

Bazy danych

dr Rafał Stęgierski,
Politechnika Lubelska

Zamiast wstępu



2

Cyfrowy odcisk palca

- Treści publikowane przez użytkownika w tym posty w mediach społecznościowych, zdjęcia, filmy, komentarze, artykuły na blogach, wpisy na forach.
- Interakcje online takie jak polubienia, udostępnienia, subskrypcje, udział w dyskusjach, czatach, grupach, recenzje i opinie (np. w sklepach internetowych).
- Aktywność transakcyjna obejmująca zakupy online, udział w programach lojalnościowych, korzystanie z usług bankowych przez Internet.
- Dane z kont i profili zawierające informacje podane przy rejestracji (np. e-mail, imię, pseudonim), zdjęcia profilowe, opisy bio, statusy.
- Udział w społecznościach na przykład obecność w serwisach randkowych, zawodowych (LinkedIn), aktywność w grach online, uczestnictwo w wydarzeniach online.

3

Cyfrowy cień

- Metadane w tym data i godzina zrobienia zdjęcia, lokalizacja GPS, dane EXIF w plikach.
- Dane z logów systemowych zawierające historię logowania, adresy IP, czas korzystania z aplikacji i usług.
- Informacje telemetryczne zbierane przez urządzenia i aplikacje lokalizacja ze smartfona, statystyki korzystania z aplikacji, dane o aktywności.
- Ślady transakcyjne takie jak płatności kartą, historia zakupów online, dane z programów lojalnościowych.
- Dane z komunikacji na przykład nagłówki e-maili, rejestry połączeń, wiadomości przesyłane przez komunikatory (nawet szyfrowane).
- Informacje z mediów społecznościowych zbierane pośrednio w tym to, jakie treści oglądasz, ile czasu spędzasz nad konkretnym postem, jakie reklamy klikniesz.
- Dane sensoryczne z urządzeń typu smartwatche, opaski fitness, samochody podłączone do sieci (np. dane o lokalizacji, prędkości, aktywności fizycznej).

4

Rozmiar śladu i cienia

- Cyfrowy ślad: od kilku MB (minimalna aktywność) do setek GB (twórcy treści wideo).
- Cyfrowy cień: od kilkuset MB do wielu TB danych, bo większość generowana jest automatycznie i masowo (np. dane lokalizacyjne, logi).
- Cień jest wielokrotnie większy niż ślad i mówi się nawet, że przeciętny użytkownik Internetu generuje 10–50 razy więcej cienia niż śladu.

Cyfrowy wszechświat

- Dane tworzone przez ludzi takie jak zdjęcia, filmy, posty, wiadomości, dokumenty.
- Dane generowane maszynowo w tym logi systemów IT, transakcje finansowe, monitoring CCTV, czujniki IoT, dane telemetryczne.
- Dane z komunikacji i sieci uwzględniające ruch internetowy, e-maile, wiadomości przesyłane przez aplikacje, rozmowy VoIP.
- Dane replikowane i kopiowane wśród nich kopie zapasowe, cache, mirrory w chmurze, dane w systemach rozproszonych.

6

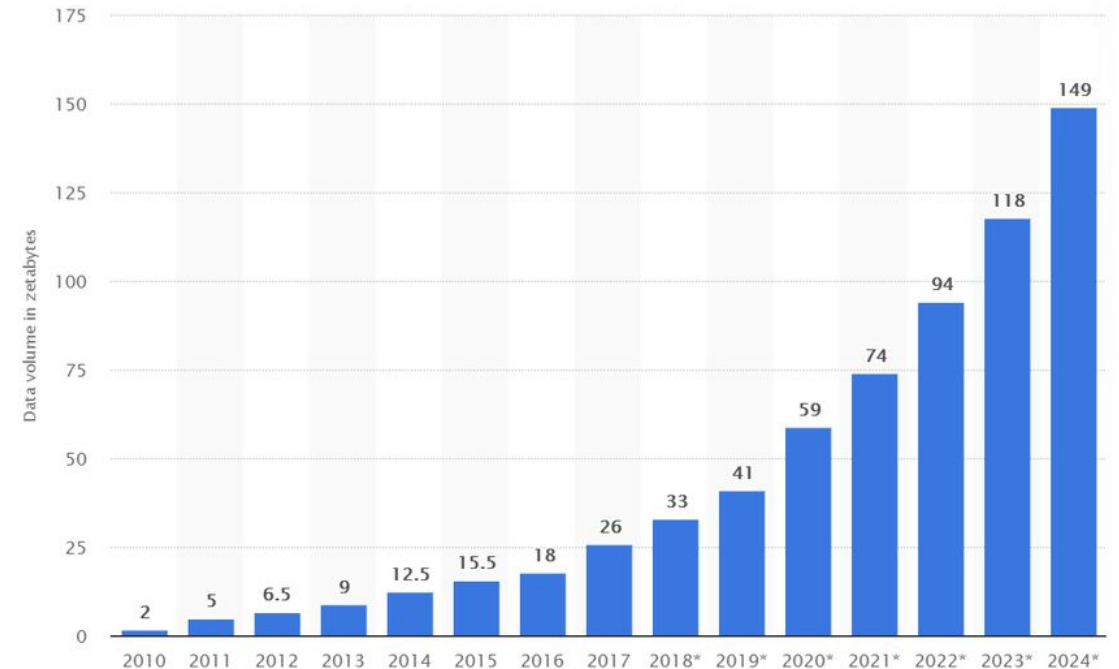
Cyfrowy wszechświat

- Eksplozja danych co oznacza, że większość danych powstaje w ostatnich latach.
- Dominacja danych nieustrukturyzowanych lub semi ustrukturyzowanych takich jak zdjęcia, filmy, dźwięki, teksty stanowią większość.
- Maszynowa produkcja z dużym wpływem rosnącego udziału IoT, AI, autonomicznych systemów, które same generują dane.
- Realna wartość biznesowa, bo jego analiza napędza AI, big data, personalizację usług, reklamę i naukę.

7

Cyfrowy wszechświat

- W 2010 roku szacowano go na ok. 1 zettabajt (10^{21} bajtów).
- W 2020 roku przekroczył 50 zettabajtów.
- Prognozy mówią, że zachowana zostanie podobna szybkość wzrostu



Bazy danych



Dane

- Dane to uporządkowany lub nieuporządkowany zbiór faktów, wartości lub opisów, które mogą być przechowywane, przetwarzane i interpretowane przez człowieka albo maszynę.
- Dane same w sobie nie muszą mieć znaczenia, bo zyskują je dopiero po przetworzeniu i osadzeniu w kontekście (wtedy stają się informacją).

Baza danych

Bazą danych jest zbiór powiązanych przy użyciu ściśle określonego klucza ze sobą dane, które reprezentują wybrany aspekt rzeczywistości, charakteryzujący się pełną koherentnością.

11

Dziedzina problemu

- Wybrany fragment rzeczywistości (aspekt świata), który ma zostać odwzorowany i opisany za pomocą bazy danych.
- W obrębie tej dziedziny określa się obiekty, ich cechy (atrybuty) oraz relacje między nimi.
- Model dziedziny problemu powinien być na tyle elastyczny, aby odzwierciedlać również zmiany zachodzące w rzeczywistości, np. aktualizacje danych czy nowe powiązania.

12

Koherentność bazy

- Baza danych jest logicznie spójnym zbiorem danych.
- Przypadkowe dane nie spełniają tego warunku, czyli nie tworzą bazy danych.
- Innymi słowy: dopiero wtedy, gdy dane są uporządkowane i mają sens w kontekście opisywanej rzeczywistości, można mówić o bazie danych.

Wykorzystanie baz danych

- Nasze codzienne działania mniej lub bardziej świadomie związane są z wykorzystaniem baz danych, poczynając od tych najmniejszych, a kończąc na ogromnych systemach.

Wykorzystanie baz danych

- Średnie bazy danych – obsługują działalność komercyjną o większej skali.
 - Typowym przykładem są systemy w hipermarketach czy dużych sieciach handlowych. Każda transakcja przy kasie powoduje aktualizację bazy danych: zmienia się stan magazynowy, zapisuje się historia zakupu, a system lojalnościowy uzupełnia punkty klienta. Tutaj baza danych musi działać szybko i niezawodnie, bo obsługuje tysiące operacji dziennie.

Wykorzystanie baz danych

- Olbrzymie bazy danych – funkcjonują w systemach państwowych i korporacyjnych, gdzie skala i złożoność są ogromne.
 - Przykładem jest system obsługi zeznań podatkowych, który gromadzi i przetwarza informacje o milionach obywateli i przedsiębiorstw. Dane te muszą być nie tylko spójne, ale także bezpieczne i zgodne z przepisami prawa. Dodatkowo systemy te często integrują się z innymi bazami (np. ZUS, urzędy gmin, banki).

16

Alternatywne zastosowania baz danych

- Bazy multimedialne
 - Gromadzą i udostępniają różne rodzaje danych multimedialnych: obrazy, dźwięki, filmy, nagrania wideo czy grafiki 3D.
 - Wykorzystywane są m.in. w bibliotekach cyfrowych, serwisach streamingowych, systemach zarządzania zasobami medialnymi czy w medycynie (np. bazy zdjęć RTG, MRI).

17

Alternatywne zastosowania baz danych

- Geograficzne systemy informacyjne (GIS – Geographic Information Systems)
 - Przechowują i analizują dane przestrzenne: mapy, informacje o terenie, dane meteorologiczne, hydrologiczne, sejsmiczne, geologiczne itp.
 - Znajdują zastosowanie w urbanistyce, transporcie, ochronie środowiska, planowaniu przestrzennym czy zarządzaniu kryzysowym.

Alternatywne zastosowania baz danych

- Hurtownie danych dla systemów OLAP (Online Analytical Processing)
 - Hurtownie danych integrują informacje z wielu źródeł, tworząc jednolity i spójny zbiór do analiz.
 - OLAP pozwala na wielowymiarowe przetwarzanie danych (np. sprzedaż wg czasu, regionu i kategorii produktów).
 - Wykorzystywane w systemach Business Intelligence do wspomagania decyzji menedżerskich i strategicznych.

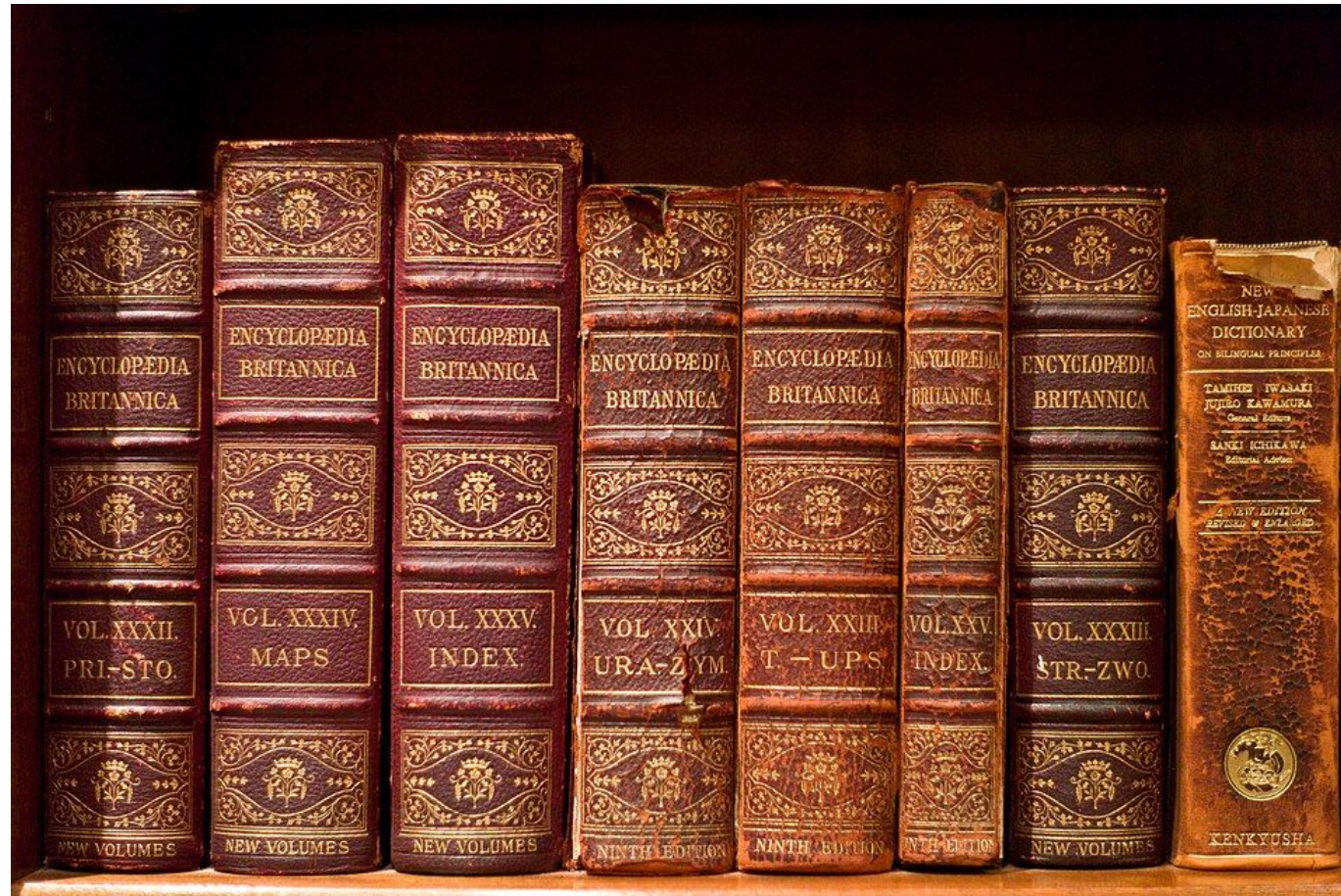
19

Alternatywne zastosowania baz danych

- Bazy danych czasu rzeczywistego
 - Obsługują procesy, w których opóźnienia w dostępie do danych mogłyby mieć krytyczne znaczenie.
 - Przykłady: sterowanie procesami przemysłowymi, monitorowanie pracy reaktorów, systemy lotnicze i kolejowe, zarządzanie ruchem ulicznym czy medyczne systemy monitorowania pacjentów.

Czy do istnienia bazy danych konieczny jest komputer?

21



Środowisko bazy danych

- Samoopisująca natura.
- Oddzielenie procesów przetwarzania od danych.
- Abstrakcja danych.
- Obsługa wielu perspektyw dla jednego zbioru danych.
- Współdzielenie danych i transakcyjność obsługi.

Samoopisująca natura

- Baza danych przechowuje nie tylko same dane, ale także ich metadane, czyli opisy struktury (np. definicje tabel, typy atrybutów, relacje, ograniczenia).
- Dzięki temu system zarządzania bazą danych (DBMS) „wie”, jak interpretować i przetwarzać przechowywane informacje.

Oddzielenie procesów przetwarzania od danych

- Aplikacje korzystające z bazy danych nie muszą znać szczegółów fizycznego sposobu zapisu danych.
- DBMS pośredniczy między użytkownikiem a danymi, co pozwala zmieniać sposób ich przechowywania bez ingerencji w kod programów korzystających z bazy.

Abstrakcja danych

- Użytkownicy mają dostęp do danych na różnych poziomach abstrakcji (np. widoki logiczne, schematy koncepcyjne, schematy fizyczne).
- Dzięki temu można ukryć szczegóły implementacyjne i prezentować tylko te informacje, które są istotne w danym kontekście.

Obsługa wielu perspektyw dla jednego zbioru danych

- Ten sam zbiór danych może być prezentowany w różny sposób różnym użytkownikom.
- Na przykład księgowy, menedżer i analityk mogą korzystać z jednej bazy, ale każdy z nich ma dostęp do innego zestawu widoków i raportów.

Współdzielenie danych i transakcyjność obsługi

- Wiele użytkowników może jednocześnie pracować na tej samej bazie danych.
- DBMS dba o to, aby operacje były wykonywane w sposób spójny i bezpieczny, wykorzystując mechanizmy transakcji (ACID: atomicity, consistency, isolation, durability, lub BASE: basically available, soft state, eventually consistent).
- Dzięki temu dane pozostają poprawne nawet w przypadku awarii czy równoczesnych modyfikacji.

28

Współdzielenie danych i transakcyjność obsługi

- Izolacyjność transakcji:
 - Gwarancja, że każda transakcja jest wykonywana w warunkach separacji od transakcji współbieżnych.
- Niepodzielność transakcji:
 - Gwarancja wykonania wszystkich operacji transakcji.
 - Jeżeli któraś operacja nie może być wykonana anulowana jest cała transakcja.

Aktorzy na scenie

- Administratorzy baz danych
- Projektanci bazy danych
- Użytkownicy końcowi
- Analitycy

30

Administratorzy (DBA)

- Dbą o zasoby:
 - Baza danych.
 - Użytkownicy.
 - Naruszenia bezpieczeństwa.
 - Problemy wydajnościowe.

Projektanci

- Identyfikują dane, które docelowo powinny być zawarte w bazie.
- Wybierają odpowiednie struktury do ich reprezentacji.
- Odpowiadają za komunikację z przyszłymi użytkownikami i zebranie informacji o ich potrzebach.
- Tworzą perspektywy.

Użytkownicy

- Osoby, które korzystają z zasobów zgromadzonych w bazie.
- Dorywczy:
 - Korzystają sporadycznie i wymagają różnych danych.
- Naiwni (parametryczni):
 - Wielokrotnie wykonują standardowe typy zapytań (transakcje zapuszkowane).
- Doświadczeni:
 - Inżynierowie, analitycy znający języki zapytań
- Samodzielni:
 - Wykorzystują własne, niezależne bazy.

33

Analitycy i inżynierowie oprogramowania

- Analiza potrzeb organizacji.
- Określają wymagania użytkowników naiwnych.
- Modelują dziedziny problemu.
- Przygotowują procedury programowe dostępu do danych.

Osoby poza sceną

- Projektanci i programiści systemów baz danych.
- Programiści narzędzi.
- Operatorzy sprzętu i administratorzy OS.

Zalety stosowania baz danych

- Kontrola nadmiarowości danych:
 - W bazach danych eliminuje się wielokrotne przechowywanie tych samych informacji, dzięki czemu unika się niespójności i oszczędza przestrzeń dyskową.
 - Przykład: dane klienta zapisane w jednej tabeli, zamiast powielane w wielu plikach.

Zalety stosowania baz danych

- Ograniczenie nieautoryzowanego dostępu do danych:
 - Systemy zarządzania bazami danych (DBMS) oferują mechanizmy kontroli dostępu (loginy, role, uprawnienia).
 - Dzięki temu każdy użytkownik ma dostęp tylko do tych danych, do których został uprawniony.

Zalety stosowania baz danych

- Trwałe przechowywanie obiektów:
 - Bazy danych umożliwiają bezpieczne i długotrwałe przechowywanie informacji, z możliwością tworzenia kopii zapasowych i odzyskiwania danych po awarii.
 - Dane są niezależne od pojedynczych aplikacji i istnieją niezależnie od tego, kto i w jaki sposób ich używa.

Zalety stosowania baz danych

- Zapewnienie efektywnego przetwarzania zapytań:
 - Dzięki indeksom, optymalizacji zapytań i specjalnym algorytmom wyszukiwanie i przetwarzanie danych jest szybkie i wydajne, nawet przy bardzo dużych zbiorach.
 - DBMS pozwala na korzystanie z języków zapytań (np. SQL), które są znacznie bardziej elastyczne niż przeszukiwanie plików tekstowych.

Zalety stosowania baz danych

- Bezpieczeństwo przechowywanych danych:
 - Mechanizmy transakcyjności (ACID), kopie zapasowe, replikacja i szyfrowanie zapewniają integralność oraz ochronę danych przed utratą i uszkodzeniem.
 - Dodatkowo systemy bazodanowe umożliwiają audyt operacji – wiadomo, kto i kiedy wprowadzał zmiany.

Zalety stosowania baz danych

- Zapewnienie interfejsów dla wielu użytkowników:
 - System bazodanowy umożliwia jednoczesny dostęp do tych samych danych przez wielu użytkowników i aplikacje.
 - Dzięki mechanizmom współbieżności oraz transakcyjności nie dochodzi do konfliktów danych.

Zalety stosowania baz danych

- Reprezentacja wielu relacji między danymi:
 - Bazy danych (zwłaszcza relacyjne) pozwalają odzwierciedlać powiązania między obiektami rzeczywistości.
 - Dzięki związkom (np. jeden-do-jednego, jeden-do-wielu, wiele-do-wielu) możliwe jest modelowanie złożonych zależności w sposób spójny i logiczny.

Zalety stosowania baz danych

- Wymuszenie integralności:
 - System bazodanowy kontroluje poprawność danych przy ich wprowadzaniu i modyfikowaniu.
 - Integralność zapewniają m.in. klucze główne, obce, ograniczenia unikalności, reguły NOT NULL czy CHECK.

Zalety stosowania baz danych

- Wnioskowanie w oparciu o zdefiniowane reguły:
 - W niektórych bazach danych (np. obiektowych czy eksperckich) można definiować reguły biznesowe i logiczne, na podstawie których system automatycznie wyciąga wnioski lub wykonuje działania.
 - Przykład: „jeżeli klient przekroczy limit kredytowy, zablokuj możliwość składania nowych zamówień”.

Zalety stosowania baz danych

- Wymuszenie stosowania standardów:
 - DBMS narzuca spójne sposoby definiowania, przechowywania i dostępu do danych (np. język SQL, standardowe mechanizmy autoryzacji).
 - Ułatwia to integrację różnych aplikacji i ogranicza chaos organizacyjny w pracy z danymi.

Zalety stosowania baz danych

- Skrócony czas tworzenia aplikacji:
 - Dzięki gotowym mechanizmom DBMS (transakcyjność, bezpieczeństwo, indeksowanie, raportowanie, zarządzanie użytkownikami) programiści nie muszą implementować ich od podstaw.
 - Pozwala to szybciej i taniej budować aplikacje, które korzystają z bazy danych.

Krótką historia baz danych

- Systemy hierarchiczne i sieciowe (lata 60. i 70.):
 - Dane były przechowywane w postaci powiązanych rekordów i struktur fizycznych.
 - Systemy hierarchiczne miały strukturę drzewa (np. IMS firmy IBM), a sieciowe umożliwiały bardziej złożone powiązania między rekordami (np. model CODASYL).
 - Wadą była duża zależność między danymi logicznymi a ich fizycznym zapisem, co utrudniało rozwój i utrzymanie.

47

Krótką historia baz danych

- Systemy relacyjne czyli RDBMS (od lat 70.):
 - Koncepcję przedstawił E.F. Codd w 1970 roku.
 - Dane zaczęto przechowywać w tabelach (relacjach), a dostęp do nich odbywał się za pomocą języka zapytań (SQL).
 - To podejście oddzieliło warstwę logiczną od fizycznej, zwiększając niezależność danych.
 - Najpopularniejszy i dominujący model baz danych.

Krótką historia baz danych

- Systemy obiektowe (lata 80. i 90.):
 - Wraz z rozwojem programowania obiektowego pojawiła się potrzeba przechowywania złożonych struktur (obiektów, klas, metod).
 - Systemy obiektowe łączyły cechy baz danych i języków obiektowych (np. ObjectDB, Versant).
 - Sprawdziły się głównie w wyspecjalizowanych zastosowaniach (CAD, multimedia).

Krótką historia baz danych

- Systemy relacyjno-obiektowe (lata 90. i 2000.):
 - Próba pogodzenia relacyjnego świata SQL z obiektowością.
 - Dodano obsługę typów złożonych, dziedziczenia, kolekcji oraz integrację z językami obiektowymi.

Krótką historia baz danych

- Systemy nierelacyjne i NoSQL (od ok. 2000):
 - Powstały w odpowiedzi na rosnące potrzeby przetwarzania wielkich zbiorów danych (Big Data) oraz aplikacji internetowych o wysokiej dostępności.
 - Modele przechowywania: klucz–wartość, dokumentowe, kolumnowe, grafowe.
 - Stawiają na skalowalność, elastyczność i szybkość, często kosztem pełnej spójności (BASE zamiast ACID).

Abstrakcja danych

- Abstrakcja danych to mechanizm ukrywający przed użytkownikiem szczegóły przechowywania informacji oraz sposób ich fizycznej organizacji.
- Dzięki temu użytkownicy i aplikacje widzą dane w formie logicznej i zrozumiałej, bez konieczności znajomości technicznych aspektów implementacji.

Abstrakcja danych

- Poziomy abstrakcji danych (wg ANSI/SPARC):
 - Poziom fizyczny opisuje, jak dane są fizycznie zapisane w pamięci (np. struktury plików, indeksy, ścieżki dostępu).
 - Poziom logiczny (konceptyjny) opisuje jakie dane są przechowywane i jakie są relacje między nimi (schemat bazy danych).
 - Poziom zewnętrzny (widoki) – określa, jak dane są widziane przez użytkowników lub aplikacje (np. raporty, formularze, widoki SQL).

Abstrakcja danych

- Użytkownik nie musi wiedzieć, czy dane są zapisane w pliku binarnym, czy w rozproszonej chmurze,
- Programista nie musi zmieniać aplikacji, gdy zmieni się sposób fizycznego przechowywania danych,
- Można łatwo udostępniać różne perspektywy tych samych danych dla różnych grup użytkowników.

Model danych

- Zbiór pojęć wykorzystywanych do opisu struktury bazy danych zawierający środki do osiągnięcia oczekiwanego poziomu abstrakcji.
- Strukturą bazy danych są typy danych, związki i ograniczenia, które muszą być spełnione przez przechowywane w bazie dane.
- Model może zawierać też aspekty dynamiczne i zachowania.

Kategorie modeli danych

- Modele niskopoziomowe (fizyczne):
 - Opisują jak dane są faktycznie przechowywane w systemie komputerowym.
 - Uwzględniają struktury plików, indeksy, adresowanie, formaty zapisu na dysku.
 - Używane głównie przez projektantów i administratorów systemów baz danych.

Kategorie modeli danych

- Modele wysokopoziomowe (konceptyjne):
 - Bliskie sposobowi, w jaki użytkownicy postrzegają dane.
 - Umożliwiają opis dziedziny problemu w sposób zrozumiały, bez wnikania w szczegóły implementacji.
 - Przykład: model związków encji (ERD).

Kategorie modeli danych

- Modele reprezentacyjne (implementacyjne):
 - Stanowią kompromis: są zrozumiałe dla użytkowników technicznych, ale jednocześnie bliskie fizycznemu zapisowi danych.
 - Mogą być bezpośrednio zaimplementowane w DBMS.
 - Najważniejszy przykład: model relacyjny.

Kategorie modeli danych

- Model obiektowy:
 - Rozszerzenie podejścia relacyjnego o cechy obiektowości (obiekty, klasy, dziedziczenie, metody).
 - Umożliwia przechowywanie i manipulowanie bardziej złożonymi strukturami danych (np. multimedialnymi).
 - Znajduje zastosowanie w systemach obiektowych i relacyjno-obiektowych (np. PostgreSQL z typami obiektowymi, Oracle z typami złożonymi).

Schemat bazy danych

- Formalny opis struktury bazy danych.
- Powstaje na etapie projektowania bazy danych i stanowi jej „plan logiczny”, według którego przechowywane i organizowane są dane.
- Jest to element względnie stabilny w przeciwieństwie do danych, które mogą zmieniać się często schemat powinien ewoluować jak narzędzie.
- Prezentuje się go w większości w postaci diagramów (np. diagram związków encji, diagram UML klas, diagramy relacyjne).

60

Stan bazy danych

- Nazywany czasem również migawką bazy.
- Bieżący zbiór danych zawartych w bazie składający się z wystąpień (egzemplarzy).
- W bazie każdy niezmienny schemat zawiera zbiór zmiennych wystąpień.
- Każda zmiana wystąpienia powoduje zmianę stanu bazy.

Stan bazy danych

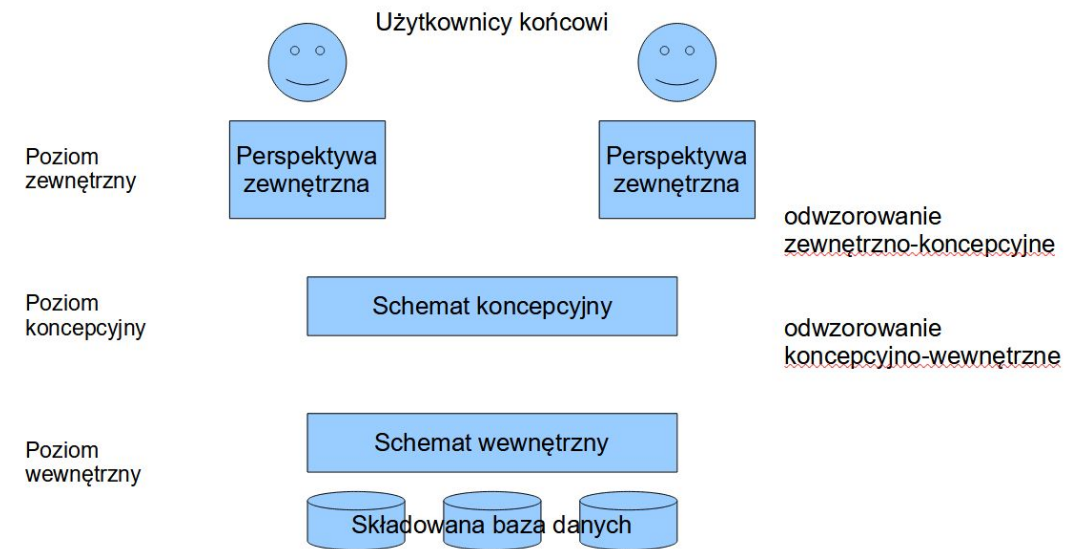
- Nowo utworzona baza jest w stanie pustym i zawiera tylko schemat.
- Pierwsze wprowadzenie wystąpienie powoduje przejście jej w stan początkowy.
- W każdym momencie stan bazy musi być stanem prawidłowym.

Reprezentacja schematu

- W systemie schemat reprezentowany jest za pomocą metadanych.
- Obecne systemy umożliwiają zmianę schematu istniejącej już bazy bez utraty danych, co określa się mianem ewolucji schematu.

Architektura trójwarstwowa

- Poziom zewnętrzny (external level).
- Poziom koncepcyjny (conceptual level).
- Poziom wewnętrzny (internal level).



Architektura trójwarstwowa

- Poziom zewnętrzny (external level):
 - Najbliższy użytkownikowi.
 - Opisuje sposób, w jaki dane są widziane przez konkretnego użytkownika lub grupę użytkowników.
 - Umożliwia definiowanie widoków, czyli wybranych fragmentów danych dopasowanych do potrzeb.

Architektura trójwarstwowa

- Poziom koncepcyjny (conceptual level):
 - Opisuje całą bazę danych w sposób logiczny.
 - Definiuje obiekty (tabele, relacje, atrybuty), ich powiązania oraz ograniczenia integralności.
 - Jest niezależny od implementacji fizycznej i pokazuje „co” przechowujemy, a nie „jak”.

Architektura trójwarstwowa

- Poziom wewnętrzny (internal level):
 - Najbliższy fizycznemu zapisowi danych.
 - Opisuje, jak dane są przechowywane w pamięci i na dysku (struktury plików, indeksy, metody adresowania, formaty zapisu).
 - Jest niewidoczny dla użytkownika końcowego.

Architektura trójwarstwowa

- Zalety architektury trójwarstwowej:
 - Oddzielenie użytkownika od szczegółów implementacyjnych.
 - Możliwość zmian na poziomie fizycznym bez wpływu na aplikacje.
 - Różni użytkownicy mogą mieć różne widoki tych samych danych.
 - Ułatwiona konserwacja, rozwój i bezpieczeństwo systemu.

Języki opisu

- SDL (Storage Definition Language).
- DDL (Data Definition Language).
- VDL (View Definition Language).
- DML (Data Manipulation Language).

Języki opisu

- **SDL (Storage Definition Language):**
 - Służy do definiowania schematu wewnętrznego, czyli sposobu fizycznego przechowywania danych.
 - Obejmuje m.in. struktury plików, metody indeksowania, strategie rozmieszczenia danych na nośnikach.
 - W praktyce rzadko dostępny bezpośrednio dla użytkownika i zwykle ukryty w DBMS.

Języki opisu

- DDL (Data Definition Language):
 - Służy do definiowania schematu koncepcyjnego bazy danych.
 - Określa strukturę bazy: tabele, atrybuty, typy danych, relacje, ograniczenia integralności.

Języki opisu

- DML (Data Manipulation Language):
 - Służy do manipulowania danymi w bazie, czyli do ich wstawiania, usuwania, aktualizowania i odczytywania.
 - W przeciwieństwie do DDL (który definiuje strukturę bazy), DML operuje na danych wewnątrz już zdefiniowanego schematu.

Języki opisu

- VDL (View Definition Language):
 - Umożliwia definiowanie schematów zewnętrznych (widoków użytkowników).
 - Pozwala tworzyć różne perspektywy danych dopasowane do potrzeb poszczególnych grup użytkowników.

Model związków encji



Model związków encji (Entity Relationship)

- Jest to wysokopoziomowy i koncepcyjny model danych, służący do opisu dziedziny problemu w sposób zrozumiały zarówno dla użytkowników, jak i projektantów systemów.
- Powstał w 1976 roku (Peter Chen) i do dziś jest podstawowym narzędziem przy projektowaniu baz danych.
- Przedstawiany najczęściej w postaci diagramów ER (ERD – Entity Relationship Diagram).

75

Podstawowe elementy modelu ER

- Encje (Entities), które reprezentują obiekty lub pojęcia ze świata rzeczywistego, które chcemy odwzorować w bazie.
- Atrybuty (Attributes) czyli cechy opisujące encje lub związki.
- Związki (Relationships), które określają powiązania między encjami.

Encja

- Dosłownie to „byt”.
- Może być to fizycznie istniejący obiekt (człowiek, rzecz, miejsce) jak i koncepcja (czynność, działanie).
- Każda encja jest opisywana przez szczegółowe właściwości czyli atrybuty.
- Wartości atrybutów są podstawowymi danymi składowanymi w bazie.

Atrybut złożony i atrybut prosty

- Atrybut złożony:
 - Można rozłożyć na prostsze grupy.
 - Każda grupa składowa ma niezależne znaczenie.
 - Może mieć strukturę hierarchiczną.
 - Ma zastosowanie gdy konieczne jest operowanie tak złożeniem, jak i poszczególnymi jego fragmentami w zależności od kontekstu.
- Atrybut niepodzielny jest atrybutem prostym, inaczej atomowym.

Atrybut jedno i wielowartościowy

- Atrybuty jednowartościowe:
 - Mają dokładnie jedną wartość dla każdej instancji encji.
 - W relacyjnych bazach danych są najczęściej odwzorowywane jako zwykłe kolumny tabeli.
- Atrybuty wielowartościowe:
 - Mogą mieć więcej niż jedną wartość dla pojedynczej encji.
 - W modelu relacyjnym nie da się ich przechowywać w pojedynczej kolumnie.

Atrybut stałe i pochodne

- Atrybuty stałe:
 - Są trwale przechowywane w bazie danych.
 - Ich wartości są wprowadzane i zapisywane bezpośrednio.
- Atrybuty pochodne:
 - Nie są przechowywane bezpośrednio w bazie, lecz mogą być wyliczane na podstawie innych danych.
 - Dzięki temu unika się redundancji i zapewnia spójność.

Typ encji

- Definiuje zbiór encji o takim samym zestawie atrybutów i nazwywa.
- Każdy jest opisywany w bazie za pomocą nazwy i atrybutów.
- Opisuje schemat i intensję kolekcji encji, które mają taką samą strukturę.

Atrybut klucza encji

- Nakładaj unikalność wartości na wybrane atrybuty.
- Klucz umożliwia jednoznaczną identyfikację poszczególnych encji.
- Encja może mieć wiele kluczy jednocześnie.
- Klucz może wykorzystywać atrybuty złożone z założeniem warunku minimalności (złożenie musi gwarantować unikalność atrybutu).
- Typ encji bez klucza jest nazywany typem słabym.

82

Związki encji

- Łączą jednoznacznie typy encji.
- Odpowiadają związkom z miniświata.
- Związki mogą łączyć wiele encji co nazywa się stopniem związku.
- Związek łączący encje tego samego typu, rekursywne, muszą mieć przypisaną rolę.

Współczynnik liczności związku

- W modelu związków encji każdy związek między dwiema encjami opisujemy za pomocą współczynnika liczności, który określa, ile obiektów jednej encji może być powiązanych z obiektami drugiej encji.

Współczynnik liczności związku

- 1:1 (jeden do jednego):
 - Każdemu obiektowi encji A odpowiada co najwyżej jeden obiekt encji B.
 - Przykład: Osoba $\leftrightarrow$ Paszport.
- 1:N (jeden do wielu):
 - Jeden obiekt encji A może być powiązany z wieloma obiektami encji B.
 - Przykład: Klient $\leftrightarrow$ Zamówienia.
- N:1 (wiele do jednego):
 - Wiele obiektów encji A może wskazywać na ten sam obiekt encji B.
 - Przykład: Zamówienie $\leftrightarrow$ Klient.
- M:N (wiele do wielu):
 - Wiele obiektów encji A może być powiązanych z wieloma obiektami encji B.
 - Przykład: Student $\leftrightarrow$ Przedmiot (student może zapisać się na wiele przedmiotów, a przedmiot może być realizowany przez wielu studentów).

85

Udział encji w związku

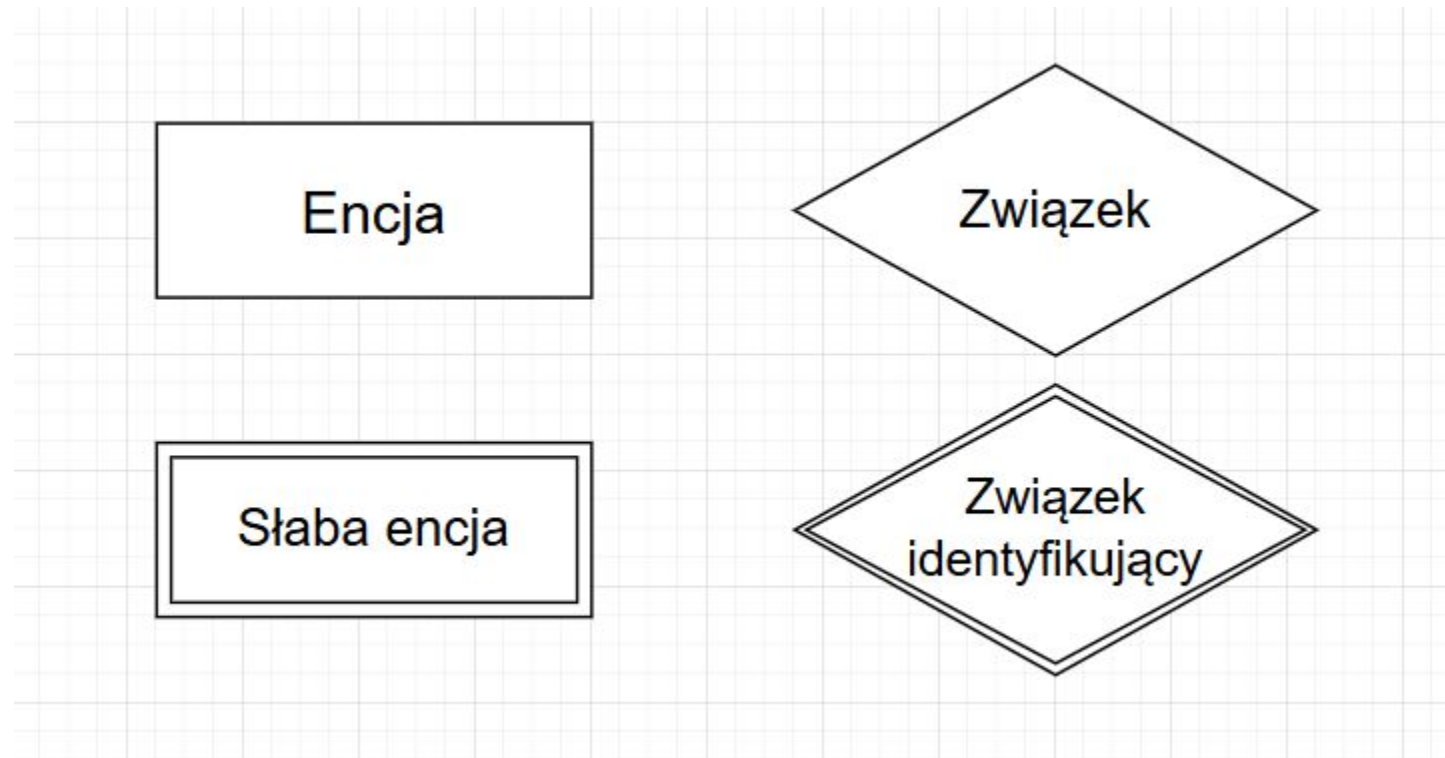
- Udział określa, czy istnienie encji jest uzależnione od powiązania ze związaną encją.
- Udział całkowity:
 - Każdy obiekt encji musi brać udział w danym związku.
 - Przykład: Student musi być przypisany do jakiegoś Kierunku studiów.
- Udział częściowy:
 - Tylko część obiektów encji bierze udział w związku.
 - Przykład: nie każdy Pracownik musi być przypisany do Projektu (np. część pracowników pełni inne funkcje).

86

Atrybuty typów związku

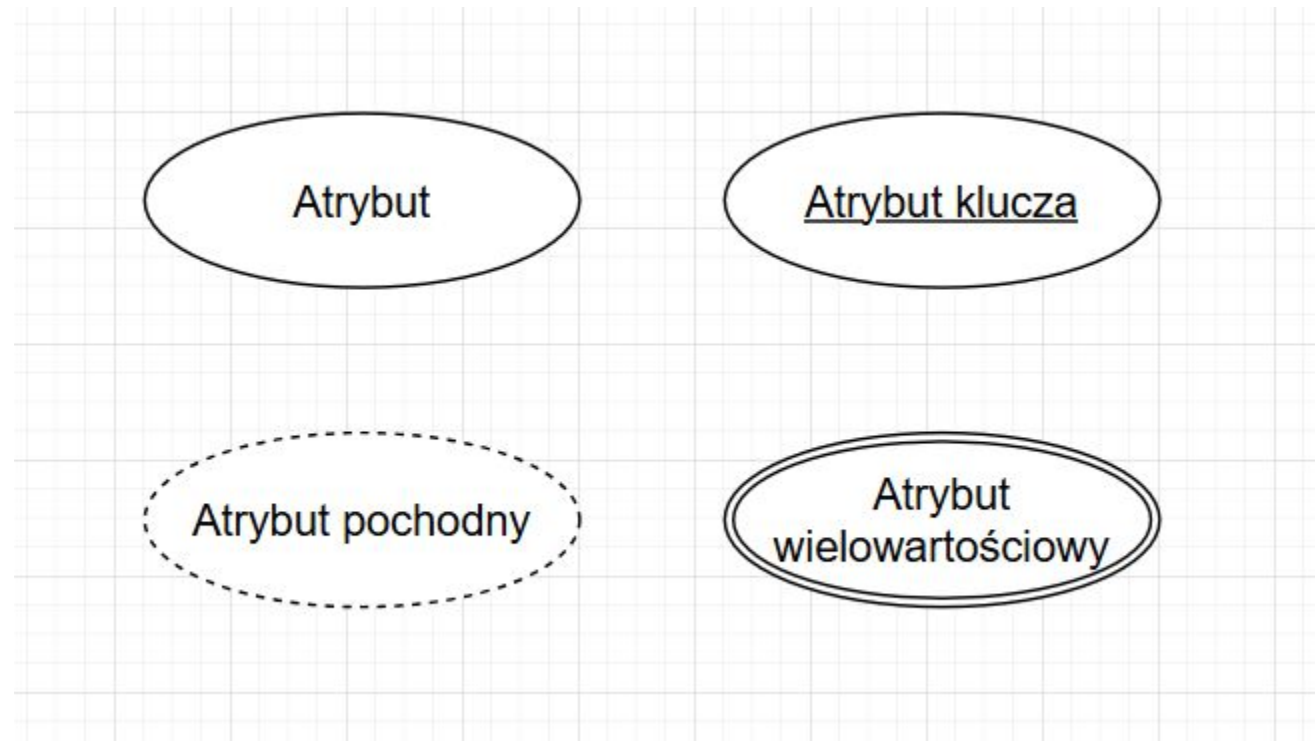
- Cechy opisujące sam związek między encjami, a nie encje osobno.
- Pojawiają się wtedy, gdy powiązanie pomiędzy encjami ma dodatkowe informacje, które nie pasują jako atrybuty samych encji.

Reprezentacja graficzna

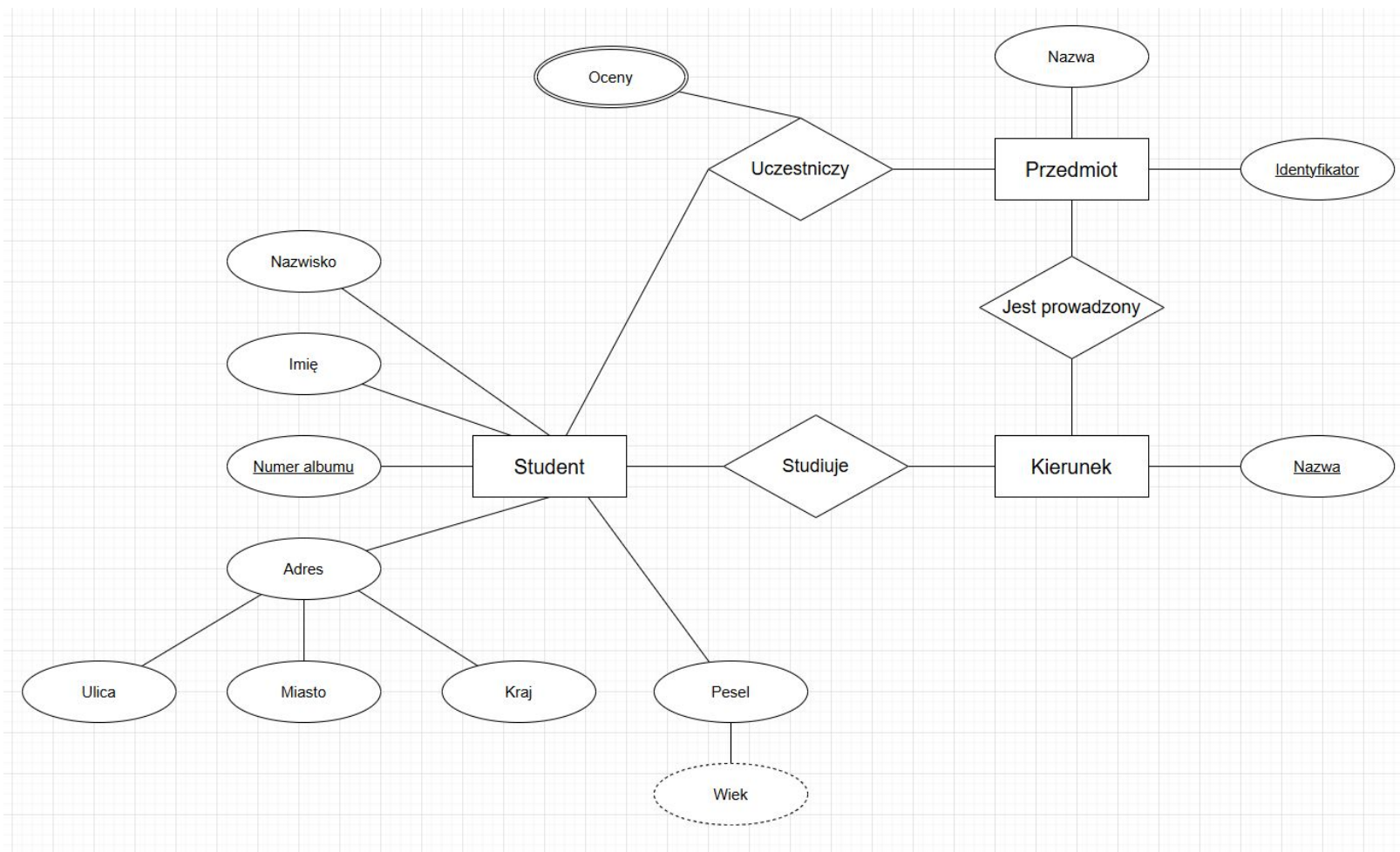


88

Reprezentacja graficzna



89



Relacyjny model danych



91

Historia RDBMS

- Edgar F. Codd (IBM Research) publikuje artykuł: “A Relational Model of Data for Large Shared Data Banks” (Communications of the ACM, 1970).
- Idea: dane i ich relacje opisane matematycznie (logicznie), nie fizycznie.
- Programista nie musi znać sposobu zapisu danych i wystarczy język zapytań.

Historia RDBMS

- IBM buduje System R (1974–1979), czyli pierwszy działający prototyp RDBMS.
- Powstaje język zapytań SEQUEL (Structured English Query Language), później nazwany SQL (bo „SEQUEL” był znakiem towarowym innej firmy).
- System R dowodzi, że relacyjne bazy mogą być wydajne praktycznie, nie tylko teoretycznie.
- Uniwersytet Kalifornijski w Berkeley rozwija Ingres, czyli inny pionierski RDBMS open source.
- Zespół Ingres później stworzy m.in. PostgreSQL i Sybase.

93

Historia RDBMS

- 1979 - Oracle V2 to pierwszy komercyjny RDBMS z SQL.
- 1981 - IBM SQL/DS to pierwszy komercyjny produkt bazodanowy IBM.
- 1983 - IBM DB2 to następca SQL/DS i jeden z pierwszych standardów korporacyjnych.
- 1989 - Microsoft SQL Server jako wynik współpracy Microsoft i Sybase.

94

Historia RDBMS

- 1995 - MySQL jako prostsza alternatywa dla mniejszych projektów, staje się popularny w Internecie.
- 1996 - PostgreSQL staje się open-source'owym następcą Postgres.
- Po 2010 - Cloud RDBMS w wydaniu Amazon RDS, Google Cloud SQL, Azure SQL Database i NewSQL, czyli nowe systemy relacyjne z architekturą rozproszoną (np. CockroachDB, YugabyteDB, TiDB).

95

Relacja? Relacyjny?

- Model relacyjny reprezentuje bazę danych jako zbiór relacji...
- ...lekką upraszczając ;-) każda relacja przypomina połączenie tabeli wartości w pliku o płaskiej organizacji danych rekordów.
- Każdy wiersz takiej tabeli reprezentuje zbiór powiązanych wartości odpowiadających wartościom atrybutów encji lub wartościom atrybutów związku.
- Nazwy tabel i kolumn ułatwiają interpretację znaczeń.

96

Dziedzina

- Dziedzina D jest zbiorem wartości atomowych w danym modelu relacyjnym.
- Dziedzina wyznacza typ danych, z których muszą pochodzić wszystkie wartości tworzące ją.
- Dziedzina ma nadaną nazwę ułatwiającą interpretowanie wartości względem reprezentowanego świata rzeczywistego.

Dziedzina

- Składa się z:
 - logicznej definicji,
 - typu lub formatu danych np.: data (rrrr-mm-dd),
 - jednostek skojarzonych np.: waga (kilogramy, funty).

Schemat relacji R

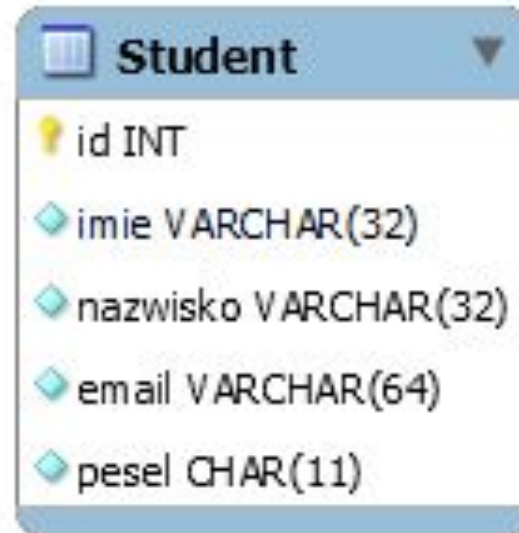
- Zapisuje się go w postaci $R(A_1, A_2, A_3, \dots, A_n)$
- A_n – n -ty atrybut, czyli nazwa roli odgrywanej przez określoną dziedzinę D w schemacie relacji R
- D nazywamy dziedziną atrybutu i zapisujemy jako: $dom(A_n)$
- R – nazwa relacji
- n – stopień relacji, czyli liczba jej atrybutów

Przykład relacji piątego stopnia

Student (id, imie, nazwisko, pesel, email)

100

Przykład relacji piątego stopnia



Student

- id INT
- imie VARCHAR(32)
- nazwisko VARCHAR(32)
- email VARCHAR(64)
- pesel CHAR(11)

101

Przykład relacji piątego stopnia

id	imie	nazwisko	email	pesel
1	Tomcio	Paluch	t.paluch@gmail.com	12345678901
2	Cali	Neczka	c.neczka@wp.pl	12345678909

Odwołanie do atrybutu

- W oparciu o nazwę.
- W oparciu o pozycję w relacji.

Stan relacji

- Stan r relacji zapisywany jako $r(R)$ dla schematu relacji $R(A_1, A_2, A_3, \dots, A_n)$ jest zbiorem m krotek,
- więc $r = \{t_1, t_2, t_3, \dots, t_m\}$
- i każda krotka jest uporządkowana listą n wartości $t = \{v_1, v_2, v_3, \dots, v_n\}$
- dodatkowo każda wartość v_i , jest elementem dziedziny $dom(A_i)$
- do i -tej wartości w krotce odwołujemy się poprzez $t[v_i]$ lub $t[i]$ przy notacji pozycyjnej.

104

Stan relacji

- R – intensja relacji.
- $r(R)$ – ekstensja relacji.

Relacja

- Stanem relacji $r(R)$ nazywamy relację matematyczną stopnia n zdefiniowaną na dziedzinach $dom(v_1), \dots, dom(v_n)$, która jest podzbiorem iloczynu kartezyjskiego dziedzin definiujących R

Relacja

$$r(R) \subseteq (dom(A_1) \times dom(A_2) \times \dots \times dom(A_n))$$

Liczność dziedziny

$$|D| = |dom(A_1)| \times |dom(A_2)| \times \dots \times |dom(A_n)|$$

Bieżący stan relacji

- Istnieje w określonym momencie.
- Zawiera jeden poprawny zbiór krotek.
- Reprezentuje konkretny stan świata rzeczywistego.
- Wraz ze zmianami świata rzeczywistego zmianom ulega też relacja.
- Schemat relacji jest względnie statyczny i nie ulega zmianom wraz ze zmianami stanu.

109

Dziedzina wspólna

- Można zdefiniować wiele atrybutów, których wartości należą do tej samej dziedziny reprezentujących jednak inne role.

Porządkowanie krotek w relacji

- Zbiór krotek, czyli relacja, nie jest uporządkowany.
- Kolejność krotek wynikająca z jej położenia np.: w pliku bazy na dysku nie jest częścią definicji relacji.
- Żaden z logicznych porządków nie jest uprzywilejowany wobec pozostałych.
- Zmiana porządku nie zmienia relacji.

111

Porządkowanie wartości wewnątrz krotek

- Każda n-krotka jest uporządkowaną listą wartości i porządek wartości w krotce jest istotny.
- Na poziomie logicznym kolejność atrybutów nie jest ważna do momentu gdy utrzymywana jest zgodność pomiędzy atrybutami i wartościami.

112

Wartości w krotkach

- Każda wartość w krotce jest atomowa.
- W modelu relacyjnym nie można stosować atrybutów złożonych i wielowartościowych co wynika z faktu, że model ten jest złożeniem postaci normalnych.
- Atrybuty wielowartościowe mogą być reprezentowane jako osobne relacje.

113

Wartości puste

- Wielość znaczeń:
 - wartość nieznana,
 - wartość niedostępna,
 - atrybut nie ma zastosowania.

Interpretacja relacji

- Schemat relacji może być interpretowany jako deklaracja lub typ twierdzenia.
- Każda krotka w relacji może być interpretowana jako fakt lub konkretny egzemplarz twierdzenia.
- Krotki mogą reprezentować zarówno fakty o istniejących encjach jak i związkach między nimi.
- Model relacyjny nie rozróżnia encji od ich związków.

115

Ograniczenia modelu relacyjnego

- Stan bazy odpowiada stanowi wszystkich powiązanych relacji.
- Trzy kategorie ograniczeń nakładanych na dane:
 - ograniczenia modelu,
 - ograniczenia schematu,
 - ograniczenia aplikacji.

116

Ograniczenia modelu relacyjnego

- Stan bazy odpowiada stanowi wszystkich powiązanych relacji.
- Trzy kategorie ograniczeń nakładanych na dane:
 - ograniczenia modelu:
 - porządkowanie krotek,
 - porządkowanie wartości w krotkach,
 - ograniczenia schematu:
 - ograniczenia dziedziny,
 - ograniczenia klucza,
 - ograniczenia wartości pustych,
 - ograniczenia aplikacji.

117

Ograniczenia dziedziny

- Wartość każdego atrybutu wewnątrz krotek musi być wartością atomową należącą do dziedziny.
- Systemy relacyjne oferują szereg typów danych powiązanych z dziedzinami, co omówimy przy SQLu.

118

Ograniczenia klucza

- Każda krotka musi być unikatowa na poziomie danej relacji.
- Nie mogą istnieć dwie krotki o takiej samej kombinacji wszystkich wartości w atrybutach.
- Istnieją podzbiory atrybutów schematu relacji, w których w danym stanie relacji nie mogą istnieć dwie krotki z taką samą kombinacją tych atrybutów.
- Zbiór takich atrybutów tworzy nadklucz **SK**.

119

Ograniczenia klucza

- Każda krotka musi być unikatowa na poziomie danej relacji.
- Nie mogą istnieć dwie krotki o takiej samej kombinacji wszystkich wartości w atrybutach.
- Istnieją podzbiory atrybutów schematu relacji, w których w danym stanie relacji nie mogą istnieć dwie krotki z taką samą kombinacją tych atrybutów.
- Zbiór takich atrybutów tworzy nadklucz **SK**.

120

Ograniczenia klucza

$$t1[SK] \neq t2[SK]$$

121

Nadklucze i klucze

- Nie mogą istnieć dwie różne krotki mające taką samą wartość nadklucza.
- Każda relacja ma co najmniej jeden nadklucz, czyli zbiór wszystkich swoich atrybutów.
- Nadklucz może zawierać atrybuty nadmiarowe.
- Klucz może zawierać tylko atrybuty nienadmiarowe.

122

Klucz

- Klucz K schematu relacji R jest nadkluczem, który po usunięciu dowolnego atrybutu A z klucza powoduje, że otrzymany zbiór atrybutów nie tworzy już nadklucza relacji.

Ograniczenia klucza

- Dwie różne krotki w żadnym stanie relacji nie mogą zawierać identycznych wartości we wszystkich atrybutach należących do klucza.
- Klucz jest minimalnym nadkluczem, czyli takim, z którego nie możemy usunąć żadnych atrybutów bez utraty unikatowości.

124

Klucz relacji

- Ograniczenie, które musi być spełnione w każdym poprawnym stanie relacji danego schematu.
- Własność ta jest niezmienna w czasie niezależnie od zmian stanu relacji.
- Może istnieć więcej niż jeden klucz dla danego schematu relacji i nazywa się je kluczami kandydującymi.
- Może istnieć tylko jeden klucz główny reprezentujący krotkę w związkach i wybierany spośród wszystkich kluczy kandydujących.

125

Schemat relacyjnej bazy danych

- Jest:
 - zbiorem relacji S ,
 - więzów integralności WI .

Stan relacyjnej bazy danych

- Zbiór stanów relacji schematu S , gdzie każde r_i jest stanem relacji spełniającym więzy WI .

Stan nieprawidłowy

- Stan bazy, który nie spełnia zdefiniowanych więzów integralności **WI** jest stanem nieprawidłowym.

128

Więzy integralności encji

- Żadna wartość atrybutu pełniącego rolę klucza głównego nie może być pusta.
- Wynika to z faktu, że wartości klucza głównego są używane do identyfikacji poszczególnych krotek.

Więzy integralności odwołań

- Wykorzystuje się je do utrzymania spójności powiązań pomiędzy krotkami należącymi do dwóch relacji.
- Odwołanie krotki z relacji do krotki z innej relacji musi wskazywać na krotkę istniejącą.
- Jest to realizowane za pośrednictwem klucza obcego.

130

Klucz obcy

- Zbiór atrybutów **FK** w schemacie relacji R_1 jest kluczem obcym tego schematu, który odwołuje się do relacji R_2 tylko gdy spełnione są warunki:
 - wartości atrybutów w zbiorze **FK** mają tą samą dziedzinę co atrybuty klucza głównego schematu relacji R_2 .
 - Wartość klucza **FK** krotki t_1 bieżącego stanu relacji $r_1(R_1)$ musi być równa wartości klucza **PK** pewnej krotki t_2 w stanie $r_2(R_2)$ albo mieć wartość pustą.
 - Mówimy, że t_1 odwołuje się do t_2 .

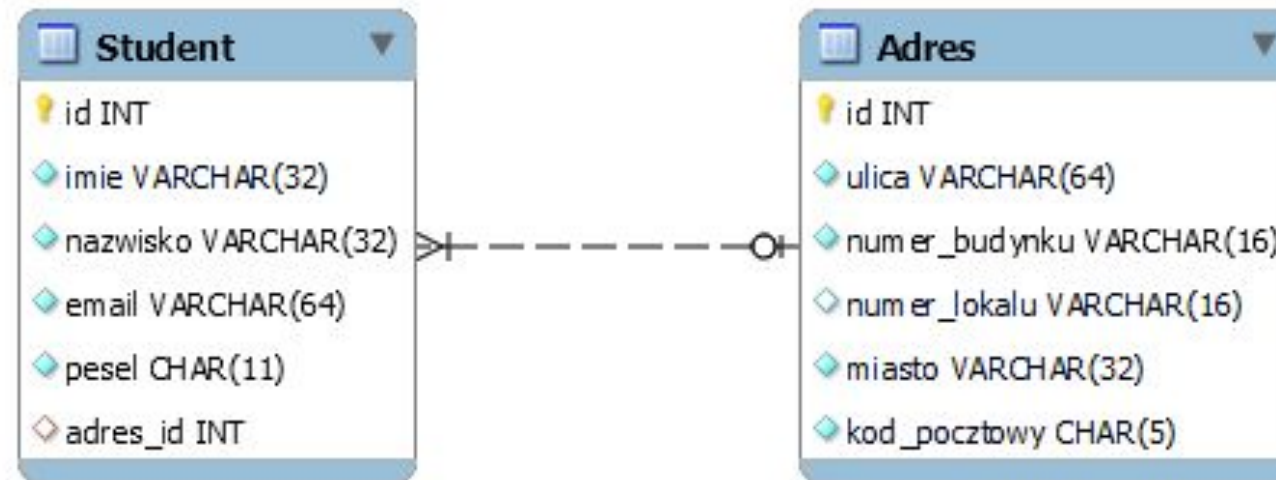
131

Klucz obcy

- Relacja R_1 jest relacją odwołującą.
- Relacja R_2 jest relacją wskazującą.
- W systemie składającym się z wielu relacji istnieje wiele różnych więzów integralności odwołań.

132

Klucz obcy



133

Określenie więzów integralności odwołań

- Odbywa się w oparciu o znaczenie i rolę encji odgrywaną w rzeczywistym świecie.
- Określamy je opierając się na powiązaniach istniejących pomiędzy encjami reprezentowanymi przez różne schematy relacji.

134

Klucz obcy

- Klucz obcy ***FK*** może wskazywać na własne relacje.
- Taką sytuację nazywa się relacją samoodwołującą lub rekurencyjną.

135

Klucz obcy



136

Więzy ograniczeń semantycznych

- Mają bardziej ogólną postać.
- Wymuszane są na poziomie aplikacji, lub wykorzystują wyzwalacze i asercje.

137

Więzy ograniczeń funkcyjnych

- Determinuje, że wartość jednego ze zbiorów atrybutów relacji zależy od wartości innego zbioru.

138

Ograniczenia stanów

- Więzy ograniczające stan relacji tak aby był on poprawny.
- Celem jest utrzymanie integralności danych nawet przy błędach użytkownika lub aplikacji.

Ograniczenia przejść

- Więzy definiujące nam sposób przejścia relacji z jednego stanu w inny.
- Monotoniczność wartości np. saldo konta nie może spaść poniżej 0.
- Kolejność stanów czyli zamówienie nie może przejść z „wysłane” do „oczekujące”.
- Czasowe ograniczenia jak w przypadku daty zakończenia $\geq$ data rozpoczęcia.
- Kontrola wersji danych gdzie nowy rekord musi mieć większy numer wersji niż poprzedni.

140

Algebra relacji



141

Algebra relacji

- Podstawowy zbiór operacji dla modelu relacyjnego.
- Umożliwia określenie podstawowych żądań wyszukiwania.
- Wynikiem działania jest nowa relacja tworzona w oparciu o zawartość jednej lub wielu istniejących relacji.

142

Algebra relacji

- Formalna podstawa do operacji modelu relacyjnego.
- Jej elementy wykorzystane zostały przy tworzeniu SQL.
- Wykorzystywana podczas optymalizacji zapytań dla relacyjnych baz danych.

143

Operacje unarne

- Operują na pojedynczych relacjach.
- Zaliczamy do nich selekcję i projekcję.

Selekcja

$$\sigma_{\langle \text{warunek selekcji} \rangle}(R)$$

$$\sigma_{\langle \text{nazwa atrybutu} \rangle \langle \text{operacja porównania} \rangle \langle \text{wartość} \rangle}(R)$$

$$\sigma_{\langle \text{nazwa atrybutu} \rangle \langle \text{operacja porównania} \rangle \langle \text{nazwa atrybutu} \rangle}(R)$$

145

Selekcja

- Nazwa atrybutu jest nazwą jednego z atrybutów R .
- Operacja porównania jest operatorem należącym do zbioru ($=$, $<$, $<=$, $>$, $>=$, $\neq$).
- Wartość jest stała z dziedziny wartości danego atrybutu
- Klauzule mogą być łączone za pomocą operatorów logicznych I, LUB oraz NIE (AND , OR , NOT).

Operatory porównania

- $=$, $<$, $<=$, $>=$, $\neq$ mają zastosowanie wyłącznie dla atrybutów z dziedzinami będącymi uporządkowanymi zbiorami wartości.
- $=$, $\neq$ mają zastosowanie dla atrybutów z dziedzinami nieuporządkowanymi.
- Operator podciągu stwierdza wystąpienie ciągu znaków w innym ciągu.

147

Wynik selekcji

- Warunek selekcji jest stosowany dla każdej krotki t w relacji R poprzez zastąpienie wystąpienia atrybutu A_i w warunku selekcji jego wartością w krotce $t[A_i]$.
- Jeżeli wynikiem będzie PRAWDA krotka t zostaje włączona do relacji wynikowej.

Przemienność selekcji

$$\sigma_{\langle \text{warunek selekcji } 1 \rangle}(\sigma_{\langle \text{warunek selekcji } 2 \rangle}(\sigma_{\langle \text{warunek selekcji } 1 \rangle}(R))) = \sigma_{\langle \text{warunek selekcji } 2 \rangle}(\sigma_{\langle \text{warunek selekcji } 1 \rangle}(\sigma_{\langle \text{warunek selekcji } 2 \rangle}(R)))$$

149

Łączność selekcji

$$\sigma_{\langle \text{warunek selekcji 1} \rangle}(\sigma_{\langle \text{warunek selekcji 2} \rangle}(R)) = \sigma_{\langle \text{warunek selekcji 1} \rangle} \circ \sigma_{\langle \text{warunek selekcji 2} \rangle}(R)$$

150

Przykłady selekcji

$$\sigma_{\text{kierunek} = 4} (STUDENT)$$

$$\sigma_{\text{średnia ocen} > 4.0} (STUDENT)$$

$$\sigma_{\text{średnia ocen} > 4.0 \text{ I } \text{średnia ocen} < 4.5} (STUDENT)$$

151

Projekcja

- Wybiera kolumny odrzucając pozostałe.
- Umożliwia wybór tylko interesujących nas atrybutów.
- Odpowiada pionowemu podziałowi relacji na dwie nowe relacje.

152

Projekcja

$$\pi_{\text{lista atrybutów}}(R)$$

Projekcja

- Stopień relacji wynikowej jest równy liczbie atrybutów na liście.
- Jeżeli wybrane atrybuty nie tworzą klucza w relacji R projekcja usuwa powtórzenia krotek.

154

Przykład projekcji

$\pi_{\text{imię, nazwisko}}$ (*STUDENT*)

Sekwencja operacji

- Można kolejno stosować wiele operacji algebry relacyjnej i układać je w sekwencje.
- Można też zapisywać je jako pojedyncze, nazwane wyrażenia algebry relacyjnej.
- Możliwe jest też działanie poprzez stworzenie sekwencji relacji pośrednich.

156

Przykłady selekcji

$$\pi_{imię, nazwisko, średnia}(\sigma_{kierunek = 4}(STUDENT))$$

$$STUDENT\ CB \leftarrow \sigma_{kierunek = 4}(STUDENT)$$

$$LISTA\ ŚREDNICH \leftarrow \pi_{imię, nazwisko, średnia}(STUDENT\ CB)$$

157

Sekwencja operacji

- Wykorzystanie pośrednich wyników jest łatwiejsze niż budowa skomplikowanych operacji złożonych.
- Możemy też używać tego w celu zmiany nazw atrybutów w pośrednich operacjach wynikowych.

158

Przykłady selekcji

CYBERSEC GRADES(first name, last name, average grades) =
 $\pi_{\text{imię, nazwisko, średnia}} (STUDENT CB)$

159

Operatory sumy, części wspólnej i różnicy

- Klasyczne operatory działania na zbiorach.
- Scalają elementy należące do dwóch relacji.
- Są operacjami binarnymi, co oznacza, że odnoszą się do dwóch krotek.
- Użycie oznacza konieczność spełnienia warunku zgodności łączenia.

160

Zgodność łączenia

- Ma miejsce gdy dla dwóch relacji $R(A_1, A_2, \dots, A_n)$ i $S(B_1, B_2, \dots, B_n)$:
 - Są tego samego stopnia n .
 - Dziedziny atrybutów są wspólne $dom(A_i) = dom(B_i)$ dla $1 \leq i \leq n$.

Operatory sumy

- Wynikiem jest relacja, która zawiera wszystkie krotki należące do jednej z relacji R i S lub obu tych relacji naraz.
- Powtórzenia krotek są automatycznie eliminowane.

162

Operator sumy

$$STUDENT\ CB \leftarrow \sigma_{kierunek = 4}(STUDENT)$$

$$STUDENT\ IT \leftarrow \sigma_{kierunek = 5}(STUDENT)$$

$$STUDENT\ NT \leftarrow STUDENT\ CB \vee STUDENT\ IT$$

163

Operator sumy

	id	imie	nazwisko	email	pesel	kierunek_id
▶	1	Tomcio	Paluch	t.paluch@gmail.com	12345678901	1
	2	Cali	Neczka	c.neczka@wp.pl	12345678909	1
	3	Bool	Lean	boolean@logic.org	12345654321	1

	id	imie	nazwisko	email	pesel	kierunek_id
▶	3	Bool	Lean	boolean@logic.org	12345654321	2
	4	Alice	Lee	a.lee@proton.mail	65432123456	2
	5	Billy	Brick	b.b@lego.com	11223311223	2

	id	imie	nazwisko	email	pesel
▶	1	Tomcio	Paluch	t.paluch@gmail.com	12345678901
	2	Cali	Neczka	c.neczka@wp.pl	12345678909
	3	Bool	Lean	boolean@logic.org	12345654321
	4	Alice	Lee	a.lee@proton.mail	65432123456
	5	Billy	Brick	b.b@lego.com	11223311223

Operator części wspólnej

	id	imie	nazwisko	email	pesel	kierunek_id
▶	1	Tomcio	Paluch	t.paluch@gmail.com	12345678901	1
	2	Cali	Neczka	c.neczka@wp.pl	12345678909	1
	3	Bool	Lean	boolean@logic.org	12345654321	1

	id	imie	nazwisko	email	pesel	kierunek_id
▶	3	Bool	Lean	boolean@logic.org	12345654321	2
	4	Alice	Lee	a.lee@proton.mail	65432123456	2
	5	Billy	Brick	b.b@lego.com	11223311223	2

	id	imie	nazwisko	email	pesel
▶	3	Bool	Lean	boolean@logic.org	12345654321

Operator różnicy

	id	imie	nazwisko	email	pesel	kierunek_id
▶	1	Tomcio	Paluch	t.paluch@gmail.com	12345678901	1
	2	Cali	Neczka	c.neczka@wp.pl	12345678909	1
	3	Bool	Lean	boolean@logic.org	12345654321	1

	id	imie	nazwisko	email	pesel	kierunek_id
▶	3	Bool	Lean	boolean@logic.org	12345654321	2
	4	Alice	Lee	a.lee@proton.mail	65432123456	2
	5	Billy	Brick	b.b@lego.com	11223311223	2

	id	imie	nazwisko	email	pesel
▶	1	Tomcio	Paluch	t.paluch@gmail.com	12345678901
	2	Cali	Neczka	c.neczka@wp.pl	12345678909

Dodatkowe założenia

- Nazwy atrybutów w relacji wynikowej są takie jak w relacji ***R***.
- Operatory sumy i części wspólnej są przemienne.
- Wszystkie trzy operatory są łączne.

Produkt: iloczyn kartezjański

- Operacja binarna.
- Nie wymaga zgodności łączenia.
- Łączy krotki pochodzące z dwóch relacji w sposób kombinatoryczny.
- Jeżeli relacja ***R*** składa się z ***n*** krotek, a ***S*** z ***m*** to wynikowa relacja ***Q*** składała się będzie z ***n*m*** krotek.
- Atrybuty są uporządkowane.

168

Produkt: iloczyn kartezjański

$$R(A_1, A_2, \dots, A_n) \times S(B_1, B_2, \dots, B_m) = Q(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$$

Produkt

	id	imie	nazwisko	email	pesel	kierunek_id
▶	1	Tomcio	Paluch	t.paluch@gmail.com	12345678901	1
	2	Cali	Neczka	c.neczka@wp.pl	12345678909	1
	3	Bool	Lean	boolean@logic.org	12345654321	1

	id	imie	nazwisko	email	pesel	kierunek_id
▶	3	Bool	Lean	boolean@logic.org	12345654321	2
	4	Alice	Lee	a.lee@proton.mail	65432123456	2
	5	Billy	Brick	b.b@lego.com	11223311223	2

	imie	nazwisko	email	pesel	nazwa
▶	Billy	Brick	b.b@lego.com	11223311223	Cyberbezpieczeństwo
	Billy	Brick	b.b@lego.com	11223311223	Informatyka Techniczna
	Alice	Lee	a.lee@proton.mail	65432123456	Cyberbezpieczeństwo
	Alice	Lee	a.lee@proton.mail	65432123456	Informatyka Techniczna
	Bool	Lean	boolean@logic.org	12345654321	Cyberbezpieczeństwo
	Bool	Lean	boolean@logic.org	12345654321	Informatyka Techniczna
	Cali	Neczka	c.neczka@wp.pl	12345678909	Cyberbezpieczeństwo
	Cali	Neczka	c.neczka@wp.pl	12345678909	Informatyka Techniczna
	Tomcio	Paluch	t.paluch@gmail.com	12345678901	Cyberbezpieczeństwo
	Tomcio	Paluch	t.paluch@gmail.com	12345678901	Informatyka Techniczna

170

Kompletny zbiór operacji

- Kompletny zbiór operacji, czyli taki, gdzie pozostałe operacje można wyrazić jako ich złożenie, tworzą:
 - projekcja,
 - selekcja,
 - suma,
 - różnica,
 - produkt.

171

Operator złączenia

- Łączy występujące w różnych relacjach, powiązane ze sobą krotki w jedną.
- Przetwarza istniejące związki między relacjami.

172

Operator złączenia

- Dla relacji $R(A_1, A_2, \dots, A_n)$ i $S(B_1, B_2, \dots, B_m)$ wynikiem operacji złączenia jest relacja Q o liczbie atrybutów $m+n$ po jednej krotce dla każdej kombinacji krotek R i S spełniającym warunek łączenia.

173

Operator złączenia

$STUDENCI \leftarrow STUDENT \bowtie KIERUNEK$

$STAROŚCI \leftarrow \sigma_{id\ starosty = id}(STUDENCI)$

174

Równozłączenie

- Jest to złączenie gdzie jedynym operatorem jest równość.
- W wyniku daje parę lub wiele par atrybutów z identycznymi wartościami we wszystkich krotkach.

175

Złączenie naturalne

- Działa jak równozłączenie tylko usuwa nadmiarowe atrybuty z krotek.
- Atrybuty w różnych relacjach, które są uwzględnione w warunku łączności muszą mieć taką samą nazwę.
- Gdy ich nazwy są inne należy wcześniej użyć operatora zmiany nazwy.

176

Selektywność złączenia

- Jeżeli n jest licznosci relacji R , a m relacji S to licznosc relacji Q powstałej w wyniku złączenia R i S jest od 0 do $m*n$.
- Oczekiwany rozmiar licznosci podzielony przez jej rozmiar maksymalny nazywamy selektywnością złączenia.

Operacje spoza algebry relacji

- Część operatorów wykorzystywanych w bazach danych wykracza poza algebrę relacji.
- Umożliwiają one specjalne działania na relacjach.
- Tak jak w algebrze relacji wynikiem jest nowa relacja.
- Do takich operacji należą między innymi agregacje, czy złączenia zewnętrzne.

Złączenie zewnętrzne

- Uwzględnia wszystkie krotki relacji R , S , lub obu bez eliminacji w sytuacji gdy brakuje pasujących wartości.

Agregacje

- Umożliwiają obliczenia statystyczne na danych zawartych w relacji.
- Najczęściej zawierają się w nich:
 - suma,
 - średnia,
 - odchylenie standardowe,
 - zliczenie,
 - wartość minimalna,
 - wartość maksymalna.

180

Operator agregacji

$A \leftarrow \text{atrybut grupowania } \mathfrak{Z}_{\text{lista funkcji}}(R)$

181

Język zapytań SQL



182

Krótką historia SQLa

- Lata 70. – w laboratoriach IBM powstaje język SEQUEL (Structured English Query Language), oparty na koncepcjach Edgara F. Codd'a dotyczących relacyjnych baz danych.
- 1974 – Donald D. Chamberlin i Raymond F. Boyce opracowują pierwszą wersję SEQUEL w projekcie System R.
- Koniec lat 70. – nazwa zostaje skrócona do SQL (z powodów prawnych).
- 1981 – Oracle wprowadza pierwszą komercyjną implementację SQL.
- 1986 – ANSI standaryzuje SQL jako język zapytań dla relacyjnych baz danych.
- 1989, 1992, 1999, 2003, 2011, 2016, 2019 – kolejne wersje standardu (m.in. SQL-89, SQL-92, SQL:1999, SQL:2003, SQL:2011, SQL:2016, SQL:2019) rozwijają m.in. obsługę XML, JSON i analityki.
- Dziś SQL pozostaje najpowszechniejszym językiem zapytań w relacyjnych bazach danych.

183

Zakres działań SQL

- Definiowanie danych.
- Manipulacja danymi.
- Kontrola dostępu.
- Kontrola transakcji.
- Zapytania i analityka w tym filtracja, grupowanie, łączenie danych, funkcje agregujące i analityczne, podzapytania.

Przyczyny popularności SQL

- Prostota i deklaratywność:
 - Użytkownik opisuje co chce uzyskać, a nie jak to zrobić.
 - Zrozumiała składnia zbliżona do języka naturalnego.
- Standaryzacja:
 - Od 1986 r. język standaryzowany przez ANSI i ISO,
- Elastyczność:
 - Obsługuje szeroki zakres operacji: od prostych zapytań po złożone analizy, widoki, transakcje i procedury.
- Wysoka wydajność i optymalizacja:
 - Systemy RDBMS zawierają zaawansowane optymalizatory zapytań i indeksy, co pozwala efektywnie przetwarzać ogromne zbiory danych.

185

Przyczyny popularności SQL

- Wsparcie i ekosystem:
 - Ogromna liczba narzędzi, bibliotek, szkoleń i specjalistów.
 - SQL stanowi fundament dla większości systemów biznesowych i analitycznych.
- Trwałość i zgodność z modelem relacyjnym:
 - Oparty na solidnej teorii (model Edgara F. Codda).
 - Gwarantuje spójność, integralność i trwałość danych (ACID).

186

Równoległość terminologii

Model relacyjny

- relacja,
- krotka,
- atrybut.

SQL

- tabela,
- wiersz,
- kolumna.

Przyczyny popularności SQL

- Wsparcie i ekosystem:
 - Ogromna liczba narzędzi, bibliotek, szkoleń i specjalistów.
 - SQL stanowi fundament dla większości systemów biznesowych i analitycznych.
- Trwałość i zgodność z modelem relacyjnym:
 - Oparty na solidnej teorii (model Edgara F. Codd).
 - Gwarantuje spójność, integralność i trwałość danych (ACID).

188

INFORMATION_SCHEMA

- Wbudowany zestaw widoków systemowych w większości relacyjnych baz danych (m.in. PostgreSQL, MySQL, SQL Server).
- Zawiera metadane, czyli dane o strukturze bazy danych.
- Jest częścią standardu SQL (ANSI/ISO).
- Zawiera informacje o:
 - tabelach i kolumnach (TABLES, COLUMNS),
 - kluczach (KEY_COLUMN_USAGE, TABLE_CONSTRAINTS),
 - widokach (VIEWS),
 - uprawnieniach (SCHEMA_PRIVILEGES, ROLE_TABLE_GRANTS).

189

Tworzenie tabel

- Definiuje nową tabelę wraz z jej nazwą.
- Zawiera informację o atrybutach.
- Definiuje więzy początkowe.
- Ograniczenia klucza, więzy integralności encji i odwołań mogą być również nałożone potem.

Ogólny schemat

```
CREATE TABLE nazwa_tabeli (  
    nazwa_kolumny typ_danych [opcje],  
    ...  
    [ograniczenia_tabeli]  
);
```

Przykład

```
CREATE TABLE `student` (  
  `id` INT NOT NULL AUTO_INCREMENT,  
  `imie` VARCHAR(32) NOT NULL,  
  `nazwisko` VARCHAR(32) NOT NULL,  
  `email` VARCHAR(64) NOT NULL,  
  `pesel` CHAR(11) NOT NULL,  
  PRIMARY KEY (`id`),  
  UNIQUE INDEX `pesel_UNIQUE` (`pesel` ASC) VISIBLE);
```

192

Tworzenie tabel

- Tabele utworzone poprzez CREATE TABLE nazywa się tabelami bazowymi.
- Możliwe jest też tworzenie tabel tymczasowych, czyli relacji wirtualnych.
- Atrybuty są uporządkowane w tabeli w kolejności podanej podczas kreacji.

Typy danych SQL

- Numeryczne.
- Ciągi znaków.
- Ciągi bitowe.
- Logiczny.
- Daty i czasu.
- Znacznika czasu.
- Przedziału czasu.

194

Typ numeryczny SQL

- Obejmuje liczby całkowite o różnym rozmiarze (INTEGER lub INT, SMALLINT, BIGINT).
- Zmiennoprzecinkowe o różnej precyzji (REAL lub FLOAT, DOUBLE PRECISION).
- Liczby formatowane definiowane przez użycie DECIMAL(i,j), DEC(i,j), NUMERIC(i,j) gdzie *i* to precyzja, a *j* to skala, czyli liczba cyfr po przecinku.

195

Ciągi znaków SQL

- Stałej długości CHAR(n) lub CHARACTER(n) gdzie **n** jest liczbą znaków.
- Zmiennej długości VARCHAR(n), CHAR VARYING(n) lub CHARACTER VARYING(n) gdzie **n** to maksymalna liczba znaków.
- Dowolnej długości TEXT.

Ciągi znaków SQL

- Zapis ciągów znaków:
 - Ograniczone apostrofem.
 - Uwzględniają wielkość liter.
 - Przy stałej długości krótsze ciągi uzupełnia się spacjami.
 - Uzupełnienia spacjami są ignorowane przy porównaniu ciągów.
 - Ciągi są porządkowane we wskazanej kolejności alfabetycznej.

197

Ciągi znaków SQL

- Funkcje operujące na ciągach znaków:
 - LENGTH(tekst) – długość ciągu.
 - LOWER(tekst) / UPPER(tekst) – zmiana wielkości liter.
 - SUBSTRING(tekst, początek, długość) – fragment tekstu.
 - REPLACE(tekst, 'stare', 'nowe') – zamiana fragmentów.
 - CONCAT('początek', 'koniec') – połączenie ciągów.
 - TRIM(' białe znaki ') – usuwa wiodące i kończące białe znaki.

Ciągi bitowe SQL

- Używa się typów bitowych do przechowywania flagi logicznej lub opcji konfiguracji ponieważ są bardziej wydajne niż BOOLEAN lub INT.
- Stałej długości BIT(n).
- Zmiennej długości BIT VARYING(n).
- Zapisuje się w konwencji **b'01010011'**.

Typ logiczny SQL

- Reprezentuje wartości TRUE i FALSE.
- Istnienie wartości pustej NULL w SQL powoduje, że jest używana logika trójwartościowa.

200

Typ daty i czasu SQL

- DATE – dziesięć pozycji, obejmuje YEAR, MONTH, DAY.
- TIME – przynajmniej osiem pozycji obejmujących między innymi: HOUR, MINUTE, SECOND.
- DATETIME – obejmuje zarówno datę, jak i czas.
- Dane do dalszego działania są dopuszczane gdy spełniają format.

201

Typ daty i czasu SQL

- TIME(i) – z precyzją ułamka sekund na i+1.
- TIME WITH TIME ZONE – uwzględnia przesunięcie względem uniwersalnej strefy czasowej w zakresie: +13:00 do – 12:59 domyślnie dla lokalnej strefy czasowej.
- DATETIME WITH TIME ZONE – j. w. ale z uwzględnieniem daty.

202

Typ daty i czasu SQL

- Poddają się porównaniom.
- Można je zapisywać jako specjalny ciąg znaków.
- Można dokonywać konwersji do wybranego formatu.

Typ znacznika czasu SQL

- **TIMESTAMP** zawiera zarówno pole **DATE** jak i **TIME** w zapisie epokowym.
- Sześć lub więcej pozycji dla części ułamkowej sekund.
- Opcjonalnie **WITH TIME ZONE**.
- Zapisuje się jako ciąg znaków: **TIMESTAMP '2008-12-01 15:00:00 123456'**.

Operacja INTERVAL SQL

- Umożliwia zmianę znacznika czasu o zadaną wartość.

```
SELECT TIMESTAMP '2025-10-21 10:00:00' - INTERVAL '2' HOUR;
```

	TIMESTAMP '2025-10-21 10:00:00' - INTERVAL '2' HOUR
▶	2025-10-21 08:00:00

205

Podstawowe ograniczenia modelu w SQL

- Ograniczenia krotek.
- Stosowania wartości pustych.
- Reguły dziedzin atrybutów.
- Ograniczenia klucza.
- Więzy integralności odwołań.

206

Ograniczenia i wartości domyślne atrybutów

- NOT NULL – ograniczenie konieczności ustawienia wartości dla atrybutu.
- Niejawnie dotyczy wszystkich kluczy głównych relacji.

Ograniczenia i wartości domyślne atrybutów

- DEFAULT – zdefiniowanie wartości domyślnej dla atrybutu.
- Wartość jest używana w sytuacji dodania krotki do relacji bez podanej wartości atrybutu.

Klauzula CHECK

- Ograniczenie nakładane na atrybuty.
- Może nakładać ograniczenia na dziedziny wartości.

Przykład

```
CREATE TABLE `student` (  
  `id` INT NOT NULL AUTO_INCREMENT,  
  `imie` VARCHAR(32) NOT NULL CHECK (LENGTH(`imie`)>0),  
  `nazwisko` VARCHAR(32) NOT NULL CHECK (LENGTH(`nazwisko`)>0),  
  `email` VARCHAR(64) NOT NULL,  
  `pesel` CHAR(11) NOT NULL DEFAULT '00000000000',  
  PRIMARY KEY (`id`),  
  UNIQUE INDEX `pesel_UNIQUE` (`pesel` ASC) VISIBLE);
```

210

Ograniczenia klucza

- Jedne z najważniejszych ograniczeń dla relacji.
- Umożliwiają określenia atrybutów składających się na klucz główny relacji.
- Umożliwiają określenie kluczy alternatywnych (kandydujących).
- Definiuje klucze obce dla relacji.

211

Przykład

```
CREATE TABLE `student` (  
  `id` INT NOT NULL AUTO_INCREMENT,  
  `imie` VARCHAR(32) NOT NULL CHECK (LENGTH(`imie`)>0),  
  `nazwisko` VARCHAR(32) NOT NULL CHECK (LENGTH(`nazwisko`)>0),  
  `email` VARCHAR(64) NOT NULL,  
  `pesel` CHAR(11) NOT NULL DEFAULT '00000000000',  
  `status` INT NOT NULL DEFAULT 0,  
  PRIMARY KEY (`id`),  
  FOREIGN KEY (`status`) REFERENCES `status` (`id`),  
  UNIQUE INDEX `pesel_UNIQUE` (`pesel` ASC) VISIBLE);
```

212

Wywołanie akcji klucza

- Domyślnie próba naruszenia więzów integralności odwołań owocuje odmową wykonania zapytania.
- Alternatywne działania mogą zostać zdefiniowane przez użytkownika w postaci akcji.
- Trzy rodzaje akcji:
 - SET NULL - ustawienie wartości pustej.
 - CASCADE - propagowanie zmiany.
 - SET DEFAULT - ustawienie wartości domyślnej.

213

Kwalifikatory akcji klucza

- Istnieją dwa przypadki, które mogą wywołać akcję klucza:
 - Usuwanie krotki wraz z jej kluczem głównym – ON DELETE.
 - Zmiana wartości atrybutu klucza w krotce – ON UPDATE.

Przykład

```
CREATE TABLE `student2` (  
  `id` INT NOT NULL AUTO_INCREMENT,  
  `imie` VARCHAR(32) NOT NULL CHECK (LENGTH(`imie`)>0),  
  `nazwisko` VARCHAR(32) NOT NULL CHECK (LENGTH(`nazwisko`)>0),  
  `email` VARCHAR(64) NOT NULL,  
  `pesel` CHAR(11) NOT NULL DEFAULT '00000000000',  
  `status` INT DEFAULT 0,  
  PRIMARY KEY (`id`),  
  FOREIGN KEY (`status`) REFERENCES `status`(`id`)  
    ON DELETE SET NULL ON UPDATE CASCADE,  
  UNIQUE INDEX `pesel_UNIQUE` (`pesel` ASC) VISIBLE);
```

215

Nadawanie nazw ograniczeniom

- Jest opcjonalne.
- Nazwy muszą być unikalne.
- Umożliwiają usuwanie i modyfikowanie poszczególnych ograniczeń.
- Wykorzystuje się w tym celu słowo kluczowe CONSTRAINT.

216

Przykład

```
CREATE TABLE `student2` (  
  `id` INT NOT NULL AUTO_INCREMENT,  
  `imie` VARCHAR(32) NOT NULL CHECK (LENGTH(`imie`)>0),  
  `nazwisko` VARCHAR(32) NOT NULL CHECK (LENGTH(`nazwisko`)>0),  
  `email` VARCHAR(64) NOT NULL,  
  `pesel` CHAR(11) NOT NULL DEFAULT '00000000000',  
  `status` INT DEFAULT 0,  
  PRIMARY KEY (`id`),  
  CONSTRAINT `student_status`  
    FOREIGN KEY (`status`) REFERENCES `status` (`id`)  
    ON DELETE SET NULL ON UPDATE CASCADE,  
  UNIQUE INDEX `pesel_UNIQUE` (`pesel` ASC) VISIBLE);
```

217

Ewolucja schematu

- Dodawanie lub usuwanie tabel, atrybutów, ograniczeń.
- Modyfikowanie innych elementów składowych schematu, w tym akcji.

218

DROP

- Usuwa nazwane elementy schematu.
- Umożliwia usuwanie też całych schematów i baz danych.

219

Możliwe zdefiniowane akcje dla DROP

- CASCADE - usuwa wraz z wszelkimi należącymi do usuwanego elementu odwołaniami w tym ograniczeniami i perspektywami.
- RESTRICT - uniemożliwia usunięcie elementu, który ma jakiegokolwiek odwołania w postaci nałożonych ograniczeń.

220

ALTER

- Umożliwia:
 - Dodawanie i usuwanie atrybutów.
 - Modyfikację definicji.
 - Zmianę ograniczeń.

221

Przykład

```
ALTER TABLE `student`  
ADD COLUMN `indeks` VARCHAR(16) NOT NULL AFTER `pesel`,  
ADD UNIQUE INDEX `indeks_UNIQUE` (`indeks` ASC);
```

```
ALTER TABLE `student`  
ALTER COLUMN `status` SET DEFAULT 100;
```


Podstawowe zapytania SQL

- Klauzula **SELECT**:
 - Mimo podobieństwa nazwy nie jest pełnym odpowiednikiem operatora selekcji z algebry relacji, ale jest złożeniem selekcji, projekcji, i w wielu przypadkach, złączenia.
 - Wynik działania nie musi dawać relacji w znaczeniu algebry relacji tylko wielozbiór (bag) krotek.
 - Zapytanie może obejmować wiele tabel.
 - W ramach zapytania możliwe jest grupowanie wyników i agregacje.

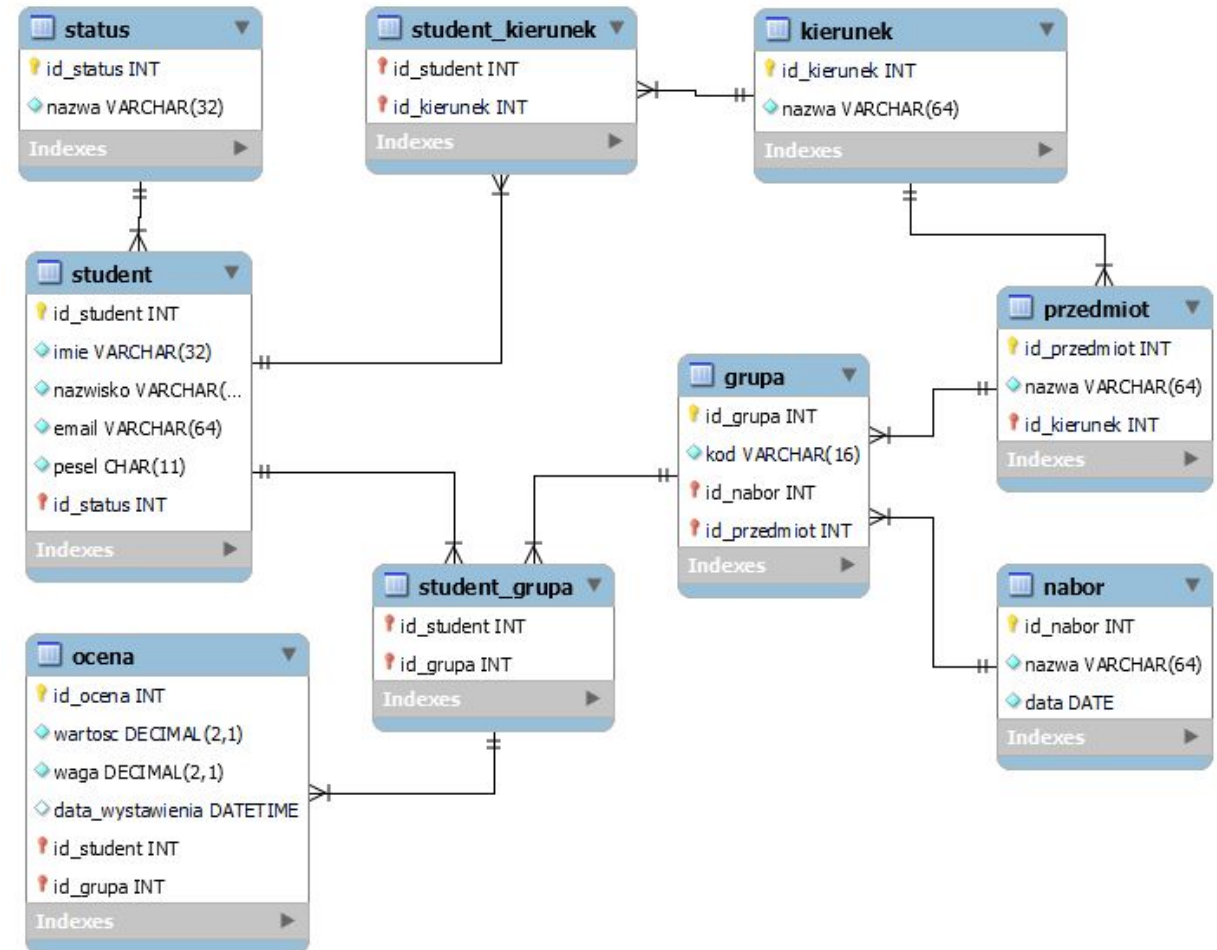
223

Ogólny schemat

```
SELECT [DISTINCT] kolumny  
FROM tabela  
[JOIN inna_tabela ON warunek_łączenia]  
[WHERE warunek_filtrowania]  
[GROUP BY kolumna]  
[HAVING warunek_dla_grup]  
[ORDER BY kolumna [ASC | DESC]]  
[LIMIT liczba_wierszy];
```

224

Baza przykładowa



225

Przykład

```
SELECT * FROM student;
```

id_student	imie	nazwisko	email	pesel	id_status
1	Hassie	Will	providenci14@example...	19213512819	1
2	Gabe	Harber	marcus.fisher@exampl...	80796331464	2
3	Tremaine	Littel	taurean36@example.org	32180144694	1
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2
5	Dion	Hegmann	loma40@example.org	96788439786	1
6	Nora	Gulgowski	deshaun.schaefer@ex...	96276650749	2
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1
8	Wellington	Howell	vlind@example.net	66587171964	2
9	Ozella	Howell	emma10@example.org	64418979671	1
10	Elmira	Larson	iwatsica@example.com	37134090138	2
11	Susanna	Turner	hilpert.santa@example...	87692682955	1
12	Virginie	Swaniawski	hhackett@example.net	66344685754	2
13	Herbert	Conroy	kiana.kemmer@exampl...	28899596987	1
14	Adelia	Little	lenny.carroll@example...	42234273856	2
15	Brock	Fay	leannon.delta@exampl...	54415787480	1

226

SELECT-FROM-WHERE

- **SELECT**
 - Określa, które kolumny mają zostać pobrane z tabeli.
 - Może zawierać wyrażenia, funkcje agregujące i aliasy.
 - Pozwala wybrać wszystkie kolumny lub tylko wybrane.
- **FROM**
 - Wskazuje źródło danych — tabelę lub kilka tabel.
 - Może obejmować złączenia (JOIN) między tabelami.
 - Jest wykonywane jako pierwszy etap przetwarzania zapytania.
- **WHERE**
 - Definiuje warunki filtrowania wierszy.
 - Zwraca tylko te rekordy, które spełniają określony warunek logiczny.
 - Może wykorzystywać operatory porównania, logiczne i wzorcowe (LIKE, BETWEEN, IN).

227

Przykład

```
SELECT * FROM student WHERE nazwisko='Schuster';
```

id_student	imie	nazwisko	email	pesel	id_status
75	Amber	Schuster	zemard@example.com	94830958998	1
94	Domingo	Schuster	eliza36@example.com	98502424418	2

Można lepiej!

228

Przykład

```
SELECT * FROM student, status WHERE student.id_status = status.id_status;
```

Można lepiej!

id_student	imie	nazwisko	email	pesel	id_status	id_status	nazwa
1	Hassie	Will	providenci14@example...	19213512819	1	1	aktywny
3	Tremaine	Littel	taurean36@example.org	32180144694	1	1	aktywny
5	Dion	Hegmann	loma40@example.org	96788439786	1	1	aktywny
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1	1	aktywny
9	Ozella	Howell	emma10@example.org	64418979671	1	1	aktywny
11	Susanna	Turner	hilpert.santa@example...	87692682955	1	1	aktywny
13	Herbert	Conroy	kiana.kemmer@exampl...	28899596987	1	1	aktywny
15	Brock	Fay	leannon.delta@exampl...	54415787480	1	1	aktywny
17	Coty	Grant	elmo.boyle@example.c...	35600584526	1	1	aktywny
19	Christelle	Schulist	eprosacco@example.net	51930582841	1	1	aktywny
21	Vince	Pagac	dwright.vandervort@e...	95710247507	1	1	aktywny
23	Jazlyn	Heidenr...	leslie03@example.com	73133915191	1	1	aktywny
25	Virginia	Kerluke	metz.selina@example....	49791040417	1	1	aktywny
27	Johathan	Schaden	moen.blaise@example....	46908194381	1	1	aktywny

229

Przykład

```
SELECT id_student, imie, nazwisko, email, status.nazwa FROM student, status  
WHERE student.id_status = status.id_status;
```

id_student	imie	nazwisko	email	nazwa
1	Hassie	Will	providenci14@example...	aktywny
3	Tremaine	Littel	taurean36@example.org	aktywny
5	Dion	Hegmann	loma40@example.org	aktywny
7	Georgia...	Bayer	rickie.veum@example....	aktywny
9	Ozella	Howell	emma10@example.org	aktywny
11	Susanna	Turner	hilpert.santa@example...	aktywny
13	Herbert	Conroy	kiana.kemmer@exampl...	aktywny
15	Brock	Fay	leannon.delta@exampl...	aktywny
17	Coty	Grant	elmo.boyle@example.c...	aktywny
19	Christelle	Schulist	eprosacco@example.net	aktywny
21	Vince	Pagac	dwright.vandervort@e...	aktywny
23	Jazlyn	Heidenr...	leslie03@example.com	aktywny
25	Virginia	Kerluke	metz.selina@example....	aktywny
27	Johathan	Schaden	moen.blaise@example....	aktywny

230

Niejednoznaczność nazw atrybutów

- Nazwa atrybutu ma zasięg lokalny (pojedyncza relacja).
- Atrybuty wielu relacji mogą mieć takie same nazwy.
- W celu jednoznacznego wskazania możliwe jest użycie kwalifikatora relacji.

```
SELECT id_student, imie, nazwisko, email, status.nazwa FROM student,  
status WHERE student.id_status = status.id_status;
```

231

Alias nazwy

- Umożliwia lokalną zmianę nazwy dla relacji i atrybutów.
- Może być zadeklarowany poprzez słowo kluczowe **AS**.
- Może być deklarowany przez umieszczenie go po relacji, której się dotyczy.

232

Przykład

```
SELECT id_student AS id, imie, nazwisko, email, status.nazwa AS status FROM  
student, status WHERE student.id_status = status.id_status;
```

```
SELECT id_student id, imie, nazwisko, email, status.nazwa status FROM  
student, status WHERE student.id_status = status.id_status;
```

id	imie	nazwisko	email	status
1	Hassie	Will	providenci14@example...	aktywny
3	Tremaine	Littel	taurean36@example.org	aktywny
5	Dion	Hegmann	loma40@example.org	aktywny
7	Georgia...	Bayer	rickie.veum@example....	aktywny
9	Ozella	Howell	emma10@example.org	aktywny

233

Operatory porównań w WHERE

- SQL daje szereg operatorów dopasowań:
 - = – równe,
 - != lub <> – różne,
 - > – większe niż,
 - < – mniejsze niż,
 - >= – większe lub równe,
 - <= – mniejsze lub równe,
 - BETWEEN ... AND ... – wartość w podanym zakresie.

234

Dopasowanie podciągów

- SQL umożliwia porównania fragmentów ciągów znakowych:
 - Używa w tym celu operatora LIKE.
 - Domyślnie jest niewrażliwy na wielkość liter, chyba że kolumna ma kolację `_cs` (case-sensitive).
 - Umożliwia użycie szblonów:
 - `%` - wiele znaków.
 - `_` - jeden znak.

235

Dopasowanie podciągów

```
SELECT DISTINCT nazwisko FROM student WHERE nazwisko LIKE '_ar%';
```

nazwisko
Harber
Larson
Ward
Carter

236

Wyrażenia regularne

- REGEXP (lub RLIKE) – operator służący do wyszukiwania wzorców w tekście za pomocą wyrażień regularnych.
- Działa na zasadzie dopasowania wzorca do całego ciągu znaków lub jego fragmentu.
- Domyślnie operator jest niewrażliwy na wielkość liter, chyba że kolumna ma kolację `_cs` (case-sensitive).
- Obsługuje składnię POSIX REGEX.

237

Wyrażenia regularne

- Podstawowe metaznaki:
 - . – dowolny znak,
 - ^ – początek ciągu,
 - \$ – koniec ciągu,
 - [] – zbiór lub zakres znaków,
 - | – alternatywa (lub),
 - *, +, ?, {n,m} – ilość wystąpień.

238

Przykład

```
SELECT DISTINCT nazwisko FROM student WHERE nazwisko RLIKE '^.ar.*$';
```

nazwisko
Harber
Larson
Ward
Carter

239

Nieokreślona klauzula WHERE

- Możliwe jest pominięcie warunku selekcji.
- Wynikiem będą wszystkie krotki wskazanej relacji.
- Jeżeli wskazano wiele relacji wynikiem działania selekcji będzie iloczyn kartezjański wskazanych relacji.

Projekcja wartości wszystkich atrybutów

- Możliwe jest otrzymanie wartości wszystkich atrybutów krotek wskazanej relacji.
- Czasem używa się projekcji wszystkiego w sytuacji gdy zależy nam na czytelniejszym zapisie zapytania w początkowym okresie jego tworzenia.

```
SELECT * FROM student;
```

241

Sortowanie wyników

- Służy do uporządkowania wyników zapytania według jednej lub wielu kolumn.
 - Składnia: **ORDER BY kolumna [ASC | DESC]**.
 - ASC (domyślnie) – sortowanie rosnące.
 - DESC – sortowanie malejące.
- Można sortować według kilku kolumn, podając je po przecinku.
- Kolejność kolumn w ORDER BY decyduje o priorytecie sortowania.
- Możliwe jest sortowanie zarówno po kolumnach tabeli, jak i po wyrażeniach lub aliasach.
- Sortowanie może być wykonywane zarówno na danych tekstowych, liczbowych, jak i datowych.

242

Przykład

```
SELECT imie, nazwisko FROM student;
```

imie	nazwisko
Hassie	Will
Gabe	Harber
Tremaine	Littel
Suzanne	West
Dion	Hegmann
Nora	Gulgowski
Georgia...	Bayer
Wellington	Howell
Ozella	Howell
Elmira	Larson

243

Przykład

```
SELECT imie, nazwisko FROM student ORDER BY nazwisko;
```

imie	nazwisko
Patricia	Abernathy
Henry	Abernathy
Colin	Bashirian
Georgia...	Bayer
Destinee	Bechtelar
Adalberto	Bechtelar
Shyanne	Bechtelar
Aurore	Bednar
Ronaldo	Bergstrom
Wendell	Bernier

244

Przykład

```
SELECT imie, nazwisko FROM student ORDER BY nazwisko, imie;
```

imie	nazwisko
Henry	Abernathy
Patricia	Abernathy
Colin	Bashirian
Georgia...	Bayer
Adalberto	Bechtelar
Destinee	Bechtelar
Shyanne	Bechtelar
Aurore	Bednar
Ronaldo	Bergstrom
Wendell	Bernier

245

Przykład

```
SELECT imie, nazwisko FROM student ORDER BY nazwisko DESC, imie;
```

imie	nazwisko
Tyrel	Zieme
Donna	Wolff
Idell	Wolff
Kavon	Wolff
Kristin	Wisozk
Hassie	Will
Suzanne	West
Ettie	Welch
Troy	Watsica
Alana	Ward

246

Operacja porównania z wartością pustą

- Wartość pusta, **NULL**, może być interpretowana na trzy sposoby:
 - Wartość nieznana (ale istniejącą).
 - Wartość niedostępna (istniejąca, ale celowo niezapisana).
 - Wartość nie mająca zastosowania w danym przypadku (niezdefiniowana dla danej krotki).

247

Operacja porównania z wartością pustą

- Możliwe jest porównywanie wartości krotek nawet w sytuacji gdy niektóre z nich zawierają wartość **NULL**.
- Wynik takiego porównania to **UNKNOWN**.
- Implikuje to fakt, że SQL korzysta z logiki trójwartościowej.

Operator IS NULL

- Umożliwia sprawdzenie czy atrybut zawiera wartość niezdefiniowaną.
- Każda wartość pusta w atrybucie jest inna od pozostałych wartości pustych, dlatego nie wolno ich porównywać bezpośrednio.

Zapytania zagnieżdżone

- Istnieje możliwość umieszczania zapytań select-from-where wewnątrz klauzuli WHERE.
- Zapytanie wewnątrz, którego następuje zagnieżdżenie nazywa się zapytaniem zewnętrznym.
- Do odniesienia się do wyniku zapytania zagnieżdżonego używa się operatora IN.

250

Przykład

```
SELECT * FROM student WHERE id_student  
IN  
(SELECT id_student FROM student_kierunek WHERE id_kierunek = 1);
```

id_student	imie	nazwisko	email	pesel	id_status
2	Gabe	Harber	marcus.fisher@exampl...	80796331464	2
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2
5	Dion	Hegmann	loma40@example.org	96788439786	1
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1
10	Elmira	Larson	iwatsica@example.com	37134090138	2
15	Brock	Fay	leannon.delta@exampl...	54415787480	1
16	Aurore	Bednar	fstehr@example.com	11670035340	2

251

Przykład

```
SELECT * FROM student WHERE id_student  
NOT IN  
(SELECT id_student FROM student_kierunek WHERE id_kierunek = 1);
```

id_student	imie	nazwisko	email	pesel	id_status
1	Hassie	Will	providenci14@example...	19213512819	1
3	Tremaine	Littel	taurean36@example.org	32180144694	1
6	Nora	Gulgowski	deshaun.schaefer@ex...	96276650749	2
8	Wellington	Howell	vlind@example.net	66587171964	2
9	Ozella	Howell	emma10@example.org	64418979671	1
11	Susanna	Turner	hilpert.santa@example...	87692682955	1
12	Virginie	Swaniawski	hhackett@example.net	66344685754	2

252

Zastąpienie IN poprzez =

- Jeżeli wewnętrzne zapytanie zwraca pojedynczy atrybut i pojedynczą krotkę dopuszczalne jest zastąpienie IN operatorem porównania.
- W przypadku gdy wewnętrzne zapytanie zwróci wiele wartości, lub zwracane jest wiele atrybutów użycie operatora porównania wygeneruje błąd.

253

ANY, SOME i ALL

- W porównaniach mogą być używane dodatkowe operatory:
 - ANY i SOME mają identyczne znaczenie i sprawdzają sytuację czy wielozbiór nie jest pusty.
 - ALL zwraca PRAWDĘ jeżeli porównanie obejmuje wszystkie krotki wielozbioru.

254

Zapytania skorelowane

- Są to podzapytania zależne od zapytania nadrzędnego.
- Wewnętrzne zapytanie (subquery) odwołuje się do kolumn z zapytania zewnętrznego.
- Dla każdego wiersza zapytania zewnętrznego, podzapytanie jest wykonywane osobno.
- Wynik podzapytania zmienia się dynamicznie w zależności od bieżącego wiersza.

255

Zapytania skorelowane

- Często wykorzystywane do porównań, sprawdzenia istnienia lub obliczeń zależnych.
- Typowe operatory w zapytaniach skorelowanych:
 - EXISTS / NOT EXISTS.
 - IN / NOT IN.
 - =, >, <, >=, <=
- Są mniej wydajne niż zapytania nieskorelowane, ponieważ wykonują się wielokrotnie.
- W wielu przypadkach można je zastąpić złączeniami dla lepszej wydajności.
- Stosowane, gdy nie da się wyrazić relacji między tabelami prostym złączeniem.

256

Eliminacja powtórzeń

- Operacja kosztowna czasowo, wymaga posortowania i przeszukania tabeli.
- W przypadku zapytań w ramach systemu relacyjnego użytkownik może oczekiwać powtórzeń o ile nie uwzględniono klucza głównego.
- Przy użyciu agregacji eliminacja powtórzeń jest nieuzasadniona.

257

Eliminacja powtórzeń

Każda tabela z krotkami zawierającymi klucz jest zbiorem, ale
wynik selekcji może być wielozbiorem.

258

Przykład

```
SELECT nazwisko FROM student ORDER BY nazwisko;
```

nazwisko
Abernathy
Abernathy
Bashirian
Bayer
Bechtelar
Bechtelar
Bechtelar
Bednar
Bergstrom

259

DISTINCT (dystynkcja)

- Umożliwia w procesie selekcji wybranie tylko krotek unikalnych.
- Redukuje, czyli eliminuje powtórzenia.
- Ma przeciwne działanie niż klauzula ALL, która jest domyślnym działaniem.

260

Przykład

```
SELECT DISTINCT nazwisko FROM student ORDER BY nazwisko;
```

nazwisko
Abernathy
Bashirian
Bayer
Bechtelar
Bednar
Bergstrom
Bernier
Boyer
Boyle

261

Algebra relacji w SQL

- Zawarto w SQL możliwość użycia algebry relacji do operacji na zbiorach.
- Daje ona możliwość użycia:
 - Sumy - **UNION**.
 - Różnicy - **EXCEPT**.
 - Iloczynu - **INTERSECT**.
- Konieczne jest używanie razem z nimi zbiorów a nie wielozbiorów co implikuje uwzględnienie klucza głównego lub użycie dystynkcji.

262

Przykład

```
SELECT DISTINCT nazwisko FROM student WHERE nazwisko LIKE 'A%'
UNION
```

```
SELECT DISTINCT nazwisko FROM student WHERE nazwisko LIKE 'B%';
```

nazwisko
Abernathy
Bayer
Bednar
Bechtelar
Bergstrom
Bashirian
Boyle
Boyer
Breitenberg
Bernier

263

Operacje na wielozbiorach

- **SQL** pozwala na działania w analogiczny do algebry relacji sposób z wykorzystaniem **UNION ALL**, **EXCEPT ALL**, **INTERSECT ALL**.

Operacje na wielozbiorach

- Funkcja **UNIQUE**:
 - Zwraca prawdę, jeżeli wynik zapytania jest unikalny.
 - Może służyć do ustalenia czy wynik jest zbiorem czy wielozbiorem.

Klauzule grupowania - GROUP BY

- Agregują wybrane podgrupy krotek danej relacji.
- Podgrupy wyznacza się w oparciu o wartości jednego lub więcej atrybutów.
- Pozwalają na zliczenia i wyznaczanie średnich.
- SQL pozwala na użycie funkcji grupującej dla każdego z atrybutów osobno.

266

Funkcje agregacyjne

- Pozwalają na wyliczenie wartości w oparciu o wartość wskazanego atrybutu.
- Operują na zgrupowanych wierszach wyniku zapytania.
- Atrybuty poddawane agregacji nie muszą pokrywać się z atrybutami użytymi do grupowania.

267

Table 12.26 Aggregate (GROUP BY) Functions

Name	Description
<u>AVG()</u>	Return the average value of the argument
<u>BIT AND()</u>	Return bitwise AND
<u>BIT OR()</u>	Return bitwise OR
<u>BIT XOR()</u>	Return bitwise XOR
<u>COUNT()</u>	Return a count of the number of rows returned
<u>COUNT(DISTINCT)</u>	Return the count of a number of different values
<u>GROUP CONCAT()</u>	Return a concatenated string
<u>JSON_ARRAYAGG()</u>	Return result set as a single JSON array
<u>JSON_OBJECTAGG()</u>	Return result set as a single JSON object
<u>MAX()</u>	Return the maximum value
<u>MIN()</u>	Return the minimum value
<u>STD()</u>	Return the population standard deviation
<u>STDDEV()</u>	Return the population standard deviation
<u>STDDEV_POP()</u>	Return the population standard deviation
<u>STDDEV_SAMP()</u>	Return the sample standard deviation
<u>SUM()</u>	Return the sum
<u>VAR_POP()</u>	Return the population standard variance
<u>VAR_SAMP()</u>	Return the sample variance
<u>VARIANCE()</u>	Return the population standard variance

Przykład

```
SELECT * FROM student, student_kierunek, kierunek
WHERE student.id_student = student_kierunek.id_student
AND student_kierunek.id_kierunek = kierunek.id_kierunek;
```



Niezmiennie można lepiej!

id_student	imie	nazwisko	email	pesel	id_status	id_student	id_kierunek	id_kierunek	nazwa
87	Lydia	Breitenberg	jackson.gusikowski@e...	81472598216	1	87	2	2	Cyberbezpieczeństwo
88	Dillan	Erdman	miller.arvilla@example....	95119047985	2	88	2	2	Cyberbezpieczeństwo
90	Joany	Gleichner	eldred15@example.org	85713688228	2	90	2	2	Cyberbezpieczeństwo
92	Karine	Franecki	lelah31@example.com	94037321068	2	92	2	2	Cyberbezpieczeństwo
93	Rowan	Greenholt	gloria85@example.org	38193916671	1	93	2	2	Cyberbezpieczeństwo
96	Ashton	Stehr	dina71@example.com	49854423348	2	96	2	2	Cyberbezpieczeństwo
97	Kristin	Wisozk	kerluke.noble@exampl...	39896127102	1	97	2	2	Cyberbezpieczeństwo
98	Tate	Nikolaus	hilpert.maximus@exam...	55908163883	2	98	2	2	Cyberbezpieczeństwo
100	Bette	Koelpin	crona.daryl@example....	83702243411	2	100	2	2	Cyberbezpieczeństwo
2	Gabe	Harber	marcus.fisher@exampl...	80796331464	2	2	1	1	Informatyka
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2	4	1	1	Informatyka
5	Dion	Hegmann	loma40@example.org	96788439786	1	5	1	1	Informatyka
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1	7	1	1	Informatyka
10	Elmira	Larson	iwatsica@example.com	37134090138	2	10	1	1	Informatyka
15	Brock	Fay	leannon.delta@exampl...	54415787480	1	15	1	1	Informatyka
16	Aurore	Bednar	fstehr@example.com	11670035340	2	16	1	1	Informatyka
18	Shyanne	Bechtelar	zemlak.megane@exam...	33538177427	2	18	1	1	Informatyka

269

Przykład

```
SELECT imie, nazwisko, nazwa  
FROM student, student_kierunek, kierunek  
WHERE student.id_student = student_kierunek.id_student  
AND student_kierunek.id_kierunek = kierunek.id_kierunek;
```



Niezmiennie można lepiej!

imie	nazwisko	nazwa
Lydia	Breitenberg	Cyberbezpieczeństwo
Dillan	Erdman	Cyberbezpieczeństwo
Joany	Gleichner	Cyberbezpieczeństwo
Karine	Franecki	Cyberbezpieczeństwo
Rowan	Greenholt	Cyberbezpieczeństwo
Ashton	Stehr	Cyberbezpieczeństwo
Kristin	Wisozk	Cyberbezpieczeństwo
Tate	Nikolaus	Cyberbezpieczeństwo
Bette	Koelpin	Cyberbezpieczeństwo
Gabe	Harber	Informatyka
Suzanne	West	Informatyka
Dion	Hegmann	Informatyka
Georgia...	Bayer	Informatyka
Elmira	Larson	Informatyka
Brock	Fay	Informatyka
Aurore	Bednar	Informatyka
Shyanne	Bechtelar	Informatyka

270

Przykład

```
SELECT nazwa, COUNT(*) `liczba studentów` FROM student, student_kierunek, kierunek  
WHERE student.id_student = student_kierunek.id_student  
AND student_kierunek.id_kierunek = kierunek.id_kierunek  
GROUP BY nazwa;
```



Niezmiennie można lepiej!

nazwa	liczba studentów
Cyberbezpieczeństwo	55
Informatyka	45

271

INSERT

- Służy do dodawania do tabeli nowych wierszy.
- Podczas dodawania wierszy sprawdzane więzy integralności nałożone na bazę.
- Konieczne jest przekazania w zapytaniu wartości dla wszystkich atrybutów, które są: obligatoryjne, ale nie posiadają wartości domyślnej.
- W niektórych SQLach, w tym MySQL, możliwe jest jednoczesne wstawianie wielu wierszy na raz.

272

INSERT (jeden wiersz)

```
INSERT INTO nazwa_tabeli (kolumna_1, kolumna_2, kolumna_3)  
VALUES (wartość_1, wartość_2, wartość_3);
```

273

INSERT (wiele wierszy)

```
INSERT INTO nazwa_tabeli (kolumna_1, kolumna_2, kolumna_3)  
VALUES
```

```
(wartość_1a, wartość_2a, wartość_3a),  
(wartość_1b, wartość_2b, wartość_3b),  
(wartość_1c, wartość_2c, wartość_3c);
```

274

Przykład

```
SELECT * FROM student;
```

```
INSERT INTO student(imie, nazwisko, email, pesel, id_status)
```

```
VALUES('Bool', 'Lean', 'bool@lean.net', '12312312312', 1);
```

275

INSERT z SELECT

- Możliwe jest wstawianie wierszy, które są wynikiem zapytania selekcji.
- Konieczna jest zgodność dziedzin dla poszczególnych kolumn.
- Wykorzystuje się takie rozwiązanie między innymi w celu:
 - Tworzenia wszelkiego rodzaju tabel pełniących rolę pamięci podręcznych.
 - Tabel z danymi archiwalnymi.
 - Tabel tymczasowych.
 - Przenoszenia danych pomiędzy bazami.

276

INSERT z SELECT

```
INSERT INTO tabela_y (kolumna_1, kolumna_2)  
SELECT kolumna_a, kolumna_b  
FROM tabela_x  
WHERE warunek;
```

277

ON DUPLICATE KEY UPDATE

- W przypadku wstawiania wiersza, który ma taką samą wartość klucza głównego jak wiersz istniejący otrzymamy błąd.
- Możliwe jest zastosowanie dodatkowego kwalifikatora akcji, który spowoduje aktualizację danych zamiast ich wstawienia.

278

Przykład

```
SELECT * FROM student WHERE nazwisko RLIKE 'Lean';
```

id_student	imie	nazwisko	email	pesel	id_status
101	Bool	Lean	bool@lean.net	12312312312	1

```
INSERT INTO student(id_student, imie, nazwisko, email, pesel, id_status)  
VALUES(101, 'Bool', 'Lean', 'bool@lean.net', '12312312312', 1);
```

Error Code: 1062. Duplicate entry '101-1' for key 'student.PRIMARY'

279

Przykład

```
INSERT INTO student(id_student, imie, nazwisko, email, pesel, id_status)
VALUES(101, 'Bool', 'Lean', 'bool@lean.net', '12312312312', 1)
ON DUPLICATE KEY UPDATE email = 'bool@lean.com';
```

```
SELECT * FROM student WHERE nazwisko RLIKE 'Lean';
```

id_student	imie	nazwisko	email	pesel	id_status
101	Bool	Lean	bool@lean.com	12312312312	1

280

INSERT IGNORE

- Przy konflikcie rekord nie zostanie dodany, ale wykonywanie kolejnych zapytań nie zostanie przerwane w przypadku użycia modyfikatora IGNORE.

INSERT IGNORE

```
INSERT INTO student(id_student, imie, nazwisko, email, pesel, id_status)  
VALUES(101, 'Bool', 'Lean', 'bool@lean.net', '12312312312', 1);
```

0 row(s) affected, 1 warning(s): 1062 Duplicate entry '101-1' for key 'student.PRIMARY'

281

DELETE

- Usuwa wiersze z tabeli.
- Zawiera klauzulę WHERE analogiczną do tej znanej z selekcji-projekcji.
- Jednocześnie możliwe jest usuwanie krotek zawartych tylko w jednej tabeli.
- Możliwe jest propagowanie operacji na krotki innych tabel jeżeli zdefiniowane zostały dla więzów odpowiednie akcje.
- Usunięcie nawet wszystkich krotek z tabeli nie usuwa jej samej (temu służy DROP).

282

DELETE

```
DELETE FROM nazwa_tabeli  
[WHERE warunek]  
[LIMIT liczba_rekordów];
```

283

Przykład

```
DELETE FROM student  
WHERE id_student = 7;
```

284

TRUNCATE

- Alias w niektórych systemach bazodanowych.
- Odpowiada usunięciu wszystkich wierszy z tabeli.

285

UPDATE

- Zmienia wartości atrybutów w jednym lub wielu wybranych wierszach.
- Selekcja wierszy odbywa się z użyciem klauzuli WHERE.
- Zmiana wartości klucza głównego jest propagowana na klucze obce tylko gdy w DDL zostały zdefiniowane odpowiednie akcje.
- W klauzuli SET wiersz może pojawiać się wielokrotnie.
- Możliwe jest ustawienie wartości domyślnej lub pustej.

286

UPDATE

```
UPDATE nazwa_tabeli  
SET kolumna1 = wartość1,  
    kolumna2 = wartość2,  
    ...  
[WHERE warunek]  
[LIMIT liczba_rekordów];
```

287

Przykłady

```
UPDATE student  
SET imie = 'Jan', id_status = 2  
WHERE id = 5;
```

```
UPDATE student AS s  
JOIN student_kierunek AS sk ON s.id_student = sk.id_student  
SET s.id_status = 2  
WHERE sk.id_kierunek = 3;
```

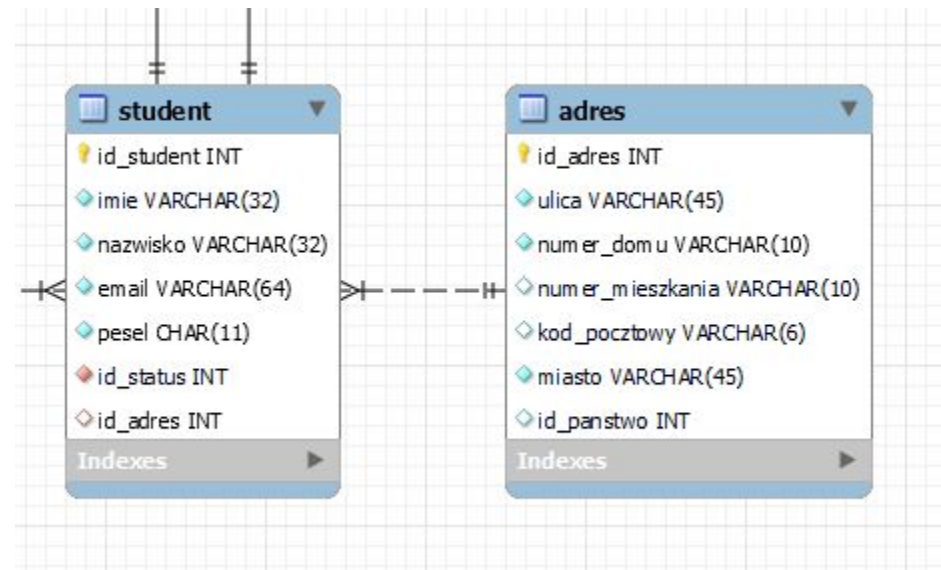
288

Złączenia

- Ponieważ mamy do dyspozycji więzy w postaci kluczy obcych dobrym (i szybkim) otrzymywaniem wyniku selekcji z wielu związanych tabel jest użycie mechanizmu złączeń.
- Może mieć zastosowanie zarówno przy zapytaniach selekcji jak i wielu tabelowych aktualizacji czy usuwania danych.

289

Złączenia



290

Złączenia

id_student	imie	nazwisko	email	pesel	id_status	id_adres
1	Hassie	Will	providenci14@example...	19213512819	1	13
2	Gabe	Harber	marcus.fisher@exampl...	80796331464	2	HULL
3	Tremaine	Littel	taurean36@example.org	32180144694	1	48
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2	13
5	Dion	Hegmann	loma40@example.org	96788439786	1	71
6	Nora	Gulgowski	deshaun.schaefer@ex...	96276650749	2	HULL
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1	85
8	Wellington	Howell	vlind@example.net	66587171964	2	HULL
9	Ozella	Howell	emma10@example.org	64418979671	1	93
10	Elmira	Larson	iwatsica@example.com	37134090138	2	35

id_adres	ulica	numer_domu	numer_mieszkania	kod_pocztowy	miasto	id_panstwo
1	Adele River	51	13	60972	Demetrishaven	1
2	Mireille Village	73	8	27837	West Pearline	3
3	Aubree Dam	36	3	97020	Janahaven	10
4	Pattie Ports	62	1	7971	Lake Ernestofort	5
5	Leda Fork	44	2	32733	West Adriel	9
6	Considine Passage	72	3	39174	Veumland	3
7	Bradtke Haven	40	7	85578	North Kelvinfurt	4
8	Purdy Light	90	8	22259	New Aniya	8
9	Huels Junction	62	8	58225	Lake Ernestina	2
10	Margaret Fields	99	13	79365	New Julenside	9

291

INNER JOIN

- Może być też po prostu użyte jako JOIN.
- Pobiera z tabeli A wszystkie wiersze mające odpowiednik w tabeli B.
- Jest to część wspólna zbiorów.
- Typowe zastosowania:
 - Danych z powiązanych tabel.
 - Tworzenie raportów łączących dane z różnych źródeł.
 - Filtrowanie rekordów na podstawie związku.

292

INNER JOIN

```
SELECT t1.kolumna, t2.kolumna  
FROM tabela1 AS t1  
INNER JOIN tabela2 AS t2  
ON t1.id = t2.id_tabela1;
```

293

INNER JOIN

id_student	imie	nazwisko	email	pesel	id_status	id_adres	id_adres	ulica	numer_domu	numer_mieszkania	kod_pocztowy	miasto	id_panstwo
1	Hassie	Will	providenci14@example...	19213512819	1	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1
3	Tremaine	Littel	taurean36@example.org	32180144694	1	48	48	Farrell Mountains	57	10	94377	South Kaelyn	5
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1
5	Dion	Hegmann	loma40@example.org	96788439786	1	71	71	Aubrey Extension	29	10	91424	Gerholdville	5
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1	85	85	Juston Crossing	48	8	85750	Krystalville	10
9	Ozella	Howell	emma10@example.org	64418979671	1	93	93	Ziemann Inlet	69	6	34714	Kertzmannhaven	2
10	Elmira	Larson	iwatsica@example.com	37134090138	2	35	35	Fahey Ridges	93	6	59695	Lake Abagail	2
11	Susanna	Turner	hilpert.santa@example...	87692682955	1	10	10	Margaret Fields	99	13	79365	New Julieside	9
12	Virginie	Swaniawski	hhackett@example.net	66344685754	2	94	94	Phoebe Throughway	60	12	7394	Stehrmouth	10
13	Herbert	Conroy	kiana.kemmer@exampl...	28899596987	1	59	59	Harris Locks	2	5	11072	Michelehaven	2
14	Adelia	Little	lenny.carroll@example...	42234273856	2	89	89	Tanya Villages	51	8	65517	Siennafort	10

294

INNER JOIN

```
SELECT * FROM student  
INNER JOIN adres  
ON student.id_adres = adres.id_adres;
```

```
SELECT * FROM student  
JOIN adres  
ON student.id_adres = adres.id_adres;
```

INNER JOIN

id_student	imie	nazwisko	email	pesel	id_status	id_adres	id_adres	ulica	numer_domu	numer_mieszkania	kod_pocztowy	miasto	id_panstwo
1	Hassie	Will	providenci14@example...	19213512819	1	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1
3	Tremaine	Littel	taurean36@example.org	32180144694	1	48	48	Farrell Mountains	57	10	94377	South Kaelyn	5
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1
5	Dion	Hegmann	loma40@example.org	96788439786	1	71	71	Aubrey Extension	29	10	91424	Gerholdville	5
7	Georgia...	Bayer	rickie.veum@example....	24511436936	1	85	85	Juston Crossing	48	8	85750	Krystalville	10
9	Ozella	Howell	emma10@example.org	64418979671	1	93	93	Ziemann Inlet	69	6	34714	Kertzmannhaven	2
10	Elmira	Larson	iwatsica@example.com	37134090138	2	35	35	Fahey Ridges	93	6	59695	Lake Abagail	2
11	Susanna	Turner	hilpert.santa@example...	87692682955	1	10	10	Margaret Fields	99	13	79365	New Julieside	9
12	Virginie	Swaniawski	hhackett@example.net	66344685754	2	94	94	Phoebe Throughway	60	12	7394	Stehrmouth	10
13	Herbert	Conroy	kiana.kemmer@exampl...	28899596987	1	59	59	Harris Locks	2	5	11072	Michelehaven	2
14	Adelia	Little	lenny.carroll@example...	42234273856	2	89	89	Tanya Villages	51	8	65517	Siennafort	10

OUTER JOIN

- Złączenie zewnętrzne.
- Łączy dane z dwóch tabel (tak jak złączenie wewnętrzne), zachowując również te rekordy, które nie mają dopasowania.
- Trzy rodzaje złączeń zewnętrznych:
 - LEFT OUTER JOIN – wszystkie rekordy z lewej tabeli + dopasowane z prawej.
 - RIGHT OUTER JOIN – wszystkie rekordy z prawej tabeli + dopasowane z lewej.
 - FULL OUTER JOIN – wszystkie z lewej i prawej tabeli z dopasowaniami (nie wspierane przez MySQL).

297

LEFT OUTER JOIN

```
SELECT t1.kolumna, t2.kolumna  
FROM tabela1 AS t1  
LEFT OUTER JOIN tabela2 AS t2  
ON t1.id = t2.id_tabela1;
```

```
SELECT t1.kolumna, t2.kolumna  
FROM tabela1 AS t1  
LEFT JOIN tabela2 AS t2  
ON t1.id = t2.id_tabela1;
```

298

LEFT OUTER JOIN

id_student	imie	nazwisko	email	pesel	id_status	id_adres	id_adres	ulica	numer_domu	numer_mieszkania	kod_pocztowy	miasto	id_panstwo
1	Hassie	Will	providenci14@example...	19213512819	1	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1
2	Gabe	Harber	marcus.fisher@exampl...	80796331464	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
3	Tremaine	Littel	taurean36@example.org	32180144694	1	48	48	Farrell Mountains	57	10	94377	South Kaelyn	5
4	Suzanne	West	ruecker.mathias@exa...	64909128687	2	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1
5	Dion	Hegmann	loma40@example.org	96788439786	1	71	71	Aubrey Extension	29	10	91424	Gerholdville	5
6	Nora	Gulgowski	deshau.n.schaefer@ex...	96276650749	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
7	Georgia...	Bayer	rickie.veum@example...	24511436936	1	85	85	Juston Crossing	48	8	85750	Krystalville	10
8	Wellington	Howell	vlind@example.net	66587171964	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
9	Ozella	Howell	emma10@example.org	64418979671	1	93	93	Ziemann Inlet	69	6	34714	Kertzmannhaven	2
10	Elmira	Larson	iwatsica@example.com	37134090138	2	35	35	Fahey Ridges	93	6	59695	Lake Abagail	2
11	Susanna	Turner	hilpert.santa@example...	87692682955	1	10	10	Margaret Fields	99	13	79365	New Julieside	9
12	Virginie	Swaniawski	hhackett@example.net	66344685754	2	94	94	Phoebe Through...	60	12	7394	Stehrmouth	10
13	Herbert	Conroy	kiana.kemmer@exampl...	28899596987	1	59	59	Harris Locks	2	5	11072	Michelehaven	2
14	Adelia	Little	lenny.carroll@example...	42234273856	2	89	89	Tanya Villages	51	8	65517	Siennafort	10

299

RIGHT OUTER JOIN

```
SELECT t1.kolumna, t2.kolumna  
FROM tabela1 AS t1  
RIGHT OUTER JOIN tabela2 AS t2  
ON t1.id = t2.id_tabela1;
```

```
SELECT t1.kolumna, t2.kolumna  
FROM tabela1 AS t1  
RIGHT JOIN tabela2 AS t2  
ON t1.id = t2.id_tabela1;
```

300

RIGHT OUTER JOIN

id_student	imie	nazwisko	email	pesel	id_status	id_adres	id_adres	ulica	numer_domu	numer_mieszkania	kod_pocztowy	miasto	id_panstwo
53	Pasquale	Glover	daniel.esmeralda@exa...	31069180825	1	1	1	Adele River	51	13	60972	Demetrishaven	1
82	Noelia	Satterfield	nfeest@example.net	37376809861	2	2	2	Mireille Village	73	8	27837	West Pearline	3
NULL	NULL	NULL	NULL	NULL	NULL	NULL	3	Aubree Dam	36	3	97020	Janahaven	10
NULL	NULL	NULL	NULL	NULL	NULL	NULL	4	Pattie Ports	62	1	7971	Lake Ernestofort	5
52	Otho	Lemke	antonietta.flatley@ex...	83404119070	2	5	5	Leda Fork	44	2	32733	West Adriel	9
77	Jermaine	Boyle	qrolfson@example.com	21931647433	1	6	6	Considine Passage	72	3	39174	Veumland	3
NULL	NULL	NULL	NULL	NULL	NULL	NULL	7	Bradtke Haven	40	7	85578	North Kelvinfurt	4
NULL	NULL	NULL	NULL	NULL	NULL	NULL	8	Purdy Light	90	8	22259	New Aniya	8
94	Domingo	Schuster	eliza36@example.com	98502424418	2	9	9	Huels Junction	62	8	58225	Lake Ernestina	2
11	Susanna	Turner	hilpert.santa@example...	87692682955	1	10	10	Margaret Fields	99	13	79365	New Julianside	9
88	Dillan	Erdman	miller.arvilla@example....	95119047985	2	10	10	Margaret Fields	99	13	79365	New Julianside	9
NULL	NULL	NULL	NULL	NULL	NULL	NULL	11	Douglas Streets	70	5	6948	Carmineshire	9
NULL	NULL	NULL	NULL	NULL	NULL	NULL	12	Curt Land	84	5	80466	Lake Annabelville	7
1	Hassie	Will	providenci14@example...	19213512819	1	13	13	Haag Islands	4	2	35893	Runolfsdottirfort	1

301

Asercje

- Definiuje ograniczenia ogólne (nie możliwe do zapisania w postaci więzów i wartości domyślnych).
- Asercje są nazwane.
- Wykorzystuje klauzule WHERE w wewnętrznej selekcji.
- Wewnętrzna selekcja opisuje kombinację krotek, które naruszają ograniczenie.

302

Asercje

- Może być nakładana na więzy atrybutów i dziedziny wartości.
- Ponieważ jej poprawność jest sprawdzana tylko przy wstawianiu i aktualizowaniu krotek jest bardziej wydajna niż ograniczenia ogólne.
- Użycie jej może mieć miejsce tylko dla tych atrybutów, krotek i dziedzin, w przypadku, których mamy pewność, że ich więzy mogą być naruszone tylko przy wstawianiu bądź aktualizacji.

303

Asercje

- Może być nakładana na więzy atrybutów i dziedziny wartości.
- Ponieważ jej poprawność jest sprawdzana tylko przy wstawianiu i aktualizowaniu krotek jest bardziej wydajna niż ograniczenia ogólne.
- Użycie jej może mieć miejsce tylko dla tych atrybutów, krotek i dziedzin, w przypadku, których mamy pewność, że ich więzy mogą być naruszone tylko przy wstawianiu bądź aktualizacji.

304

Asercje

- Asercję usuwa się za pomocą DROP.
- Nie umożliwia działań złożonych.
- Nie daje możliwości informowania o próbie naruszenia ograniczenia.

305

CREATE ASSERTION

```
CREATE ASSERTION nazwa_asercji  
CHECK (warunek_logiczny);
```

```
CREATE TABLE konto (  
    id INT PRIMARY KEY,  
    saldo DECIMAL(10,2) CHECK (saldo >= 0)  
);
```

MySQL nie obsługuje asercji!

306

Widoki

- Pojedyncza, wirtualna tabela wywiedziona z innych tabel.
- Inne tabele mogą być zarówno tabelami bazowymi jak i innym widokami.
- Nie ma (standardowo) postaci fizycznej.
- Z reguły definiuje nam zakresy danych, do których zależy nam na częstym dostępie.
- Upraszcza i przyspiesza tworzenie zapytań.

307

CREATE VIEW

```
CREATE [OR REPLACE] VIEW nazwa_widoku [(lista_kolumn)] AS  
SELECT kolumny  
FROM tabela  
[WHERE warunek]  
[GROUP BY ...]  
[HAVING ...]  
[ORDER BY ...];
```

308

Widoki

- Ma nadaną nazwę.
- Zawiera listę atrybutów.
- Atrybuty mogą powstawać zarówno w wyniku kopiowania atrybutów z istniejących tabel, jak i poprzez operacje arytmetyczne, agregacyjne i logiczne.

309

Przykład

```
CREATE OR REPLACE VIEW student_z_adresem
```

```
(id_student, imie, nazwisko, ulica, numer_domu, numer_mieszkania, kod_pocztowy, miasto)
```

```
AS
```

```
SELECT id_student, imie, nazwisko, ulica, numer_domu, numer_mieszkania, kod_pocztowy, miasto
```

```
FROM student
```

```
JOIN adres
```

```
ON student.id_adres = adres.id_adres;
```

310

Przykład

```
SELECT * FROM student_z_adresem;
```

id_student	imie	nazwisko	ulica	numer_domu	numer_mieszkania	kod_pocztowy	miasto
1	Hassie	Will	Haag Islands	4	2	35893	Runolfsdottirfort
3	Tremaine	Littel	Farrell Mountains	57	10	94377	South Kaelyn
4	Suzanne	West	Haag Islands	4	2	35893	Runolfsdottirfort
5	Dion	Hegmann	Aubrey Extension	29	10	91424	Gerholdville
7	Georgianna	Bayer	Juston Crossing	48	8	85750	Krystalville
9	Ozella	Howell	Ziemann Inlet	69	6	34714	Kertzmannhaven
10	Elmira	Larson	Fahey Ridges	93	6	59695	Lake Abagail
11	Susanna	Turner	Margaret Fields	99	13	79365	New Julienside
12	Virginie	Swaniawski	Phoebe Throughway	60	12	7394	Stehrmouth
13	Herbert	Conroy	Harris Locks	2	5	11072	Michelehaven
14	Adelia	Little	Tanya Villages	51	8	65517	Siennafort

311

Materializacja widoku

- Tworzy fizyczną tymczasową tabelę.
- Zakłada, że następne zapytania będą korzystały z tak utworzonej pamięci podręcznej.
- Konieczne jest zastosowanie mechanizmu automatycznej aktualizacji.
- Stosuje się technikę aktualizacji przyrostowej uwzględniającej wstawianie, zmianę i usuwanie krotek z tabel bazowych.
- Gdy tabela tymczasowa jest przez pewien czas nie używana zostaje usunięta.
- Nie jest dostępna w MySQL.

312

Aktualizacja poprzez widok

- Przy widoku, który nie korzysta z agregacji i dotyczy pojedynczej tabeli można wykonać ją poprzez proste rzutowanie.
- Przy perspektywie wiążącej wiele tabel aktualizacja może odbywać się na kilka różnych sposobów, ale zasadniczo możliwa jest tylko wtedy gdy nie ma niejednoznaczności mapowania.

313

Aktualizacja poprzez widok

- Widok oparty o jedną tabelę bazową może być aktualizowany gdy:
 - jej atrybuty obejmują klucz główny tabeli,
 - zawiera wszystkie atrybuty zadeklarowane z ograniczeniem NOT NULL bez zdefiniowanej wartości domyślnej.

314

Aktualizacja poprzez widok

- Widok nie jest aktualizowalny, jeśli:
 - zawiera JOIN,
 - zawiera GROUP BY lub DISTINCT,
 - zawiera UNION lub UNION ALL,
 - używa funkcji agregujących (SUM, AVG, COUNT itd.),
 - ma subquery w SELECT lub WHERE,
 - zawiera LIMIT,
 - ma aliasy kolumn niepowiązane z tabelami źródłowymi.
- Jeśli chcesz mieć pewność, że dane wprowadzane przez widok spełniają jego warunek, użyj WITH CHECK OPTION.

315

Wyzwalacze

- Obiekt związany z tabelą wykonujący określoną akcję na tabeli (lub tabelach) bazy.
- Określa jasno moment wykonania akcji.
- Działa na wszystkie wiersze tabeli dla której został zdefiniowany.
- Definiowany jest dla określonego rodzaju akcji.
- Jedna tabela może mieć przypisanych wiele wyzwalaczy, ale z pewnymi ograniczeniami.

316

Wyzwalacze

- Nie przyjmuje parametrów.
- Nie może zatwierdzać transakcji (COMMIT) i ich wycofywać (ROLLBACK).

317

CREATE TRIGGER

```
CREATE TRIGGER nazwa_wyzwalacza  
{BEFORE | AFTER} {INSERT | UPDATE | DELETE}  
ON nazwa_tabeli  
FOR EACH ROW  
BEGIN  
    -- instrukcje SQL  
END;
```

318

Wyzwalacze

- Czas wywołania:
 - BEFORE – przed modyfikacją wiersza.
 - AFTER – po modyfikacji wiersza.
- Akcja wywołująca:
 - INSERT: INSERT, LOAD DATA, REPLACE.
 - UPDATE.
 - DELETE: DELETE, REPLACE.
 - Nie jest wywoływany wyzwalacz dla DROP TABLE i TRUNCATE.

319

Wyzwalacze

- Liczba wyzwalaczy na tabelę:
 - Dwa czasy wykonania na akcję co daje 6 wyzwalaczy.
 - Nie może być dwóch wyzwalaczy z tym samym czasem i akcją.

320

Wyzwalacze

- Zapytania SQL dla wyzwalacza:
 - Zamknięte w bloku pomiędzy słowami kluczowymi BEGIN i END.
 - Mogą odnosić się do wielu tabel.

321

Przykład

DELIMITER \$\$

```
CREATE TRIGGER student_status  
AFTER INSERT ON student_kierunek  
FOR EACH ROW  
BEGIN  
    UPDATE student  
    SET id_status = 2  
    WHERE id_student = NEW.id_student;  
END$$
```

DELIMITER ;

322

Normalizacja baz danych

- Celem jest uporządkowanie struktury bazy danych, aby uniknąć nadmiarowości i anomalii aktualizacji co daje:
 - redukcję duplikacji danych,
 - poprawę integralności (dane nie sprzeczne ze sobą),
 - łatwiejszą konserwacja i rozwój bazy,
 - większą przejrzystość projektu.
- Dzielimy duże tabele na mniejsze, powiązane relacjami, zachowując spójność danych.

323

Postać normalna

- Rozłożenie danych w pewien logiczny, nie posiadający powtórzeń sposób umożliwiający złożenie jej ponownie w wyjściową całość.
- Integralna część koncepcji relacyjnych baz danych.
- Działania na danych są w założeniu nieproceduralne.
- Zaproponowane przez Edgara F. Coda (jakżeby inaczej).

324

Postać normalna

- Rozłożenie danych w pewien logiczny, nie posiadający powtórzeń sposób umożliwiający złożenie jej ponownie w wyjściową całość.
- 6 postaci normalnych w tym 3 używane standardowo dlatego baza w 3 postaci normalnej uważana jest za w pełni znormalizowaną.

325

Postać normalna

- Założenia wstępne:
 - Tabela powinna opisywać jedną i tylko jedną encję bez skrótów czy połączeń.
 - Wszystkie wiersze muszą być unikalne.
 - Kolejność wierszowa nie ma znaczenia.

326

Postać normalna

- Identyfikacja encji:
 - Podstawa normalizacji.
 - Istnieje grupa encji oczywistych.
 - Część encji wynika z procesu optymalizacji.

327

Postać normalna

- Pierwsza postać normalna:
 - Gwarantuje niepodzielność danych, ich samodzielności i niezależności.
 - Eliminuje powtarzające się grupy danych.
 - Przełamywane są kolumny łączące dane na osobne wiersze lub kolumny.

328

Postać normalna

- Powtarzające się dane:
 - Charakterystyczna cecha baz o płaskiej organizacji.
 - Marnują przestrzeń dyskową.
 - Powodują rozbudowanie mechanizmów przenoszenia danych i tracenie przepustowości tak systemu jak i mediów.
 - Mogą prowadzić do naruszenia integralności danych lub paradoksów.
 - Konieczne jest budowanie rozbudowanych mechanizmów parsowania, których wykonania kosztuje czas.

329

Postać normalna

- Dopuszczenie do pierwszej postaci normalnej:
 - Grupy danych ulegają powtórzeniom, co oznacza konieczność wydzielenia ich do osobnej tabeli.
 - Kolumny zawierające wartości nieatomowe muszą być podzielone na reprezentację w postaci złożenia kolumn z wartościami atomowymi.

330

Postać normalna

- Nowe tabele muszą posiadać klucz główny utworzony syntetycznie lub wybrany z kluczy kandydujących.
- Musi być zapewnienie odwołania do danych z tabeli głównej.
- Redukuje występowanie sprzecznych informacji poprzez eliminację powtórzeń.
- Przenoszone są tylko powtarzające się grupy danych.

331

Postać normalna

- Druga postać normalna:
 - Tabela musi być w pierwszej postaci normalnej.
 - Każda kolumna musi być zależna od całego klucza.

332

Postać normalna

- Wydzielenie tabel:
 - Proces określany przełamaniem tabeli.
 - Tabela główna zawiera dane przechowywane jednokrotnie.
 - Tabela szczegółowa zawiera te informacje, które mogą występować wielokrotnie.

333

Postać normalna

- Trzecia postać normalna:
 - Tabela musi być w drugiej postaci normalnej.
 - Żadna kolumna nie może zależeć od innej kolumny nie będącej kluczem.
 - Tabela nie może zawierać danych wyliczonych lub zagregowanych.
 - Otrzymywana poprzez usunięcie kolumn z wartościami pochodnymi.

334

Postać normalna

- Czwarta postać normalna:
 - Tabela jest nie zawiera zależności wielowartościowych.
 - W jednym wierszu nie powinno być niezależnych list wartości.
 - Unika powielania kombinacji niezależnych danych.

335

Postać normalna

- Piąta postać normalna (postać projekcyjno-skalowalna, PJNF):
 - Tabela i każda zależność złączeniowa (join dependency) wynika z kluczy kandydujących.
 - Tabela nie powinna być możliwa do bezstratnego rozbicia na mniejsze tabele w sposób bardziej złożony niż zwykłe zależności funkcyjne.
 - Eliminuje nadmiarowości wynikającej z bardziej złożonych relacji wielotabelowych.
 - Rzadko stosowana w praktyce.

336

Postać normalna

- Postać Boyce'a-Codda (BCNF):
 - Między trzecią a czwartą postacią normalną.
 - Dla każdej zależności funkcyjnej $X \rightarrow Y$, zbiór X musi być kluczem kandydującym.
 - Usuwa pewne niejednoznaczności, których nie eliminuje trzecia postać normalna.

337

Denormalizacja

- Celowe odejście od postaci normalnej jest dopuszczalne w przypadku gdy:
 - Wartość wyliczona przyspieszy generację odpowiedzi na zapytanie (należy dbać o utrzymanie spójności).
 - Redukuje w znaczący sposób liczbę złączeń.
 - Operuje się na danych historycznych, które w mniejszym stopniu są podatne na utratę spójności.
 - W znacznym stopniu upraszcza to tworzenie zapytań.

338

Problem

- Plik zawiera dane, które mogą być zarówno o charakterze uporządkowanym jak i nie.
- Jak dokonać efektywnego wyszukiwania rekordu znajdującego się w nim dla zadanej wartości atrybutu?

339

Rozwiązanie

- Utworzyć dodatkowy plik, który będzie zdefiniowany na atrybucie wykorzystanym do specyfikacji kryterium wyszukiwania.
- Zawiera on rekordy odpowiadające poszukiwanym wartościom pierwszych rekordów w poszczególnych blokach pliku danych.

340

Rozwiązanie

- Rekordy mają postać:
 - pierwszy klucz w bloku -> wskaźnik do bloku.
 - Plik dodatkowo jest wewnętrznie uporządkowany według poszukiwanych wartości.
 - Plik taki określany jest mianem indeksu.

341

Rola indeksu

- Dodatkowa struktura fizyczna reprezentacji danych.
- Celem jest przyspieszenie dostępu do danych.
- Tworzone są dla pojedynczych atrybutów lub ich grup.
- Atrybuty takie nazywane są atrybutami indeksowanymi.

342

Indeks

- Ułatwia dotarcie “na skróty” do miejsca fizycznego przechowywania danych.
- Skracają, często znacznie, czas wykonania zapytania.

343

Indeks

- Sortowawnie danych w bazie:
 - Binarne – według reprezentacji znaków.
 - Słownikowe – zależne od wielkości liter i przyjętego kodowania charakterystycznego dla określonego miejsca na świecie.
 - Wrażliwe (CS, AS) lub niewrażliwe (CI, AI) na wielkość liter i akcenty.

344

Indeksy wywiedzione z ograniczeń

- Tworzone są automatycznie przez system bazodanowy.
- Zależą od kluczy PK i wartości kluczy kandydujących (UNIQUE).
- Należy o nich pamiętać i nie duplikować ich w postaci indeksów tworzonych ręcznie.

345

Selektywność indeksu

- Poziom selektywności indeksu wynika z procentowego udziału wartości unikalnych w kolumnie.
- Wyższy procent udziału wartości unikalnych oznacza większa selektywność.
- Przekłada się na tworzenie indeksów odpowiedniego typu.
- Dla selektywności mniejszej niż 90-95% nie warto budować indeksów nieklastrowych.

346

Indeks klastrowy

- Tylko jeden.
- Domyślnie klucz główny jest tworzony z indeksem klastrowym.
- Jeżeli ma on dotyczyć innej kolumny niż klucz główny konieczne jest dyrektywy NONCLUSTERED.
- Zmiana wartości kolumny indeksowanej klastrowo powoduje jego przebudowę.
- Zalecane dla wyszukikań zakresów, agregacji min-max i zliczeń.

347

Indeks klastrowy

- Niezalecany gdy istnieje ryzyko wprowadzania danych w kolejności niesekwencyjnej do kolumny objętej indeksem.
- Zalecane jest stworzenie syntetycznego klucza głównego dodawanego w kolejności sekwencyjnej lub nie używanie indeksów klastrowych.

348

Indeks nieklastrowy

- Struktura danych przechowująca oddzielną kopię wybranych kolumn tabeli wraz z wskaźnikami (odwołaniami) do pełnych danych w tabeli.
- Tworzy osobną strukturę (np. drzewo B-tree), która przechowuje wartości klucza + wskaźnik (adres) do rekordu w tabeli.
- Tabela może mieć wiele indeksów nieklastrowych.
- Usprawnia wyszukiwanie, sortowanie i filtrowanie danych.
- Nie zmienia fizycznego porządku danych w tabeli.

349

Indeks nieklastrowy

- Zalety:
 - Przyspiesza zapytania selekcji z warunkami WHERE, ORDER BY, JOIN.
 - Może obejmować kilka kolumn (indeks złożony).
- Wady:
 - Zajmuje dodatkowe miejsce na dysku,
 - Spowalnia operacje INSERT, UPDATE, DELETE (bo trzeba aktualizować indeks).

350

CREATE INDEX

```
CREATE [UNIQUE | FULLTEXT | SPATIAL] INDEX nazwa_indeksu  
ON nazwa_tabeli (kolumna1 [ASC|DESC], kolumna2 [ASC|DESC], ...);
```

351

CREATE INDEX

- Wszystkie inne (INDEX, UNIQUE, FULLTEXT, SPATIAL) są indeksami nieklastrowymi.
- Poza UNIQUE wszystkie indeksy klastrowe dopuszczają duplikaty.
- Indeks FULLTEXT jest adresowany do typów tekstowych.
- Indeks SPATIAL do danych geolokalizacyjnych i przestrzennych.
- INDEX i UNIQUE są oparte na B-drzewie, FULLTEXT na indeksowaniu odwrotnym, a SPATIAL na R-drzewie.

352

Indeksy

- Kiedy warto tworzyć indeksy:
 - Kolumny używane często w warunkach WHERE.
 - Kolumny wykorzystywane w złączeniach.
 - Kolumny poddawane grupowaniu i porządkowaniu.
 - Kolumny przeszukiwane z warunkami porównania i zakresu.
 - Klucze obce.
 - Kolumny o wysokiej kardynalności.
 - Kolumny przeszukiwane pełnotekstowo.

353

Indeksy

- Kiedy nie warto tworzyć indeksu:
 - Dla tabel zawierających mało danych.
 - Dla tabel o niskiej zmienności.
 - Dla tabel, w których często wstawiasz, aktualizujesz lub usuwasz dane, bo każdy indeks wymaga wtedy aktualizacji.

354

Autonomiczność

- Istota transakcji.
- Zakłada wykonanie kilku kolejnych kwerend jako jednej i niepodzielnej całości.
- Poprawne wykonanie całości jest zależne od poprawnego wykonania każdej kwerendy składowej.

355

Transakcja

- Wykorzystywana w sytuacji gdy chcemy zachować spójność danych w bazie i mieć pewność, że powiązane kwerendy zostaną wykonane według reguły „wszystko albo nic”.
- Zabezpiecza to spójność danych w sytuacji awarii na poziomie zarówno sprzętowym jak i programowym.

356

Transakcja

- Ma jasno określona granicę.
- W przypadku pojedynczej kwerendy jest ona implicite nałożoną na nią.
- W przypadku wielu kwerend mających spełniać warunek wykonania autonomicznego granica musi być określona explicite.

357

Transakcja w systemie relacyjnym

- ACID:
 - Atomicity (Atomowość) – wszystko albo nic.
 - Consistency (Spójność) – pełna aktualizacja z uwzględnieniem powiązań i nałożonych ograniczeń.
 - Isolation (Izolacja) – każda transakcja jest izolowana od innych transakcji.
 - Durability (Trwałość) – po zakończeniu transakcji efekty są na stałe zapisywane w bazie.

358

Granice transakcji

- Początek.
- Punkt pośredni.
- Koniec lub cofnięcie transakcji.

359

Początek transakcji

- Wskazuje jednoznacznie punkt początkowy istnienia jednostki autonomicznej.
- Wszystko od tego punktu co nie zostanie zatwierdzone nie znajdzie się w danych bazy.

START TRANSACTION [<nazwa>|<@zmienna>]

360

Punkt pośredni transakcji

- Poprzez jego nazwę możemy cofnąć wykonanie transakcji do określonego miejsca

SAVE TRANSACTION [<nazwa>|<@zmienna>].

361

Koniec transakcji

- Dwa możliwe scenariusze:
 - Zamknięcie i zapisanie zmian do bazy.
 - Porzucenie zmian i pozostawienie bazy w stanie sprzed transakcji.

362

Zatwierdzenie transakcji

- Wszystkie zmiany wynikające z kwerend jednostki autonomicznej zostają utrwalone w danych bazy.
- Cofnięcie tej operacji jest możliwe tylko poprzez budowę odpowiedniej transakcji odwrotnej.

COMMIT [<nazwa>|<@zmienna>].

363

Cofnięcie transakcji

- Zmiany wprowadzone do danych zostają cofnięte do początku transakcji lub punktu pośredniego wskazywanego przez nazwę.

ROLLBACK [<nazwa>|<nazwa punktu pośredniego> | <@zmienna> | <@zmienna punktu pośredniego>].

Przykład

START TRANSACTION;

```
INSERT INTO student (imie, nazwisko, email, pesel, id_status)  
VALUES ('Jan', 'Nowak', 'jan@nowak.pl', '10192912312', 1);
```

```
INSERT INTO student_kierunek (id_student, id_kierunek)  
VALUES (LAST_INSERT_ID(), 2);
```

COMMIT;

365

Struktura przechowywania

- Dane składające się na bazę mogą znajdować się w:
 - Indeksach klastrowych lub tabelach zapisanych w stronach zakresów.
 - Buforze dziennika transakcji w postaci brudnych stron.

366

Kontrola

- Proces zapisu brudnych stron (zawierających dane zmienione poprzez transakcję) w bazie.
- Odbywa się w określonych odstępach czasu bez ingerencji ze strony użytkownika.

367

Wywołanie kontroli

- Kwerenda CHECKPOINT:
 - Zakończenie pracy serwera (o ile nie użyto NOWAIT).
 - Przy zmianie trybu działania bazy.
 - Przy użyciu prostego odtwarzania gdy dziennik jest wypełniony w minimum 70%.
 - Gdy ilość danych w dzienniku przekracza rozmiar, który może być obsłużony w przedziale czasu przeznaczonym na odtwarzanie.

368

Odtwarzanie

- Ma miejsce przy każdym uruchomieniu serwera niezależnie od sposobu zakończenia jego pracy w poprzedniej sesji.
- Zapisuje wszystkie zatwierdzone zmiany w dzienniku od czasu ostatniej kontroli.
- Niezatwierdzone zmiany zostają cofnięte.

369

Domniemanie transakcji

- W systemach DB2 i Oracle transakcje rozpoczynają się automatycznie wraz z wykonaniem pierwszej kwerendy modyfikującej dane i nie wymagają jawnego polecenia rozpoczęcia transakcji.
- Transakcja kończy się w momencie jej zatwierdzenia, cofnięcia lub rozpoczęcia nowej transakcji.
- W MySQL i Microsoft SQL Server w domyślnej konfiguracji każda pojedyncza kwerenda jest traktowana jako osobna, automatycznie zatwierdzana transakcja.

370

Współbieżność

- Systemy bazodanowe są współbieżne.
- Wiele równoległych transakcji i wielu użytkowników jednocześnie operuje na tych samych danych.
- Wymaga to wprowadzenia mechanizmów konkurencyjności.
- Realizuje się to stosując mechanizmy blokowania.

371

Problem czytelników i pisarzy

- Wiele osób może czytać równocześnie dane.
- Jedna osoba może zmieniać dane.
- Nikt nie może odczytywać danych, które właśnie ulegają zmianie.

372

Izolowanie transakcji

- Grupa mechanizmów zapobiegających w bazach danych:
 - Brudnym odczytom.
 - Niepowtarzalnym odczytom.
 - Fantomom.
 - Utraconym aktualizacjom.

373

Brudne odczyty

- Mają miejsce gdy transakcja odczytuje dane zmienione przez inną i niezakończoną jeszcze transakcją.
- Możliwe jest wtedy odczytanie wartości danych, które mogą ulec odwołaniu.
- Zapobiega temu izolowanie przez ustawienie w systemie zabezpieczenia READ COMMITTED.

374

Niepowtarzalne odczyty

- Mają miejsce gdy w transakcji wielokrotnie odczytujemy te same dane, których wartość w międzyczasie uległa zmianie przez inną transakcję.
- Zapobiega temu ustawienie poziomu izolacji transakcji na REPEATABLE READ lub SERIALIZABLE.

375

Fantomy

- Ma miejsce gdy zmiana poprzez UPDATE nie zostaje dla wierszy wykonana poprawnie.
- Bardzo rzadka sytuacja gdy podczas transakcji zmieniającej wartości wywołana zostanie równoległe transakcja wstawiająca nowe wiersze.
- Zapobiega temu ustawienie poziomu izolacji transakcji na **SERIALIZABLE**.

376

Utrata aktualizacji

- Ma miejsce gdy dwie operacje odczytają te same dane, dokonają ich aktualizacji i zapisania w bazie.
- Trudne do ominięcia gdy transakcja nie jest ciągła.

377

Zasoby podlegające blokowaniu

- Baza danych (tylko zmiana schematu).
- Tabela i wszelkie obiekty powiązane.
- Zakres (zazwyczaj wszystkie strony).
- Strona i wszystkie dane lub klucze na niej się znajdujące.
- Klucz.
- Wiersz.

378

Rozwijanie blokad

- Zależy od liczby blokowanych elementów.
- Zarządza nim menedżer blokad dbając o odpowiednią granulację i czas trwania.
- Gdy liczba utrzymywanych blokad osiągnie określoną wartość blokada jest rozwijana na wyższy poziom (wiersz->strona->zakres->...).

379

Blokady współdzielone

- Używane przy odczytach.
- Zapobiegają brudnym odczytom.
- Nie pozwalają na zapis do momentu ich zniesienia.

380

Blokady wyłączone

- Nie pozwalają na dostęp do danych nikomu poza transakcją nakładającą blokadę.
- Dbają o wyłączny zapis.

381

Blokady aktualizacji

- Pozwalają na sekwencję odczytu (klauzula WHERE w ramach UPDATE) i zapisu tych samych danych (SET) dla danej transakcji.
- Zabezpiecza przed zakleszczeniem (deadlock) czyli problemem uwolnieniu zasobu z blokady ze względu na nałożenie blokady równoległej.

382

Inne typy blokad

- Zamierzone, które obsługują hierarchię na poziomie strony i tabeli zwalniając z konieczności sprawdzania czy jest nałożona blokada na wiersz.
- Schematu, które zapobiegają modyfikacji schematów bazy.
- Aktualizacji masowej, które pozwala na równoległe ładowanie danych do bazy

383

Poziom izolowania

- Określa sposób nakładania izolacji na transakcję.
- Zabezpiecza przed wystąpieniem błędów transakcji.
- Im bardziej restrykcyjny tym wydajność systemu mniejsza.

384

Poziom izolowania

- READ UNCOMMITTED.
- READ COMMITTED (domyślny).
- REPEATABLE READ.
- SERIALIZABLE.

385

Zalecenia do budowy transakcji

- Utrzymuj transakcje najkrócej jak to możliwe.
- Używaj najwyższego z najniższych akceptowalnych poziomów izolowania.
- Nie pozostawiaj transakcji otwartych.

386

Procedury składowane

- Persistent Stored Modules:
 - Część standardu SQL (SQL/PSM), która określa, jak tworzyć i wykonywać proceduralny kod w bazie danych w tym procedury, funkcje, wyzwalacze i kursory.
 - W MySQL PSM jest częściowo zaimplementowany, co oznacza, że możesz używać większości składni proceduralnej:
 - Instrukcji sterujące.
 - Zmienne.
 - Obsługi wyjątków.
 - Transakcji.

387

Schemat definiowania

DELIMITER \$\$

CREATE PROCEDURE nazwa_procedury([parametry])

BEGIN

-- instrukcje SQL

END\$\$

DELIMITER ;

388

Procedury składowane

- DELIMITER \$\$ zmienia separator poleceń, by MySQL nie zakończył procedury po pierwszym średniku.
- Parametry mogą być:
 - IN – dane wejściowe.
 - OUT – dane wyjściowe.
 - INOUT – jednocześnie wejście i wyjście.

389

Procedury składowane

- DELIMITER \$\$ zmienia separator poleceń, by MySQL nie zakończył procedury po pierwszym średniku.
- Parametry mogą być:
 - IN – dane wejściowe.
 - OUT – dane wyjściowe.
 - INOUT – jednocześnie wejście i wyjście.

390

Przykład

DELIMITER \$\$

```
CREATE PROCEDURE znajdz_studenta(IN pesel_in VARCHAR(11))
BEGIN
    SELECT imie, nazwisko, email
    FROM student
    WHERE pesel = pesel_in;
END$$
```

DELIMITER ;

Przykład

```
CALL znajdz_studenta('64418979671');
```

imie	nazwisko	email
Ozella	Howell	emma10@example.org

392

Przykład

DELIMITER \$\$

```
CREATE PROCEDURE policz_studentow(IN id_kierunek_in INT, OUT liczba INT)
BEGIN
    SELECT COUNT(*) INTO liczba
    FROM student_kierunek
    WHERE id_kierunek = id_kierunek_in;
END$$
```

DELIMITER ;

393

Przykład

```
CALL policz_studentow(2, @wynik);  
SELECT @wynik;
```

@wynik
56

394

Zalety procedur składowanych

- Redukują ruch między aplikacją a bazą (logika po stronie serwera),
- Zwiększają spójność i bezpieczeństwo,
- Pozwalają na kontrolę transakcji (START TRANSACTION, COMMIT, ROLLBACK),
- Ułatwiają ponowne użycie kodu.

395

PSM w MySQL - zmienne lokalne

```
DECLARE liczba INT DEFAULT 0;  
SET liczba = 10;
```

396

PSM w MySQL - pętla WHILE

```
SET licznik = 1;  
WHILE licznik <= 5 DO  
    INSERT INTO log VALUES (licznik);  
    SET licznik = licznik + 1;  
END WHILE;
```

397

PSM w MySQL - pętla LOOP

```
SET i = 1;  
petla: LOOP  
    IF i > 5 THEN  
        LEAVE petla;  
    END IF;  
    SET i = i + 1;  
END LOOP petla;
```

398

PSM w MySQL - instrukcja CASE

CASE wartosc

```
WHEN 1 THEN SET opis = 'pierwszy';  
WHEN 2 THEN SET opis = 'drugi';  
ELSE SET opis = 'inny';  
END CASE;
```

399

PSM w MySQL - obsługa wyjątku

```
DECLARE CONTINUE HANDLER FOR SQLEXCEPTION
```

```
BEGIN
```

```
    INSERT INTO log_bledu VALUES (NOW(), 'Błąd w procedurze');
```

```
END;
```

400

Przykład

DELIMITER \$\$

```
CREATE PROCEDURE dodaj_studenta  
( IN p_imie VARCHAR(50), IN p_nazwisko VARCHAR(50), IN p_email VARCHAR(100), IN p_pesel VARCHAR(11), IN  
p_id_kierunek INT )
```

BEGIN

```
    DECLARE nowy_id INT;  
    START TRANSACTION;  
    INSERT INTO student (imie, nazwisko, email, pesel) VALUES (p_imie, p_nazwisko, p_email, p_pesel);  
    SET nowy_id = LAST_INSERT_ID();  
    INSERT INTO student_kierunek (id_student, id_kierunek) VALUES (nowy_id, p_id_kierunek);  
    COMMIT;
```

END\$\$

401

DELIMITER ;

Bezpieczeństwo MySQL

- Bezpieczeństwo serwera MySQL obejmuje:
 - Silne hasła i nadawanie użytkownikom tylko niezbędnych uprawnień.
 - Ochronę instalacji, czyli plików danych, dzienników i aplikacji przed nieautoryzowanym dostępem.
 - Kontrolę dostępu w bazie danych (użytkownicy, uprawnienia, widoki, procedury).
 - Bezpieczeństwo sieciowe i prawidłową konfigurację uprawnień użytkowników.
 - Regularne kopie zapasowe danych, konfiguracji i logów.

402

Bezpieczeństwo MySQL

- Wytyczne ogólne:
 - Dostęp do tabeli użytkowników ma tylko konto root.
 - Zarządzaj uprawnieniami przez GRANT i REVOKE.
 - Przyznawaj tylko niezbędne uprawnienia i nigdy wszystkim hostom.
 - Nie przechowuj haseł w postaci jawnej, zawsze używaj funkcji haszujących, np. SHA2().

403

Bezpieczeństwo MySQL

- Dodatkowe mechanizmy:
 - Używaj firewalla, który blokuje większość typowych ataków.
 - Umieść MySQL za zaporą lub w DMZ, odizolowaną od sieci publicznej.
 - Zablokuj port 3306 dla niezaufanych hostów.

404

Bezpieczeństwo MySQL

- Zalecenia dla obsługi haseł:
 - Hasła użytkowników są przechowywane w **mysql.user**, do którego dostęp tylko dla administratora.
 - Hasła mogą wygasać i użytkownicy muszą je okresowo zmieniać.
 - Wtyczka **validate_password** wymusza politykę silnych haseł.
 - Chroń katalog wtyczek (**plugin_dir**) i plik **my.cnf**, by uniemożliwić podmianę wtyczek.
 - Należy zabezpieczać pliki logów, które mogą zawierać hasła.

405

Bezpieczeństwo MySQL

- MySQL umożliwia tworzenie kont użytkowników z określonymi uprawnieniami do danych i operacji administracyjnych.
- System uprawnień odpowiada za uwierzytelnianie użytkownika i kontrolę dostępu do poleceń (SELECT, INSERT, UPDATE, DELETE).
- Dostęp określają instrukcje CREATE USER, GRANT i REVOKE.
- Uprawnienia zależą od nazwy użytkownika i hosta, z którego następuje połączenie.

406

Bezpieczeństwo MySQL

- MySQL identyfikuje użytkowników po nazwie i hoście, ponieważ ta sama nazwa może oznaczać różne osoby na różnych komputerach.
- Informacje o uprawnieniach są przechowywane w tabelach systemowych bazy mysql.
- Przy uruchomieniu serwer ładuje te dane do pamięci i używa ich do kontroli dostępu.

407

Bezpieczeństwo MySQL

- Kontrola dostępu MySQL przebiega w dwóch etapach:
 - Uwierzytelnienie, podczas którego serwer sprawdza tożsamość użytkownika i hasło.
 - Autoryzacja mająca miejsce przy każdej instrukcji weryfikowane są odpowiednie uprawnienia (np. SELECT, DROP).

408

Bezpieczeństwo MySQL

- Nazwy użytkowników MySQL są niezależne od kont systemowych (Windows/Unix).
- Domyślna nazwa logowania to bieżący użytkownik systemu, ale można ją zmienić.
- Każdy może podać dowolną nazwę użytkownika, dlatego wszystkie konta MySQL muszą mieć ustawione hasła.

409

Bezpieczeństwo MySQL

- Nazwy użytkowników MySQL mogą mieć do 32 znaków (niezależnie od limitów systemowych).
- Przy logowaniu MySQL używa haseł z tabeli mysql.user, chyba że zastosowano inną wtyczkę uwierzytelniającą.
- W przypadku wtyczek możliwe jest użycie zewnętrznego hasła do logowania.

410

Bezpieczeństwo MySQL

- Uprawnienia MySQL określają dozwolone operacje konta.
- Dzielią się na poziomy:
 - Administracyjne – globalne, dotyczą pracy serwera.
 - Baz danych – dla wybranej lub wszystkich baz.
 - Obiektów – dla konkretnych tabel, widoków, indeksów lub procedur.

411

Bezpieczeństwo MySQL

- Uprawnienia MySQL określają dozwolone operacje konta.
- Dzielią się na poziomy:
 - Administracyjne – globalne, dotyczą pracy serwera.
 - Baz danych – dla wybranej lub wszystkich baz.
 - Obiektów – dla konkretnych tabel, widoków, indeksów lub procedur.

412

Bezpieczeństwo MySQL

- Uprawnienia mogą być statyczne (wbudowane) lub dynamiczne (dodawane w czasie działania).
- Typ uprawnienia wpływa na jego dostępność dla użytkowników i ról.

413

Bezpieczeństwo MySQL

- Przyznawaj tylko niezbędne uprawnienia.
- Zachowaj ostrożność przy uprawnieniach administracyjnych:
 - FILE – umożliwia odczyt dowolnych plików serwera.
 - GRANT OPTION – pozwala przekazywać uprawnienia innym.
 - ALTER – może zmieniać system uprawnień.
 - SHUTDOWN – pozwala wyłączyć serwer.
 - PROCESS – ujawnia treść wykonywanych zapytań.
 - SUPER – umożliwia kończenie sesji i modyfikację działania serwera.

414

Bezpieczeństwo MySQL

- Nazwy kont używane w CREATE USER, GRANT, SET PASSWORD mają format 'użytkownik'@'host'.
- Część '@'host' jest opcjonalna – sam 'użytkownik' oznacza 'użytkownik'@'% '.
- Używaj cudzysłowów, jeśli nazwa zawiera znaki specjalne.
- CURRENT_USER() zwraca nazwę bieżącego użytkownika i hosta.

415

Bezpieczeństwo MySQL

- Role MySQL to nazwane zbiory uprawnień.
- Mają taką samą składnię i zasady jak nazwy kont użytkowników i są przechowywane w tych samych tabelach uprawnień.

416

Bezpieczeństwo MySQL

- Role nie mogą być anonimowe (część użytkownika nie może być pusta).
- Pominięcie hosta oznacza wartość '%', ale bez funkcji symbolu wieloznacznego.
- Maski sieci w części hosta roli są ignorowane.

417

Bezpieczeństwo MySQL

- Wiersz w tabeli systemowej **mysql.user** może służyć zarówno jako konto, jak i jako rola.
- W takim przypadku żadne specjalne właściwości dopasowania nazwy użytkownika lub hosta nie mają zastosowania w kontekstach, w których nazwa jest używana jako nazwa roli.

418

Bezpieczeństwo MySQL

- Rola to nazwany zestaw uprawnień.
- Role można nadawać i odbierać jak uprawnienia.
- Konto może mieć wiele ról, co ułatwia zarządzanie dostępem bez przypisywania pojedynczych uprawnień.

419

Bezpieczeństwo MySQL

- Polecenie CREATE ROLE służy do utworzenia jednej lub kilku ról.
- DROP ROLE usuwa jedną lub więcej ról.

420

Bezpieczeństwo MySQL

- CREATE USER dodaje dla każdego konta nowy wiersz w tabeli mysql.user, zawierający określone atrybuty.
- Właściwości niepodane w poleceniu przyjmują wartości domyślne:
 - Uwierzytelnianie: domyślna wtyczka, brak poświadczeń.
 - Rola domyślna: NONE.
 - SSL/TLS: brak wymagań.
 - Limity zasobów: nieograniczone.
 - Zarządzanie hasłami: wszystkie opcje DEFAULT (EXPIRE, HISTORY, REUSE INTERVAL, REQUIRE CURRENT).
 - Śledzenie logowań: wyłączone.
 - Stan konta: ACCOUNT UNLOCK.

421

Bezpieczeństwo MySQL

- ALTER USER służy do modyfikacji istniejących kont MySQL.
- Pozwala zmieniać ustawienia uwierzytelniania, ról, SSL/TLS, limitów zasobów, haseł, komentarzy i atrybutów.
- Może także blokować lub odblokowywać konta.

422

Bezpieczeństwo MySQL

- GRANT nadaje uprawnienia i role kontom oraz rolom MySQL.
- Pozwala administratorom nadawać uprawnienia lub role użytkownikom i rolom MySQL.
- Nie można mieszać uprawnień i ról w jednej instrukcji i każdą należy nadać osobno.
- Uprawnienia i role można przyznawać temu samemu kontu, ale za pomocą oddzielnych poleceń GRANT.
- Obecność klauzuli ON oznacza nadawanie uprawnień, a jej brak nadawanie ról.

423

Bezpieczeństwo MySQL

- REVOKE pozwala administratorom cofać uprawnienia i role nadane kontom lub rolom MySQL.
- Cofnięcie roli działa natychmiast, a bieżące sesje użytkownika zostają zaktualizowane przy następnym poleceniu.
- Odwołanie roli usuwa samą rolę, a nie indywidualne uprawnienia, które ona obejmuje.

424

Bezpieczeństwo MySQL

- SET PASSWORD ustawia lub zmienia hasło użytkownika MySQL.
- Hasło można podać ręcznie lub pozwolić MySQL wygenerować je automatycznie.
- Polecenie może zawierać klauzulę weryfikacji bieżącego hasła oraz opcję określającą, czy konto ma mieć dodatkowe hasło.

425

Przykład

```
CREATE USER 'rafal'@'localhost' IDENTIFIED BY 'BardzoSilneHaslo123!';
```

```
CREATE ROLE 'rola_studenta', 'rola_administratora';
```

```
GRANT SELECT, INSERT ON uczelnia.student TO 'rola_studenta';
```

```
GRANT ALL PRIVILEGES ON uczelnia.* TO 'rola_administratora';
```

```
GRANT 'rola_administratora' TO 'rafal'@'localhost';
```

426

Kopie zapasowe

- Umożliwiają odzyskanie danych po awarii systemu, sprzętu lub przypadkowym usunięciu informacji.
- Są niezbędne przed aktualizacją MySQL, aby uniknąć utraty danych.
- Ułatwiają migrację instalacji na inny serwer oraz konfigurację replikacji.

427

Kopie zapasowe

- Typy backupów dla MySQL:
 - Fizyczny/Logiczny.
 - Online/Offline.
 - Lokalny/Zdalny.
 - Migawki.
 - Pełny/Przyrostowy.

428

Kopie zapasowe

- Fizyczne kopie zapasowe to surowe kopie plików i katalogów bazy danych.
- Stosowane przy dużych i krytycznych bazach, gdzie liczy się szybkie przywracanie.
- Logiczne kopie zapasowe zapisują strukturę (CREATE DATABASE, CREATE TABLE) i dane (INSERT lub pliki tekstowe).
- Sprawdzają się przy mniejszych bazach, gdy potrzebna jest edycja danych, struktury lub odtworzenie na innej platformie.

429

Kopie zapasowe

- Kopie zapasowe online (gorące) wykonywane podczas działania serwera MySQL, pozwalają na dostęp do danych w trakcie pracy.
- Kopie zapasowe offline (zimne) tworzone po zatrzymaniu serwera.
- Kopie ciepłe tworzy się gdy serwer działa, ale dane są zablokowane przed modyfikacją podczas tworzenia kopii.

430

Kopie zapasowe

- Niektóre systemy plików obsługują migawki (snapshots) to logiczne kopie stanu danych w danym momencie.
- Wykorzystują technikę copy-on-write, kopiując tylko zmienione fragmenty po utworzeniu migawki.
- MySQL nie posiada wbudowanej funkcji migawek, ale można je tworzyć przy użyciu narzędzi zewnętrznych, takich jak Veritas, LVM czy ZFS.

431

Kopie zapasowe

- Pełna kopia zapasowa obejmuje wszystkie dane przechowywane przez serwer MySQL w danym momencie.
- Przyrostowa kopia zapasowa zawiera tylko zmiany dokonane między dwoma punktami w czasie.
- MySQL umożliwia tworzenie pełnych kopii różnymi metodami.
- Kopie przyrostowe można wykonywać dzięki dziennikowi binarnemu, który zapisuje wszystkie modyfikacje danych.

432

Kopie zapasowe

- Narzędzia do backupów MySQL:
 - mysqldump – tworzy logiczne kopie zapasowe (struktura + dane w SQL); proste, uniwersalne.
 - mysqlpump – nowsze narzędzie, szybsze, obsługuje równoległy eksport i eksport uprawnień.
 - mysqlhotcopy – wykonuje fizyczne kopie plików (tylko dla MyISAM).
 - mysqlbackup – (Enterprise) pełne, różnicowe i przyrostowe kopie fizyczne, z kompresją i szyfrowaniem.
 - Percona XtraBackup – darmowe narzędzie open source do fizycznych i przyrostowych backupów InnoDB.
 - Dzienniki binarne (binlog) – rejestrują wszystkie zmiany w danych, używane do backupów przyrostowych i replikacji.
 - Zewnętrzne rozwiązania – LVM, ZFS, Veritas – migawki systemu plików.

433

Dziękuję za uwagę