawanie informacji powoduje wielokrotne powtarzanie tych samych danych, może prowadzić do błędów. Nie jest możliwe dodanie wybranych danych (np. samego wychowawcy).
* Usuwanie jednych informacji może doprowadzić do utraty innych informacji (np. usuwanie informacji o uczniach może doprowadzić do utraty informacji o zawodach albo wychowawcach).
* Modyfikowanie danych wymaga zmiany tych samych informacji w każdym rekordzie.

W celu wyeliminowania powyższych problemów tablicę można podzielić na mniejsze tablice np.

W projektowaniu tabel pomaga tzw proces normalizacji tabel.

Normalizując tabele przechodzimy przez kolejne normalizacje uzyskując najpierw pierwszą postać normalną (1NF), drugą postać normalną (2NF), trzecią postać normalną (3NF), itd.

Pierwsza postać normalna (1NF)

Mówimy, że tabela jest w pierwszej postaci normalnej jeżeli

* Każda kolumna (atrybut) tabeli powinna zawierać elementarne (atomowe, niepodzielne) dane. Nie może być tablicą, listą czy też czymś, co posiada własną strukturę

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | Grafi | sok z gumijagód | Gumisiowa Dolina | Bunia, Zami |
| 2 | Sani | sok z gumijagód | Gumisiowa Dolina | Zami, |
| 3 | Kabi | sok z gumijagód | Gumisiowa Dolina | Bunia, Kevin |

Te wartości można podzielić

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | Grafi | sok z gumijagód | Gumisiowa Dolina | Bunia |
| 2 | Grafi | sok z gumijagód | Gumisiowa Dolina | Zami |
| 3 | Sani | sok z gumijagód | Gumisiowa Dolina | Zami |
| 4 | Kabi | sok z gumijagód | Gumisiowa Dolina | Bunia |
| 5 | Kabi | sok z gumijagód | Gumisiowa Dolina | Kevin |

* Kolejność wierszy nie może mieć znaczenia.
* Kolejność kolumn nie może mieć znaczenia (każda kolumna powinna mieć unikatową nazwę).
* Nie może zawierać kolekcji (powtarzających się grup informacji).

!

Druga postać normalna (2NF)

Ta postać określa podstawa dobrego projektowania bazy. Mówi o tym, że **każda tabela powinna przechowywać dane dotyczące tylko konkretnej klasy obiektów**.

Tabela jest w drugiej postaci normalnej jeżeli

* Jest w 1NF
* Wszystkie kolumny zależą tylko od klucza głównego

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **bajkowo** | | | | | | |
| **nr** | **uczen** | **klasa** | **specjalizacja** | **adres** | **nauczyciel** | **email\_naucz** |
| 1 | Grafi | 1a | sok z gumijagód | Gumisiowa Dolina | Bunia | bunia@gmail.com |
| 2 | Tami | 1b | sok z gumijagód | Gumisiowa Dolina | Zami | zami@gmail.com |
| 3 | Sani | 1b | sok z gumijagód | Gumisiowa Dolina | Zami | zami@gmail.com |
| 4 | Kabi | 1a | sok z gumijagód | Gumisiowa Dolina | Bunia | bunia@gmail.com |
| 5 | Marshall | 2a | ratownictwo | Adventure Bay | Ryder | ryder@gmail.com |
| 6 | Rubble | 2a | ratownictwo | Adventure Bay | Ryder | ryder@gmail.com |
| 7 | Skye | 2b | ratownictwo | Adventure Bay | Pani Goodway | panig@gmail.com |
| 8 | Zuma | 2b | ratownictwo | Adventure Bay | Pani Goodway | panig@gmail.com |
| 10 | Muminek | 3a | wędrówki | Dolina Muminków | Włóczykij | wloczykij@gmail.com |
| 11 | Mała Mi | 3a | wędrówki | Dolina Muminków | Włóczykij | wloczykij@gmail.com |
| 12 | Ryjek | 3a | wędrówki | Dolina Muminków | Włóczykij | wloczykij@gmail.com |

|  |  |  |
| --- | --- | --- |
| **uczen** | | |
| id\_ucznia | nazwa | klasa |
| 1 | Grafi | 1a |
| 2 | Tami | 1b |
| 3 | Sani | 1b |
| 4 | Kabi | 1a |
| **specjalizacje** | | | |
| id\_specjal | nazwa | adres | |
| 1 | sok z gumijagód | Gumisiowa Dolina | |
| 2 | ratownictwo | Adventure Bay | |
| 3 | wędrówki | Dolina Muminków | |

|  |  |  |
| --- | --- | --- |
| **nauczyciele** | | |
| id\_naucz | nazwa | email |
| 1 | Bunia | bunia@gmail.com |
| 2 | Ryder | ryder@gmail.com |
| 3 | Włóczykij | wloczykij@gmail.com |
| 4 | Zami | zami@gmail.com |
| 5 | Pani Goodway | panig@gmail.com |

|  |  |
| --- | --- |
| **kwalifikacje** | |
| id\_naucz | id\_specjal |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 1 |
| 5 | 2 |

|  |  |
| --- | --- |
| **klasy** | |
| nazwa | id\_naucz |
| 1a | 1 |
| 1b | 4 |
| 2a | 2 |
| 2b | 5 |
| 3a | 3 |

Trzecia postać normalna 3NF

Tabela jest w 3NF jeżeli

* jest w 2NF
* żadna kolumna kolumny nie należąca do klucza głównego nie zależy od innej kolumny nie będącej kluczem głównym.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | pracownik | data zatrudn. | stanowisko | stawka |
| 1 | Severus Snape | 1980-09-01 | nauczyciel | 2400 |
| 2 | Minerwa McGonagall | 1980-09-01 | nauczyciel | 2400 |
| 3 | Argus Filch | 1980-09-01 | woźny | 1800 |
| 4 | Rubeus Hagrid | 1980-09-01 | gajowy | 2200 |
| 5 | Albus Dumbledore | 1950-09-01 | dyrektor | 5000 |

W powyższym przykładzie można wprowadzić oddzielną tabelę z danymi dotyczącymi stanowisk.

W praktyce nie zawsze warto stosować 3NF. Np. Kolumna z kodem pocztowym zależy od adresu jednak nie koniecznie musimy wprowadzać oddzielną tabelę z kodami. Niekiedy lepszym rozwiązaniem jest stosowanie trzeciej postaci normalnej tylko do danych często zmienianych.

Kolejne postacie normalne

Kolejne postacie normalne (czwarta postać normalna, nazywana też postacią BCNF Boyce Codd Normal Form, piąta postać normalna) istnieją, ale rzadko są brane pod uwagę w rzeczywistych projektach.