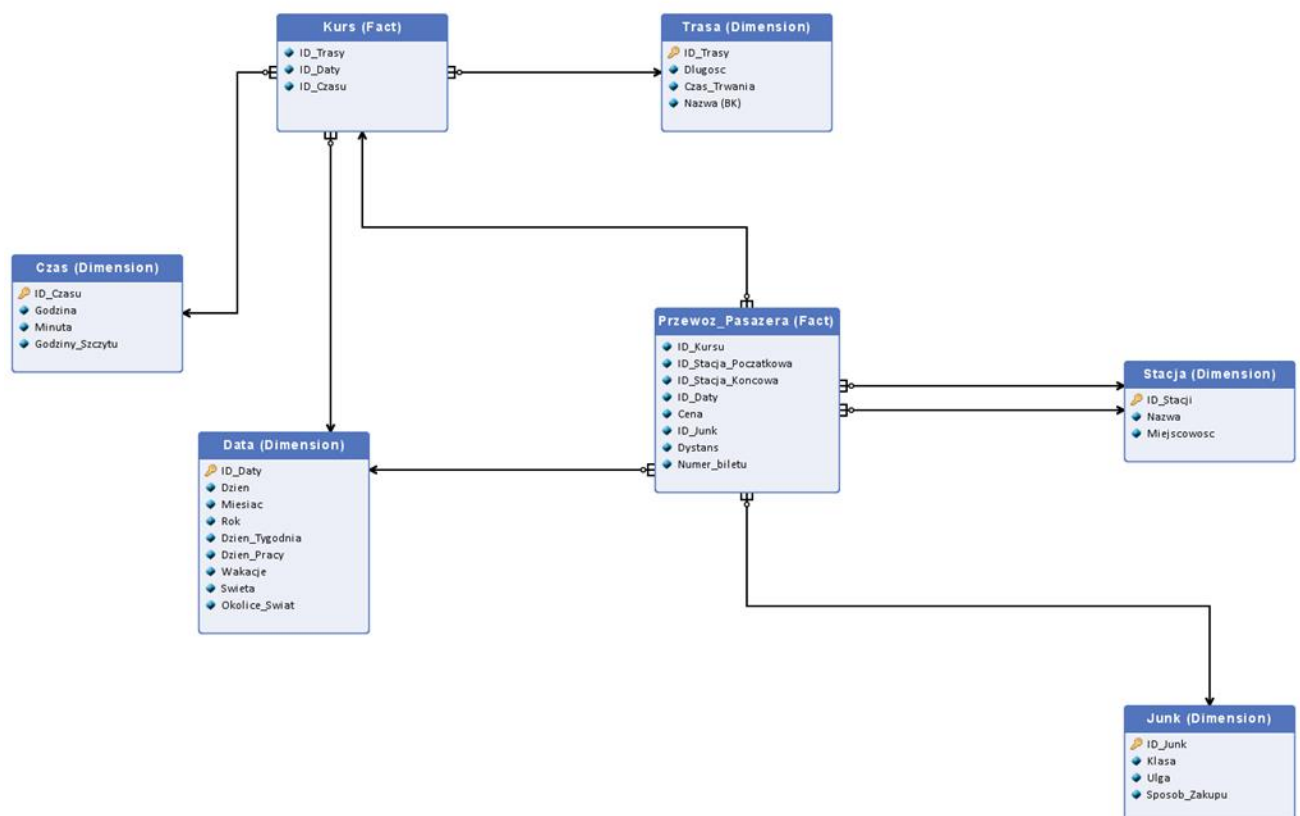


# Twoja Kolej - Projekt Hurtowni Danych

Hurtownia danych projektowana jest dla procesu biznesowego przewozu pasażerów. Proces ten jest opisany w dokumencie *requirementsProcessSpecification.pdf*.

## *Schemat relacyjnej bazy danych*



Przewoz_Pasazera (Fact)			
Jeden rekord opisuje jeden fakt przewiezienia pasazera			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
Numer_biletu	Nie	Numeric	Numer identyfikacyjny biletu
ID_Kursu	Nie	Numeric	Numer identyfikacyjny kursu
ID_Stacja_Pocatkowa	Nie	Numeric	Numer identyfikacyjny stacji na ktorej wsiada pasazer
ID_Stacja_Koncowa	Nie	Numeric	Numer identyfikacyjny stacji na ktorej wysiada pasazer
ID_Daty	Nie	Numeric	Numer identyfikacyjny daty zakupu
Cena	Nie	Money	Cena biletu
Dystans	Nie	Decimal	Dystans pomiędzy stacją początkową i końcową
ID_Junk	Nie	Numeric	Numer identyfikacyjny Junk

Kurs (Fact)			
Jeden rekord opisuje fakt wystapienia kursu			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID_Trasy	Nie	Numeric	Numer identyfikacyjny trasy
ID_Daty	Nie	Numeric	Numer identyfikacyjny daty rozpoczecia kursu
ID_Czasu	Nie	Numeric	Numer identyfikacyjny czasu rozpoczecia kursu

Trasa (Dimension)			
Jeden rekord opisuje jedna trase			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID_Trasy	Tak	Numeric	Numer identyfikacyjny trasy
Nazwa (BK)	Nie	Varchar(30)	Nazwa trasy składająca się z nazwy stacji początkowej i końcowej oddzielonych myślnikiem.  <b>Klucz biznesowy</b>
Dlugosc	Nie	Varchar(11)	Długość trasy (0-50, 50-100, 100-200, 200-500, powyżej 500)  Atrybut zmieniający się w czasie ( <b>SCD 1</b> )
Czas_Trwania	Nie	Varchar(11)	Czas trwania trasy (0-1h, 1-3h, 3-5h, 5-10h, powyżej 10h)  Atrybut zmieniający się w czasie ( <b>SCD 1</b> )

Stacja (Dimension)			
Jeden rekord opisuje jedna stacje			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID_Stacji	Tak	Numeric	Numer identyfikacyjny stacji
Nazwa	Nie	Varchar(30)	Pełna nazwa stacji
Miejscowosc	Nie	Varchar(30)	Miejscowosc w ktorej znajduje sie stacja

Data (Dimension)			
Jeden rekord opisuje jeden dzien			

Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID_Daty	Tak	Numeric	Numer identyfikacyjny daty
Dzien	Nie	2 Digits	Numer dnia (01 - 31)
Miesiac	Nie	Varchar(11)	Nazwa miesaca (Styczen, Luty, Marzec, Kwiecien, Maj, Czerwiec, Lipiec, Sierpien, Wrzesien, Pazdziernik, Listopad, Grudzien)
Rok	Nie	4 Digits	Numer roku
Dzien_Tygodnia	Nie	Varchar(12)	Nazwa dnia tygodnia (Poniedzialek, Wtorek, Sroda, Czwartek, Piątek, Sobota, Niedziela)
Dzien_Pracy	Nie	Varchar(11)	Czy jest to dzien pracy (Dzień pracy/Dzień wolny)
Wakacje	Nie	Varchar(11)	Czy jest to dzien w trakcie wakacji (Wakacje/Ferie/Rok szkolny)
Swieta	Nie	Varchar(50)	Czy jest to dzien swiateczny (Boze Narodzenie/ Wielkanoc/Wszystkich Swietych/ Nowy Rok/ Brak swiat)
Okolice_Swiat	Nie	Varchar(60)	Czy jest to dzien w okolicach swiat (Okolice Bozego Narodzenia/ Okolice Wielkanocy/Okolice Wszystkich Swietych/ Okolice Nowego Roku/ Brak swiat)

Czas (Dimension)
Jeden rekord opisuje jedna minute

Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID_Czasu	Tak	Numeric	Numer identyfikacyjny czasu
Godzina	Nie	2 Digits	Godzina (00 - 23)
Minuta	Nie	2 Digits	Minuta (00 - 59)
Godziny_Szczytu	Nie	Varchar(21)	Czy czas jest w godzinach 7 - 9 lub 15 – 17, czy w pozostałych (Godziny szczytu/Godziny poza szczytem)

Junk (Dimension)			
Rekordy opisują wszystkie możliwe kombinacje wartości dotyczących zakupionego biletu na przewóz			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID_Junk	Tak	Numeric	Numer identyfikacyjny Junk
Klasa	Nie	Varchar(2)	Klasa podrózna wybrana przez pasażera (I/II)
Ulga	Nie	Varchar(4)	Procent zniżki na cenie biletu (0%/25%/33%/37%/49%/50%/51%/78%/80%/93%/95%/100%)
Sposob_Zakupu	Nie	Varchar(9)	Sposób zakupu biletu (Kasa/Konduktor/Internet)

## ***Model wielowymiarowy***

Definicje faktów:

**Fakt 1 - Przewóz:** Przewóz osoby na określonym kursie, pomiędzy określonymi stacjami, w określonej klasie, uwzględniając określoną ulgę i określony sposób zakupu.

Tabela faktu: Przewoz\_Pasazera

Ziarnistość:

- Konkretny kurs, na danej trasie, wyruszający określonego dnia o określonej godzinie
- Konkretna stacja początkowa i końcowa znajdujące się w określonych miejscowościach
- Konkretna klasa wagonu na zakupione miejsce
- Określona ulga na cenę przejazdu
- Określony sposób zakupu biletu
- Określona data zakupu biletu

Miary i funkcje agregujące:

- Liczba przewiezionych osób - COUNT(1)
- Zysk ze sprzedanych biletów - SUM(Cena)
- Łączny dystans – SUM(Dystans)
- Średni dystans – SUM(Dystans)/COUNT(1) (Miara wyliczalna)

**Fakt 2 – Kurs:** Pojedynczy przejazd pociągu na określonej trasie, wyruszający określonego dnia o określonej godzinie.

Tabela faktu: Kurs

Ziarnistość:

- Konkretna trasa przejazdu pociągu
- Konkretna data wyruszenia
- Konkretna godzina wyruszenia

Miary i funkcje agregujące:

- Liczba kursów - COUNT(1)
- Liczba obsługiwanych tras – DISTINCT COUNT(ID\_Trasy)

**Definicje wymiarów:**

Wymiary dla faktu 1:

Wymiar/Atrybut wymiaru	Tabela/Pole w tabeli	Typ
Numer biletu	Przewoz_Pasazera.Numer_biletu	Wymiar zdegenerowany
Trasa	Trasa	Wymiar (SCD 1)
Nazwa	Trasa.Nazwa	Atrybut wymiaru
Długość	Trasa.Dlugosc	Atrybut wymiaru (SCD 1)
Czas trwania	Trasa.Czas_Trwania	Atrybut wymiaru (SCD 1)
Stacja Początkowa	Stacja	Wymiar
Nazwa	Stacja.Nazwa	Atrybut wymiaru
Miejscowość	Stacja.Miejscowosc	Atrybut wymiaru
Stacja Końcowa	Stacja	Wymiar
Nazwa	Stacja.Nazwa	Atrybut wymiaru
Miejscowość	Stacja.Miejscowosc	Atrybut wymiaru

Data zakupu biletu	Data	Wymiar
Dzień	Data.Dzien	Atrybut wymiaru
Miesiąc	Data.Miesiac	Atrybut wymiaru
Rok	Data.Rok	Atrybut wymiaru
Dzień Tygodnia	Data.Dzien_Tygodnia	Atrybut wymiaru
Dzień Pracy	Data.Dzien_Pracy	Atrybut wymiaru
Wakacje	Data.Wakacje	Atrybut wymiaru
Święta	Data.Swieta	Atrybut wymiaru
Okolice Świąt	Data.Okolice_Swiat	Atrybut wymiaru
Inne	Junk	Wymiar
Klasa	Junk.Klasa	Atrybut wymiaru
Ulga	Junk.Ulga	Atrybut wymiaru
Sposób zakupu biletu	Junk.Sposob_Zakupu	Atrybut wymiaru
Bilet	<ul style="list-style-type: none"> <li>• Junk.Ulga</li> <li>•• Junk.Klasa</li> <li>••• Junk.Sposob_Zakupu</li> </ul>	Wymiar Hierarchiczny
Data przejazdu	Data	Wymiar
Dzień	Data.Dzien	Atrybut wymiaru
Miesiąc	Data.Miesiac	Atrybut wymiaru
Rok	Data.Rok	Atrybut wymiaru
Dzień Tygodnia	Data.Dzien_Tygodnia	Atrybut wymiaru
Dzień Pracy	Data.Dzien_Pracy	Atrybut wymiaru
Wakacje	Data.Wakacje	Atrybut wymiaru
Święta	Data.Swieta	Atrybut wymiaru
Okolice Świąt	Data.Okolice_Swiat	Atrybut wymiaru
Czas przejazdu	Czas	Wymiar
Godzina	Czas.Godzina	Atrybut wymiaru
Minuta	Czas.Minuta	Atrybut wymiaru
Godziny Szczytu	Czas.Godziny_Szczytu	Atrybut wymiaru

Wymiary dla faktu 2:

Wymiar/Atrybut wymiaru	Tabela/Pole w tabeli	Typ
Trasa	Trasa	Wymiar (SCD 1)
Nazwa	Trasa.Nazwa	Atrybut wymiaru
Długość	Trasa.Dlugosc	Atrybut wymiaru
Czas trwania	Trasa.Czas_Trwania	Atrybut wymiaru
Czas	Czas	Wymiar
Godzina	Czas.Godzina	Atrybut wymiaru
Minuta	Czas.Minuta	Atrybut wymiaru
Godziny Szczytu	Czas.Godziny_Szczytu	Atrybut wymiaru
Czas wyruszenia	<ul style="list-style-type: none"> <li>• Czas.Godzina</li> <li>•• Czas.Minuta</li> </ul>	Wymiar Hierarchiczny
Data	Data	Wymiar

Dzień	Data.Dzien	Atrybut wymiaru
Miesiąc	Data.Miesiac	Atrybut wymiaru
Rok	Data.Rok	Atrybut wymiaru
Dzień Tygodnia	Data.Dzien_Tygodnia	Atrybut wymiaru
Dzień Pracy	Data.Dzien_Pracy	Atrybut wymiaru
Wakacje	Data.Wakacje	Atrybut wymiaru
Święta	Data.Swieta	Atrybut wymiaru
Okolice Świat	Data.Okolice_Swiat	Atrybut wymiaru
Kursy w święta	<ul style="list-style-type: none"> <li>• Data.Rok</li> <li>•• Data.Miesiac</li> <li>••• Data.Swieta</li> </ul>	Wymiar Hierarchiczny

## ***Sprawdzenie wykonalności zapytań w oparciu o model wielowymiarowy***

**Dlaczego w zeszłym miesiącu zysk ze sprzedaży biletów wzrósł/zmalał:**

1. Porównaj łączny dystans przejechany przez osoby w I klasie w tym miesiącu w porównaniu z poprzednim.

Miara: łączny dystans

Wymiar: Junk(atrybuty wymiaru: Klasa)

Wymiar: Data przejazdu (atrybuty wymiaru: Miesiąc)

2. Porównaj średnią długość trasy na bilecie w tym miesiącu w porównaniu z poprzednim.

Miara: Średni dystans

Wymiar: Data przejazdu(atrybuty wymiaru: Miesiąc)

3. Porównaj przychód ze sprzedaży biletów na poszczególnych trasach w tym miesiącu w porównaniu z poprzednim.

Miara: Zysk ze sprzedanych biletów

Wymiar: Trasa(atrybuty wymiaru: Nazwa)

Wymiar: Data zakupu biletu(atrybuty wymiaru: Miesiąc)

4. Porównaj liczbę sprzedanych biletów w dniach pracy w tym miesiącu w porównaniu z poprzednim.

Miara: Liczba przewiezionych osób



Wymiar: Data zakupu biletu (atrybuty wymiaru: Miesiąc, Dzień Pracy)

5. Jaka była średnia długość trasy na bilecie, którego stacja początkowa to XXX w tym miesiącu w porównaniu z poprzednim?

Miara: Średni dystans

Wymiar: Stacja Początkowa(atrybuty wymiaru: Nazwa)

Wymiar: Data przejazdu(atrybuty wymiaru: Miesiąc)

### **Dlaczego liczba pasażerów w zeszłym miesiącu wzrosła/zmaląa:**

1. Porównaj liczbę sprzedanych biletów ulgowych w tym miesiącu w porównaniu z poprzednim.

Miara: Liczba przewiezionych osób

Wymiar: Junk(atrybuty: Ulga)

Wymiar: Data zakupu biletu(atrybuty wymiaru: Miesiąc)

2. Jaka była najczęściej wybierana stacja końcowa/początkowa w tym i w zeszłym miesiącu?

Miara: Liczba sprzedanych biletów

Wymiar: Stacja Początkowa(atrybuty wymiaru: Nazwa) / Stacja Końcowa (atrybuty wymiaru: Nazwa)

Wymiar: Data zakupu biletu(atrybuty wymiaru: Miesiąc)

3. Które trasy były najpopularniejsze w godzinach szczytu w tym i w zeszłym miesiącu.

Miara: Liczba sprzedanych biletów

Wymiar: Trasa(atrybuty wymiaru: Nazwa)

Wymiar: Czas przejazdu(atrybuty wymiaru: Godziny Szczytu)

Wymiar: Data przejazdu(atrybuty wymiaru: Miesiąc)

4. Które trasy były najpopularniejsze w okolicy świąt w tym i w zeszłym miesiącu.

Miara: Liczba sprzedanych biletów

Wymiar: Trasa(atrybuty wymiaru: Nazwa)

Wymiar: Data przejazdu(atrybuty wymiaru: Miesiąc, Okolice Świąt)

5. Porównaj liczbę sprzedanych biletów I klasy w tym miesiącu w porównaniu z poprzednim.

Miara: Liczba sprzedanych biletów

Wymiar: Junk(atrybuty wymiaru: Klasa)

Wymiar: Data zakupu biletu(atrybuty wymiaru: Miesiąc)

## ***Sprawdzenie czy w źródłach danych są dane, którymi wypełnimy hurtownię danych***

Nazwa tabeli	Atrybut	Skąd pobierane są dane
Przewoz_Pasazera	Krotki odpowiadają faktom przewiezienia pasażera	
	ID_Kursu	Klucz obcy pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu Kurs zapisanego w tabeli Bilet w źródle Twój Bilet.
	Stacja_Poczatkowa	Klucz obcy pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu Od zapisanego w tabeli Bilet w źródle Twój Bilet.
	Stacja_Koncowa	Klucz obcy pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu Do zapisanego w tabeli Bilet w źródle Twój Bilet.
	ID_Daty	Klucz obcy pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu Data zakupu zapisanej w tabeli Bilet w źródle Twój Bilet.
	Cena	Cena biletu. Wartość jest pobierana z atrybutu Cena zapisanego w tabeli Bilet w źródle Twój Bilet.
	Dystans	Dystans pomiędzy stacją początkową i końcową mierzony po torach. Wartość jest pobierana z kolumny Dystans z pliku <i>distances.csv</i>

	ID_Junk	Klucz obcy jest pobierany z tabeli wymiarów. Jego wartość wynika ze sposobu zakupu, klasy i ulgi pobieranych z tabeli Bilet w źródle Twój Bilet.
Kurs	Krotki odpowiadają faktom wystąpienia kursu na trasie	
	ID_Trasy	Klucz obcy jest pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu ID pobieranego z tabeli Trasa w źródle Twój Bilet.
	ID_Daty	Klucz obcy pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu Data wyruszenia zapisanej w tabeli Kurs w źródle Twój Bilet.
	ID_Czasu	Klucz obcy pobierany z tabeli wymiarów. Jego wartość wynika z atrybutu Godzina wyruszenia zapisanej w tabeli Kurs w źródle Twój Bilet.
Trasa	Krotki odpowiadają poszczególnym trasom	
	ID_Trasy	Klucz surogatowy – generowany przez bazę danych.
	Nazwa	Nazwa trasy składająca się z nazwy stacji początkowej i końcowej oddzielonych myślnikiem. Wartości są pobierane z atrybutu Nazwa_Stacji zapisanego w tabeli Stacja_Trasa w źródle Twój Bilet, gdzie dla danego atrybutu ID_Trasy atrybut Numer_Stacji_Na_Trasie jest minimalny i maksymalny.
	Dlugosc	Wartość jest wyliczana z atrybutu Długość zapisanego w tabeli Trasa w źródle Twój Bilet.

		Przyjmuje wartości: 0-50, 50-100, 100-200, 200-500, powyżej 500.
	Czas_Trwania	Czas potrzebny na zrealizowanie trasy. Wartość jest wyliczana z atrybutu Czas Trwania zapisanego w tabeli Trasa w źródle Twój Bilet. Przyjmuje wartości: 0-1h, 1-3h, 3-5h, 5-10h, powyżej 10h
Stacja	Krotki odpowiadają poszczególnym stacjom	
	ID_Stacji	Klucz surogatowy – generowany przez bazę danych.
	Nazwa	Wartość pobierana z atrybutu Nazwa z tabeli Stacja w źródle Twój Bilet.
	Miejscowosc	Nazwa miejscowości, w której znajduje się stacja. Wartość jest pobierana z atrybutu Miejscowość zapisanego w tabeli Stacja w źródle Twój Bilet.
Data	Krotki odpowiadają konkretnej dacie. Wszystkie dane w tej tabeli są generowane krotka po krotce na podstawie dowolnego kalendarza przed procesem ETL	
Czas	Krotki odpowiadają konkretnym godzinom dnia (niezależnie od daty). Wszystkie dane w tej tabeli są generowane krotka po krotce wg zasady działania zegara przed procesem ETL.	
Junk	Krotki odpowiadają wszystkim możliwym kombinacjom wartości dla kolumn Sposób_Zakupu, Klasa i Ulga i są generowane przed procesem ETL.	
	ID_Junk	Klucz surogatowy – generowany przez bazę danych.