

Instrukcja laboratoryjna 1	Bazy Danych
	Temat: Normalizacje
	Przygotował: mgr inż. Maciej Lasota

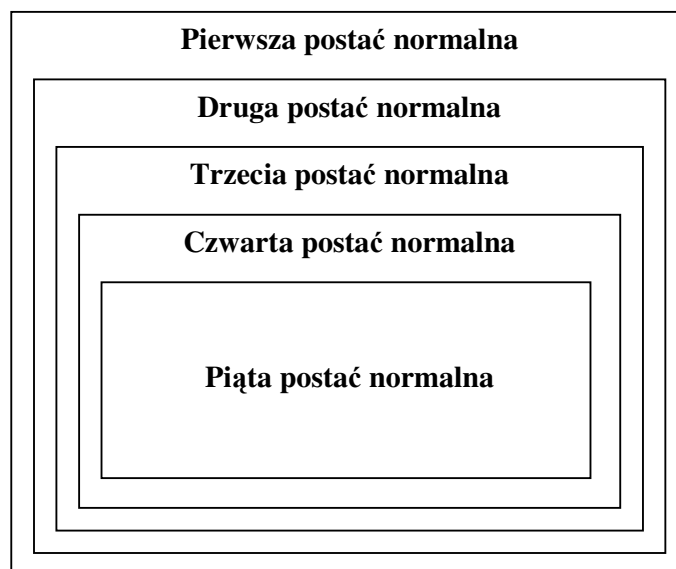
1) Wprowadzenie.

Normalizacja to proces organizacji danych w bazie danych. Polega on na tworzeniu tabel i ustanawianiu pomiędzy nimi powiązań według reguł obowiązujących zarówno przy ochronie danych, jak i uelastycznianiu bazy danych przez eliminowanie powtarzających się i niespójnych zależności. Powtarzające się dane niepotrzebnie zajmują miejsce na dysku i są przyczyną powstawania problemów z obsługą.

Dwa główne etapy normalizacji to ustalenie zbioru danych, które mają być przechowywane w bazie danych i przekształcenie ich do kolejnych postaci normalnych. Zebranie danych polega na określeniu wszystkich danych, jakie mają być przechowywane w bazie danych. Dane te należy zebrać w postaci jednej tabeli, która jest następnie dzielona na mniejsze tabele przy pomocy operacji rzutu, z zachowaniem związków pomiędzy danymi.

Pierwsza, druga i trzecia postać normalna dotyczą zależności funkcyjnych pomiędzy danymi, postać czwarta i piąta dotyczą zależności niefunkcyjnych.

Postać nieznormalizowana



Pierwsza postać normalna (1PN)

Pierwsza postać normalna dotyczy powtarzających się grup danych, jeżeli zależności funkcyjne pomiędzy danymi wskazują na związki jeden do wielu, wskazuje to na istnienie grup danych. Definicja pierwszej postaci normalnej:

Relacja jest w pierwszej postaci normalnej wtedy i tylko wtedy gdy każdy atrybut niekluczowy jest funkcyjnie zależny od klucza głównego. Pierwsza postać normalna wymaga aby żaden atrybut (kolumna) w tabeli nie był podzielny dalej, a także aby nie było powtarzających się grup.

Druga postać normalna (2PN)

Druga postać normalna dotyczy zależności funkcyjnych od części klucza złożonego, normalizacja do drugiej postaci ma na celu wykluczenie takich zależności. Definicja drugiej postaci normalnej :

Relacja jest w drugiej postaci normalnej wtedy i tylko wtedy, gdy jest w pierwszej postaci normalnej i każdy atrybut niekluczowy, czyli nie należący do zadanego klucza, jest w pełni funkcyjnie zależny od klucza głównego. Druga postać normalna mówi, że żadna informacja w wierszu nie może zależeć tylko od części klucza podstawowego (głównego).

Trzecia postać normalna (3PN)

Trzecia postać normalna ma na celu wykluczenie zależności przechodnich pomiędzy danymi poprzez określenie danych determinujących spośród niekluczowych atrybutów relacji. Definicja trzeciej postaci normalnej :

Relacja jest w trzeciej postaci normalnej wtedy i tylko wtedy, gdy jest w drugiej postaci normalnej i każdy niekluczowy atrybut jest bezpośrednio, a nie przechodnio, zależny od klucza głównego. Trzecia postać mówi, że żadna informacja w kolumnie, która nie jest kluczem podstawowym, nie może zależeć od niczego innego, jak tylko od klucza podstawowego.

Postać normalna Boyce'a-Codda (BCNF)

Definicja postaci normalnej Boyce'a-Codda stanowi właściwie sumę pierwszej, drugiej i trzeciej postaci normalnej. Posługując się wyłącznie tą definicją można stwierdzić, czy relacja spełnia warunki relacyjnego modelu danych.

Relacja jest w postaci normalnej Boyce'a-Codda wtedy i tylko wtedy, kiedy każdy jej atrybut zależy funkcjonalnie tylko od jej klucza głównego.

2) Przykłady normalizacji.

Postać nieznormalizowana:

Tabela student					
Nr_Studenta	Opiekun	Pokój	Klasa_1	Klasa_2	Klasa_3
1022	Nowak	412	101-07	143-01	159-02
4123	Kowalski	216	201-01	211-02	214-01

Pierwsza postać normalna:

Brak powtarzających się grup, tabele powinny mieć tylko dwa wymiary. Ponieważ jeden student ma kilka klas, klasy powinny znajdować się w oddzielnej tabeli. Występowanie pól Klasa_1, Klasa_2 i Klasa_3 w powyższych rekordach jest oznaką problemów podczas projektowania.

Tabela student			
Nr_Studenta (PK)	Opiekun	Pokój	Nr_Klasy
1022	Nowak	412	101-07
1022	Nowak	412	143-01
1022	Nowak	412	159-02
4123	Kowalski	216	201-01
4123	Kowalski	216	211-02
4123	Kowalski	216	214-01

Druga postać normalna:

Eliminowanie powtarzających się danych. W powyższej tabeli dla każdego pola Nr_Studenta istnieje wiele wartości w polach Nr_Klasy. Pole Nr_Klasy nie jest zależne od pola Nr_Studenta (klucz podstawowy), dlatego ta relacja nie znajduje się w drugiej postaci normalnej. Drugą postać normalną przedstawiono na następujących dwóch tabelach: studenci i rejestracja.

Tabela student		
Nr_Studenta (PK)	Opiekun	Pokój
1022	Nowak	412
4123	Kowalski	216

Tabela rejestracja	
Nr_Studenta	Nr_Klasy (PK)
1022	101-07
1022	143-01
1022	159-02
4123	201-01
4123	211-02
4123	214-01

Trzecia postać normalna:

Eliminowanie danych, które nie zależą od klucza. W ostatnim przykładzie pole Pokój (numer pokoju opiekuna) jest zależne od atrybutu Opiekun. Rozwiązaniem jest przeniesienie tego atrybutu z tabeli Studenci do tabeli Wydział, tak jak to przedstawiono poniżej:

Tabela wydział		
Opiekun	Pokój	Wydział (PK)
Nowak	412	40
Kowalski	216	41

Tabela student	
Nr_Studenta (PK)	Wydział (FK)
1022	40
4123	41

Tabela rejestracja	
Nr_Studenta (FK)	Nr_Klasy(PK)
1022	101-07
1022	143-01
1022	159-02
4123	201-01
4123	211-02
4123	214-01

Bibliografia:

- [1] „Systemy baz danych”; Paul Beynon-Davies; Wydawnictwa Naukowo-Techniczne, Warszawa 1998.
- [2] „Systemy baz danych pełny wykład”, Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom; Wydawnictwa Naukowo-Techniczne, Warszawa 2006.
- [3] „SQL dla każdego”; Rafe Coburg, Wydawnictwa Helion, Gliwice 2000.
- [4] „Bazy danych”; Piotr Gawrysiak; Politechnika Warszawska Zakład Systemów Informacyjnych.