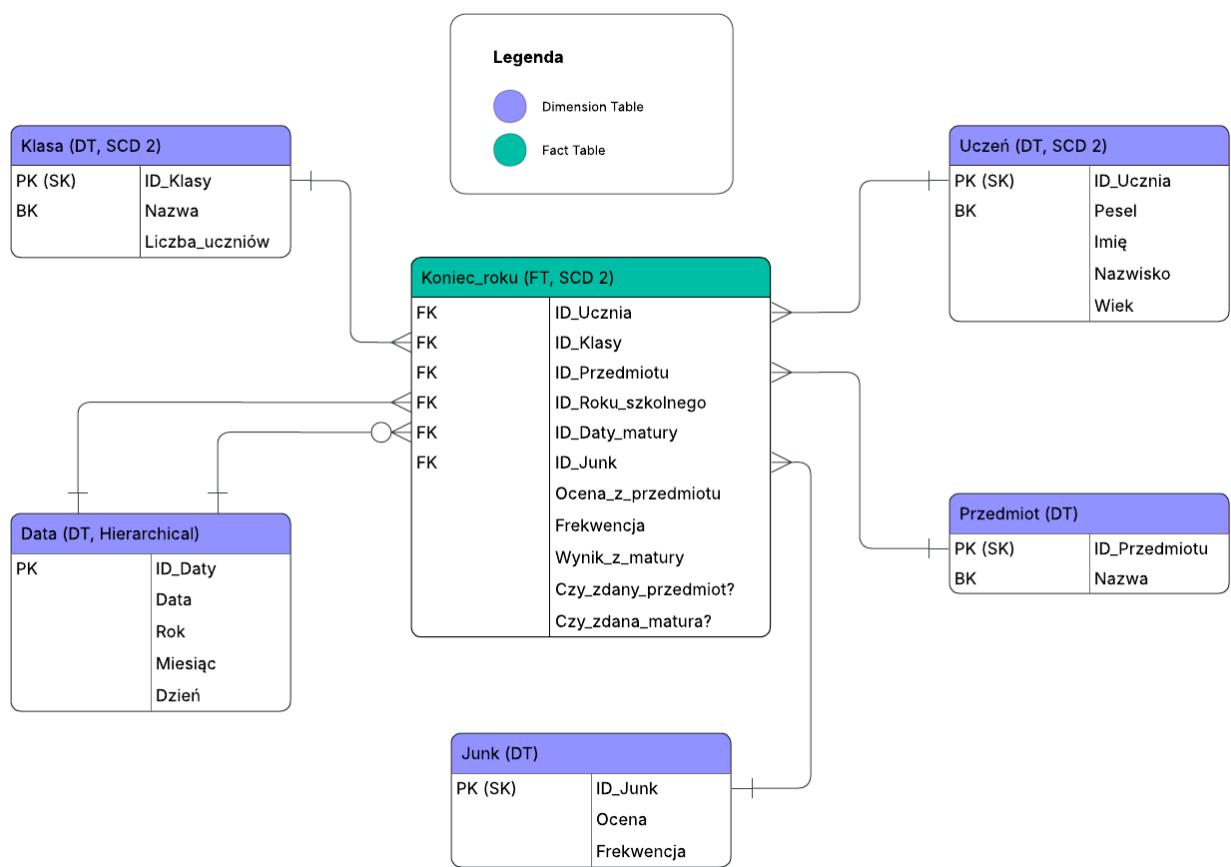


Design hurtowni danych

Jakub Nowak 197860, Piotr Staszko 197938

1. Proces biznesowy

Realizacja zajęć dydaktycznych i przygotowanie do egzaminu maturalnego – nauczanie



2. Model relacyjny hurtowni danych

3. Opis tabel

Nazwa tabeli	Atrybut	Typ atrybutu	Opis
Koniec_roku (FT)	Jedna krotka oznacza rezultat pracy ucznia z danego przedmiotu w danym roku szkolnym. Uwzględnia podejście do egzaminu maturalnego.		
	ID_Ucznia	Liczbowy	FK Uczeń, unikalny identyfikator ucznia

Nazwa tabeli	Atrybut	Typ atrybutu	Opis
	ID_Klasy	Liczbowy	FK Klasa, unikalny identyfikator klasy
	ID_Predmiotu	Liczbowy	FK Przedmiot, unikalny identyfikator przedmiotu
	ID_Roku_szkolnego	Liczbowy	FK Data, identyfikator do daty, roku, którego dotyczy zakończenie roku szkolnego
	ID_Daty_matury	Liczbowy	FK Data, identyfikator do daty, w której napisany został egzamin maturalny
	ID_Junk	Liczbowy	FK Junk, unikalny identyfikator do tabeli Junk
	Ocena_z_predmiotu	1 cyfra	Ocena wystawiona z przedmiotu na koniec roku Szkolnego. Możliwe wartości z przedziału 1-6
	Frekwencja	Liczbowy	Informacja o ilości obecności ucznia wyrażona procentowo
	Wynik_z_matury	Liczbowy	Wynik z matury z konkretnego przedmiotu wyrażony procentowo
Uczeń (DT)	<i>Jedna krotka opisuje konkretnego ucznia w szkole</i>		
	ID_Ucznia	Liczbowy	PK (surrogate key), unikalny identyfikator ucznia
	Pesel	11 cyfr	Pesel ucznia
	Imię	Tekst - 20	Imię ucznia
	Nazwisko	Tekst - 30	Nazwisko ucznia
	Wiek	Liczbowy	Wiek ucznia obliczany za pomocą pierwszych dwóch cyfr peselu
Klasa (DT)	<i>Jedna krotka opisuje konkretną klasę w szkole</i>		
	ID_Klasy	Liczbowy	PK (surrogate key), unikalny identyfikator klasy
	Nazwa	Tekst - 2	Nazwa klasy np. 4A
	Liczba_uczniów	Liczbowy	Liczba uczniów w klasie
Przedmiot (DT)	<i>Jedna krotka opisuje konkretny przedmiot w szkole</i>		

Nazwa tabeli	Atrybut	Typ atrybutu	Opis
	ID_Przedmiotu	Liczbowy	PK (surrogate key), unikalny identyfikator przedmiotu
	Nazwa	Varchar(30)	Nazwa przedmiotu
Data (DT)	ID_Daty	Liczbowy	PK (surrogate key), ID daty
	Data	Data	Konkretna data
	Rok	4 cyfry	Rok
	Miesiąc	2 cyfry	Miesiąc
	Dzień	2 cyfry	Dzień
Junk (DT)	ID_Junk	Liczbowy	PK (surrogate key), identyfikator junk
	Ocena	1 cyfra	Możliwa wartość z przedziału 1-6 przedstawiająca ocenę
	Frekwencja	Tekst - 5	Przedstawia przedział wartości odpowiadający frekwencji ucznia. Dopuszczane wartości: 0-49, 50-69, 70-89, 90-100

4. Model wymiarów

Definicja faktu

Fakt – Koniec roku szkolnego: Uczeń kończy konkretny rok szkolny, otrzymuje z danego przedmiotu ocenę i wylicza się jego frekwencje. Jeżeli jest w 4 klasie, może pisać również maturę z danego przedmiotu. Zapisana jest wówczas data napisania egzaminu oraz wynik.

Fact table: Koniec_roku

Ziarnistość:

- Dany uczeń
- Dana klasa
- Dany przedmiot
- Dany rok szkolny
- Rok podejścia do matury (jeśli był w 4 klasie i napisał egzamin)

Miary:

- Wynik z matury: AVG, COUNT
- Liczba uczniów: COUNT(1)
- Frekwencja: AVG
- Ocena: AVG
- Czy zdana matura? COUNT

5. Definicje wymiarów

Wymiar/Atrybut wymiaru	Tabela/Kolumna	Typ
Uczeń	Uczeń	Wymiar
Pesel	Uczeń.Pesel	Atrybut wymiaru
Imię	Uczeń.Imię	Atrybut wymiaru
Nazwisko	Uczeń.Nazwisko	Atrybut wymiaru
Wiek	Uczeń.Wiek	Atrybut wymiaru
Klasa	Klasa	Wymiar
Nazwa	Klasa.Nazwa	Atrybut wymiaru
Liczba uczniów	Klasa.LiczbaUczniów	Atrybut wymiaru
Przedmiot	Przedmiot	Wymiar
Nazwa	Przedmiot.Nazwa	Atrybut wymiaru
Data	Data	Wymiar
Data	Data	Atrybut wymiaru
Rok	Data.Rok	Atrybut wymiaru
Miesiąc	Data.Miesiąc	Atrybut wymiaru
Dzień	Data.Dzień	Atrybut wymiaru
Junk	Junk	Wymiar
Ocena	Junk.Ocena	Atrybut wymiaru
Frekwencja	Junk.Frekwencja	Atrybut wymiaru

Wymiar/Atrybut wymiaru	Tabela/Kolumna	Typ
Data matury (Hierarchiczny)	• Data.Rok •• Data.Miesiąc ••• Data.Dzień	Wymiar hierarchiczny
Data roku szkolnego (Hierarchiczny)	• Data.Rok •• Data.Miesiąc ••• Data.Dzień	Wymiar hierarchiczny

6. Sprawdzenie wykonalności zapytań

1. Jak zmienia się średni wynik matur z poszczególnych przedmiotów w porównaniu do średnich wyników z ostatnich 3 lat?

Miara: Średni wynik z matury

Wymiar: Przedmiot (Atrybut: Nazwa)

Wymiar: Data – rok pisania matury (Atrybut: Rok)

2. Ile uczniów miało wynik z matury <10% w tym roku?

Miara: Wynik z matury (<10%)

Wymiar: Data – rok pisania matury (Atrybut: Rok)

3. Czy uczniowie, którzy wcześniej zaczęli chodzić do szkoły (są młodsi) lepiej / gorzej napisali maturę?

Miara: Średni wynik z matury

Wymiar: Uczeń (Atrybut: Wiek)

4. Jak zmieniła się ilość osób, które nie przystąpiły do matury w danym roku szkolnym w porównaniu do poprzedniego?

Miara: Liczba uczniów

Wymiar: Data – rok pisania matury (Atrybut: Rok)

5. Jak rozmiar klasy miał znaczenie na wyniki maturalne?

Miara: Średni wynik z matury

Wymiar: Klasa (Atrybut: Liczba_uczniów)

6. Jak indywidualne średnie wyniki maturalne uczniów, którzy mają wysoką frekwencję, różnią się od wyników uczniów z niską frekwencją?

Miara: Średni wynik z matury

Wymiar: Junk (Atrybut: Frekwencja)

7. Jak zmieniła się średnia ocen maturzystów w danym roku w porównaniu do poprzedniego?

Miara: Średnia ocen
Wymiar: Data – rok szkolny (Atrybut: Rok)

8. Czy istnieje zależność między średnią frekwencją wszystkich uczniów a średnią wyników maturalnych w szkole?

Miara: Średnia frekwencja
Miara: Średni wynik z matury

9. Czy uczniowie którzy dostali ocenę 5 lub 6 z danego przedmiotu mieli lepszy wynik na maturze?

Miara: Wynik z matury
Wymiar: Przedmiot (Atrybut: Nazwa)
Wymiar: Junk (Atrybut: Ocena)

10. Czy uczniowie którzy dostali ocenę 2 z danego przedmiotu nie zdali z niego matury?

Miara: Czy zdana matura?
Wymiar: Przedmiot (Atrybut: Nazwa)
Wymiar: Junk (Atrybut: Ocena)

7. Sprawdzenie czy istnieją dane do wypełnienia hurtowni danych

Nazwa tabeli	Kolumna	Źródło
Koniec_roku	Jedna krotka oznacza rezultat pracy ucznia z danego przedmiotu w danym roku szkolnym. Uwzględnia podejście do egzaminu maturalnego.	
	ID_Ucznia	ID Ucznia, unikalny identyfikator ucznia, FK wzięty z tabeli wymiaru. Na podstawie Peselu przechowywanego w tabeli Uczeń w bazie danych.
	ID_Klasy	ID Klasy, unikalny identyfikator klasy, FK wzięty z tabeli wymiaru. Na podstawie ID_Klasy w tabeli Klasa w bazie danych
	ID_Przedmiotu	ID Przedmiotu, unikalny identyfikator przedmiotu, FK wzięty z tabeli wymiaru. Na podstawie ID_Przedmiotu w tabeli Przedmiot w bazie danych.
	ID_Roku_szkolnego	ID Daty roku szkolnego, identyfikator do daty, roku, którego dotyczy zakończenie roku szkolnego, FK wzięty z tabeli wymiaru. Na podstawie

Nazwa tabeli	Kolumna	Źródło
		Rok_szkolny z tabeli Klasa w bazie danych.
	ID_Daty_matury	ID Daty matury, identyfikator do daty, w której napisany został egzamin maturalny. FK wzięty z tabeli wymiaru. Na podstawie pliku excel Wyniki.
	ID_Junk	ID Junk, unikalny identyfikator do tabeli Junk, FK wzięty z tabeli wymiaru,
	Ocena_z_przedmiotu	Ocena wystawiona z przedmiotu na koniec roku Szkolnego. Możliwe wartości z przedziału 1-6. Na podstawie kolumny Ocena z tabeli Koniec_roku.
	Frekwencja	Informacja o ilości obecności ucznia wyrażona procentowo. Na podstawie kolumny Frekwencja z tabeli Koniec_roku.
	Wynik_z_matury	Wynik z matury z konkretnego przedmiotu wyrażony procentowo. Na podstawie kolumny Wynik z pliku excel Wyniki (łączenie po Peselu)
	Czy_zdany_przedmiot?	Zmienna logiczna, czy przedmiot został zaliczony, oparta na ocenie z przedmiotu. (ocena ≥ 2)
	Czy_zdana_matura?	Zmienna logiczna, czy maturzysta zdał maturę oparta na wyniku matury. (wynik $\geq 30\%$)
Uczeń	Jedna krotka opisuje konkretnego ucznia w szkole (Implementacja SCD2)	
	ID_Ucznia	PK (surrogate key), unikalny identyfikator ucznia.
	Pesel	Pesel ucznia. Z kolumny Pesel z tabeli Uczeń.
	Imię	Imię ucznia. Z kolumny Imię z tabeli Uczeń.
	Nazwisko	Nazwisko ucznia. Z kolumny Nazwisko z tabeli Uczeń.
	Wiek	Wiek ucznia obliczany za pomocą pierwszych dwóch cyfr peselu (Implementacja SCD2).

Nazwa tabeli	Kolumna	Źródło
Klasa	Jedna krotka opisuje konkretną klasę w szkole (Implementacja SCD2)	
	ID_Klasy	PK (surrogate key), unikalny identyfikator klasy.
	Nazwa	Nazwa klasy np. 4A. Na podstawie ID_Klasy z tabeli Klasa.
	Liczba_uczniów	Liczba uczniów w klasie (Implementacja SCD2)
Przedmiot	Jedna krotka opisuje konkretny przedmiot w szkole	
	ID_Przedmiotu	PK (surrogate key), unikalny identyfikator przedmiotu. Z kolumny IDPrzedmiotu z tabeli PRzedmiot
	Nazwa	Nazwa przedmiotu. Z kolumny Nazwa z tabeli Przedmiot.
Data	Jedna krotka opisuje jeden dzień. Wszystkie dane w tej tabeli są generowane krotka po krotce na podstawie dowolnego kalendarza, przed procesem ETL	
Junk	Krotki odpowiadają “wszystkim” możliwym kombinacjom wartości dla Ocena i Frekwencja i są generowane przed procesem ETL	