

Lecture 2

October 13, 2019

```
[3]: %%HTML
<style>
.rendered_html table, .rendered_html th, .rendered_html tr, .rendered_html td {
    font-size: 80%;
}
</style>
```

<IPython.core.display.HTML object>

1 Analiza i bazy danych

1.1 Bazy danych

1.1.1 mgr inż. Waldemar Bauer

1.2 Narzędzia używane na zajęciach



1.3 Dlaczego używamy baz danych

- Organizują dane w ustandaryzowany sposób
- System organizacji danych odpowiada ludzkiej logice
- Umożliwia szybkie przeszukiwanie, łączenie i filtrowanie danych
- Ułatwia dzielenie się danymi przez ich użytkowników

1.4 Definicja bazy danych

Nieformalnie: baza danych to dane i specjalny program umożliwiający ich gromadzenie i przetwarzanie.

Formalnie: Baza danych jest zbiorem powiązanych informacji zorganizowanych w strukturę umożliwiającą łatwe przetwarzanie.

Definicja prawna: Zależna od organizacji lub kraju

1.5 Rodzaje baz danych

Bazy proste: - Kartotekowe - Hierarchiczne

Bazy złożone: - Relacyjne - Obiektowe - Relacyjno-obiektowe - Strumieniowe - Temporalne - Nierelacyjne (NoSQL)

1.6 Przetwarzanie danych w bazie danych

- Wprowadzanie danych,
- Zapisywanie i przechowywanie danych,
- Wyszukiwanie i prezentacja danych,
- Wstawianie i usuwanie danych,
- Aktualizacje danych,
- Przetwarzanie arytmetyczne, statystyczne, algebraiczne i logiczne,
- Operacje oparte na algebrze relacji,
- Kodowanie i dekodowanie danych

1.7 Cele systemu bazodanowego

- Centralizacja
- Dostęp selektywny
- Utrzymanie spójności danych
- Normalizacja
- Dostęp wielokrotny
- Widoki
- Dostęp równoległy
- Niezależność danych
- Optymalizacja

1.8 Database Management System (DMS)

- Zestaw programów do zarządzania bazą danych

- Systemy open source: PostgreSQL, MySQL, Hadoop, Spark, Firebird,...
- Systemy handlowe: Oracle, IBM DB, MS SQL Server,...
- Systemy biurowe: OOBase, MS Access,...

1.9 Funkcjonalności DMS

- Umożliwienie zaprojektowania i wdrożenia nowej bazy danych przy użyciu narzędzi i Data Definition Language (DDL).
- Umożliwienie selektywnego dostępu do danych przy użyciu języka zapytań (query).
- Wykonanie określonych operacji na danych przy użyciu języka manipulacji danych (Data Manipulation Language, DML).
- Obsługa przechowywania dużych zestawów danych, zapewnienie niezawodności i wydajności obsługi operacji na danych zapisanych na dysku.
- Zapewnienie integralności danych (na poziomie tabeli, na poziomie bazy danych).

1.10 Funkcjonalności DMS - cd.

- Ochrona dostępu do danych - różne obszary i poziomy dostępu (prowadzenie kont użytkowników, przydzielanie uprawnień, tworzenie widoków).
- Zapewnianie dostępu dla wielu użytkowników i synchronizacja dostępu w przypadku dostępu kooperacyjnego.
- Zapewnienie możliwości komunikacji z innymi systemami.
- Dostarczenie opisu i dokumentacji (schemat i struktura).
- Optymalizacja pracy (minimalizacja czasu dostępu lub obsługi zadań), optymalizacja dostępu dla poszczególnych użytkowników, optymalizacja zarządzania zasobami i organizacją bazy danych.

1.11 Realcyjny model bazy danych

- Twórca Edgar Frank Codd, 1970, IBM
- Model organizacji danych bazujący na matematycznej teorii mnogości, w szczególności na pojęciu relacji.
- Dane są przedstawione w postaci relacyjnej.
- Relacje jest reprezentowan przez tabelę.

1.12 12 Postulatów Codd

0. System musi być kwalifikowany jako relacyjny, jako baza danych i jako system zarządzania
1. Postulat informacyjny - dane są reprezentowane jedynie poprzez wartości atrybutów w wierszach tabel,

2. Postulat dostępu - każda wartość w bazie danych jest dostępna poprzez podanie nazwy tabeli, atrybutu oraz wartości klucza podstawowego,
3. Postulat dotyczący wartości NULL - dostępna jest specjalna wartość NULL dla reprezentacji wartości nieokreślonej jak i nieadekwatnej, inna od wszystkich i podlegająca przetwarzaniu

1.13 12 Postulatów Codda cd

4. Postulat dotyczący katalogu - wymaga się, aby system obsługiwał wbudowany katalog relacyjny z bieżącym dostępem dla uprawnionych użytkowników używających języka zapytań,
5. Postulat języka danych - system musi dostarczać pełnego języka przetwarzania danych, który może być używany w trybie interaktywnym jak i w obrębie programów aplikacyjnych, obsługuje operacje definiowania danych, operacje manipulowania danymi, ograniczenia związane z bezpieczeństwem i integralnością oraz operacje zarządzania transakcjami,
6. Postulat modyfikowalności widoków - system musi umożliwiać modyfikowanie widoków,
7. Postulat modyfikowalności danych - system musi umożliwiać operacje modyfikacji danych, musi obsługiwać operatory INSERT, UPDATE oraz DELETE,

1.14 12 Postulatów Codda cd

8. Postulat fizycznej niezależności danych - zmiany fizycznej reprezentacji danych i organizacji dostępu nie wpływają na aplikacje,
9. Postulat logicznej niezależności danych - zmiany wartości w tabelach nie wpływają na aplikacje,
10. Postulat niezależności ograniczeń - ograniczenia są definiowane w bazie i nie zależą od aplikacji,
11. Postulat niezależności dystrybucyjnej - działanie aplikacji nie zależy od modyfikacji i dystrybucji bazy,
12. Postulat bezpieczeństwa względem operacji niskiego poziomu - operacje niskiego poziomu nie mogą naruszać modelu relacyjnego i więzów spójności.