

Dokumentacja projektu zaliczeniowego

Kurs: Bazy Danych

1. Spis użytych technologii

W projekcie wykorzystano następujące technologie:

- **Bazy danych:**
 - MariaDB — baza danych uruchomiona na serwerze `giniewicz.it`
 - SQLTools, ERD Editor — rozszerzenia do Visual Studio Code
- **Generowanie danych (Python):**
 - `mysql.connector`, `random`, `math`
 - `faker`, `unicodedata`, `datetime`
- **Analiza danych (Python):**
 - `mysql.connector`, `numpy`, `pandas`
 - `matplotlib`, `seaborn`, `tabulate`

2. Lista plików i opis ich zawartości

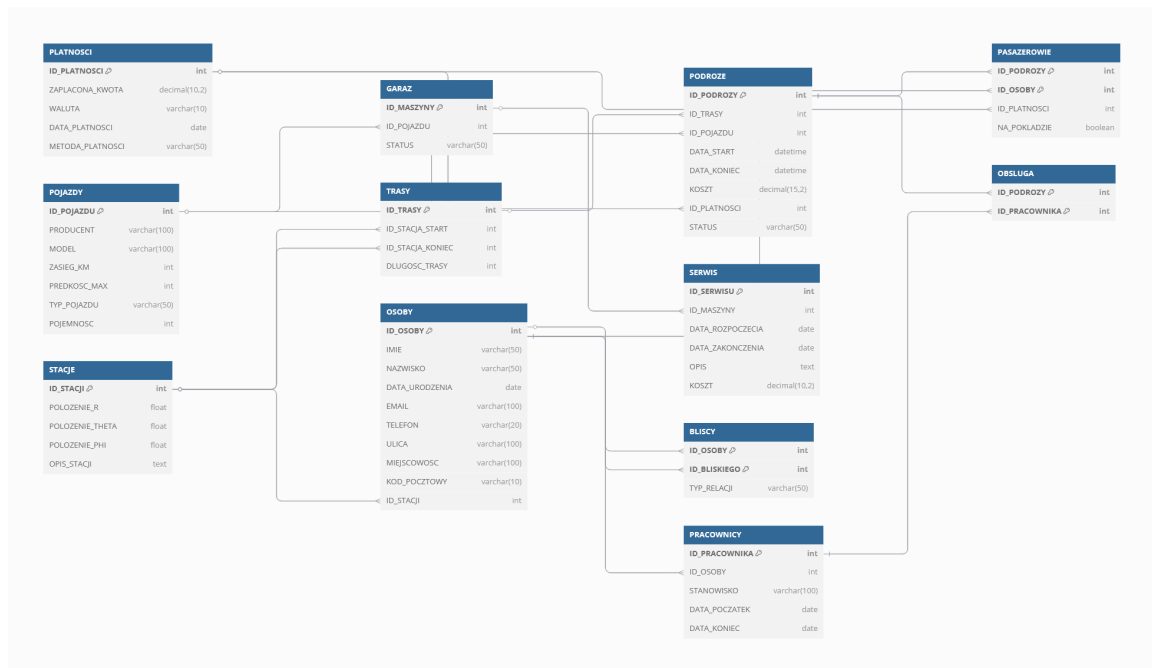
- `struktura.sql` — definicja struktury bazy danych z kluczami obcymi
- `generowanie_danych.ipynb` — notebook Pythona do generowania danych testowych
- `analiza_danych.ipynb` — notebook analizujący dane w bazie
- `raport.html`, `raport.pdf` — raport podsumowujący analizę

3. Kolejność uruchamiania

Aby otrzymać działający projekt:

1. Uruchomić `struktura.sql` na serwerze bazy danych
2. Wygenerować dane za pomocą `generowanie_danych.ipynb`
3. Przeanalizować dane przy użyciu `analiza_danych.ipynb`
4. Wyeksportować raport do pliku PDF lub HTML

4. Schemat bazy danych



5. Zależności funkcyjne

Dla każdej tabeli:

- **OSOBY:** $ID_OSOBY \rightarrow IMIE, NAZWISKO, DATA_URODZENIA, EMAIL, TELEFON, ULICA, MIEJSCOWOSC, KOD_POCZTOWY, ID_STACJI$
- **PRACOWNICY:** $ID_PRACOWNIKA \rightarrow ID_OSOBY, STANOWISKO, DATA_POCZATEK, DATA_KONIEC$
- **BLISCY:** $(ID_OSOBY, ID_BLISKIEGO) \rightarrow TYP_RELACJI$
- **PASAZEROWIE:** $(ID_PODROZY, ID_OSOBY) \rightarrow ID_PLATNOSCI, NA_POKLADZIE$
- **PLATNOSCI:** $ID_PLATNOSCI \rightarrow ZAPLACONA_KWOTA, WALUTA, DATA_PLATNOSCI, METODA_PLATNOSCI$
- **PODROZE:** $ID_PODROZY \rightarrow ID_TRASY, ID_POJAZDU, DATA_START, DATA_KONIEC, KOSZT, ID_PLATNOSCI, STATUS$
- **OBSLUGA:** $(ID_PODROZY, ID_PRACOWNIKA) \rightarrow \text{brak}$
- **TRASY:** $ID_TRASY \rightarrow ID_STACJA_START, ID_STACJA_KONIEC, DLUGOSC_TRASY$
- **STACJE:** $ID_STACJI \rightarrow POLOZENIE_R, POLOZENIE_THETA, POLOZENIE_PHI$
- **POJAZDY:** $ID_POJAZDU \rightarrow PRODUCENT, MODEL, ZASIEG_KM, PREDKOSC_MAX, TYP_POJAZDU, POJEMNOSC$

- **GARAZ:** ID_MASZYNY \rightarrow ID_POJAZDU, STATUS
- **SERWIS:** ID_SERWISU \rightarrow ID_MASZYNY, DATA_ROZPOCZECIA, DATA_ZAKONCZENIA, OPIS, KOSZT

6. Uzasadnienie, że baza spełnia EKNF

- Każda tabela posiada klucz główny jednoznacznie identyfikujący wiersz.
- Wszystkie zależności funkcyjne są postaci: klucz główny \rightarrow pozostałe atrybuty.
- Brak zależności przechodnich i zależności od atrybutów niekluczowych.
- Dane zostały znormalizowane (np. osobne tabele dla stacji, płatności, tras).

7. Największe trudności

- Stworzenie spójnej struktury odpowiadającej rzeczywistej organizacji
- Wygenerowanie realistycznych danych testowych
- Wybór interesujących aspektów do analizy danych