

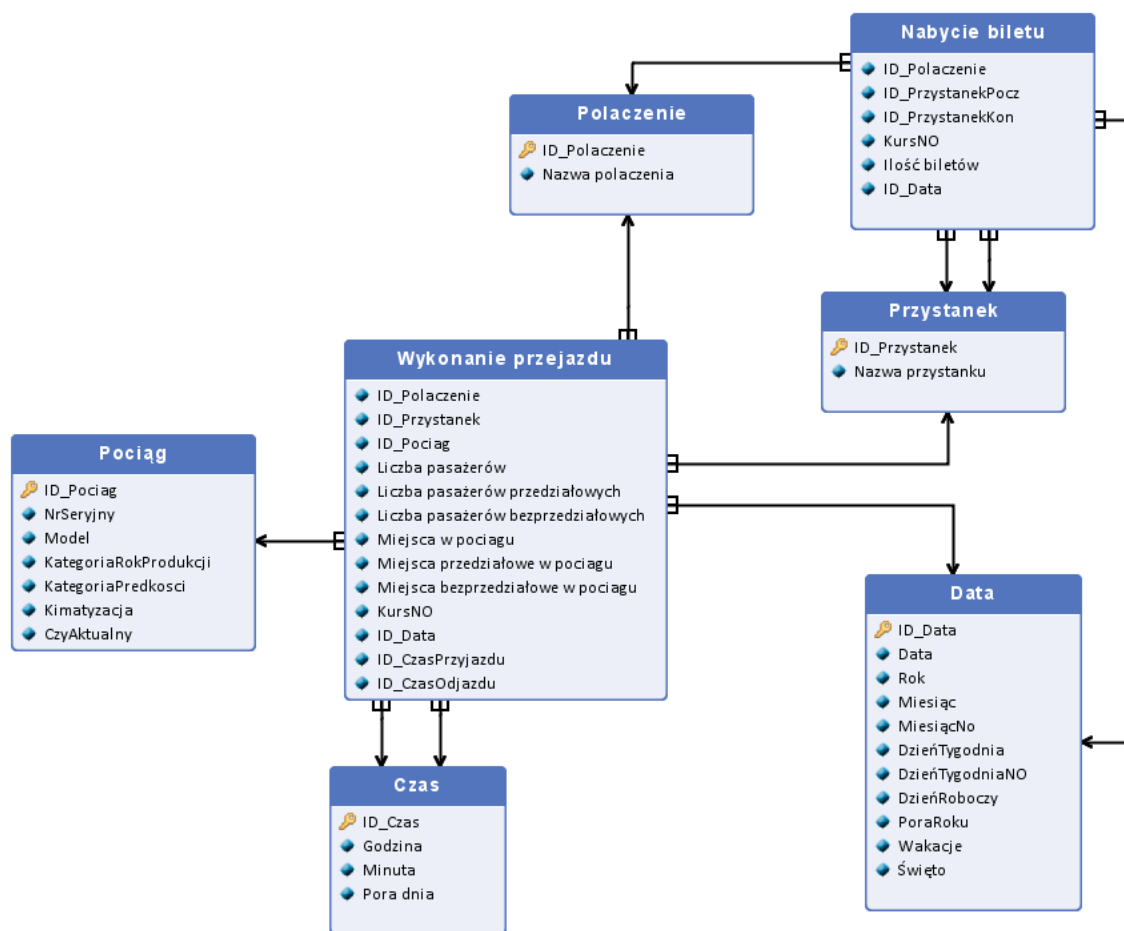
## Polskie Koleje – Projekt hurtowni danych

Paweł Kuszczak, Jan Kalwasiński

### Proces biznesowy

Hurtownia danych jest zaprojektowana pod proces przewozu pasażerskiego. Proces ten jest opisany w dokumencie *Specyfikacja wymagań dla procesu biznesowego "Przewóz pasażerski"*.

### Schemat relacyjnej bazy danych



Nazwa Tabeli	Atrybut	Typ atrybutu	Opis
Wykonanie przejazdu (Tabela faktów)	Jeden wiersz opisuje fakt przejazdu pociągu na dany przystanek w ramach danego kursu.		

	ID_Polaczenie	Numeryczny	FK Polaczenie, ID wykonywanego połączenia
	ID_Przystanek	Numeryczny	FK Przystanek, ID analizowanego przystanku
	ID_Pociag	Numeryczny	FK Pociąg, ID pociągu realizującego kurs
	Liczba pasażerów	Numeryczny	Suma z liczby pasażerów przedziałowych i bezprzedziałowych
	Liczba pasażerów przedziałowych	Numeryczny	Liczba pasażerów na miejscach przedziałowych w pociągu w momencie odjeżdżania z przystanku
	Liczba pasażerów bezprzedziałowych	Numeryczny	Liczba pasażerów na miejscach bezprzedziałowych w pociągu w momencie odjeżdżania z przystanku
	Miejsca w pociągu	Numeryczny	Suma z liczby dostępnych miejsc przedziałowych i bezprzedziałowych w pociągu
	Miejsca przedziałowe w pociągu	Numeryczny	Liczba dostępnych miejsc przedziałowych w pociągu
	Miejsca bezprzedziałowe w pociągu	Numeryczny	Liczba dostępnych miejsc bezprzedziałowych w pociągu
	KursNO	Numeryczny	Numer kursu
	ID_Data	Numeryczny	FK Data, Data przyjazdu na przystanek
	ID_CzasPrzyjazdu	Numeryczny	FK Czas, Czas przyjazdu na przystanek
	ID_CzasOdjazdu	Numeryczny	FK Czas, Czas przyjazdu na przystanek
<b>Nabycie biletu (Tabela faktów)</b>	<b>Jeden wiersz opisuje ilość nabytych biletów na dany kurs z danym przystankiem początkowym i końcowym.</b>		
	ID_Polaczenie	Numeryczny	FK Polaczenie, ID wykonywanego połączenia
	ID_PrzystanekPocz	Numeryczny	FK Przystanek, ID Przystanku początkowego

	ID_PrzystanekKon	Numeryczny	FK Przystanek, ID Przystanku końcowego
	KursNO	Numeryczny	Numer kursu
	Ilość biletów	Numeryczny	ilość biletów na dany kurs z danym przystankiem początkowym i końcowym
	ID_Data	Numeryczny	FK Data, Data kursu
<b>Pociąg (Tabela wymiarów)</b>	<b>Jeden wiersz opisuje jeden pociąg. Do pociągu może zostać dodana klimatyzacja (SCD 2)</b>		
	ID_Pociąg NrSeryjny	Numeryczny Varchar(20)	PK BK Unikalny numer przypisany konkretnemu pociągowi
	Model	Varchar(10)	Model pociągu
	KategoriaRokProdukcji	Varchar(10)	Kategorie roku produkcji pociągu Dozwolone wartości: Stare (lata 1990-1999), Nowe (lata 2000-2009), Nowsze (lata 2010-2019), Najnowocześniejsze (lata 2020+)
	KategoriaPrędkości	Varchar(10)	Kategorie maksymalnej osiąganej prędkości Dozwolone wartości: Wolne (do 149km/h) Szybkie (od 150km/h do 179km/h) Najszybsze (od 180km/h)
	Klimatyzacja	Varchar(20)	Czy pociąg jest wyposażony w klimatyzację Dozwolone wartości: Klimatyzowany, Nieklimatyzowany
	CzyAktualny	Boolean	1 jeśli informacje aktualne, 0 jeśli nie (implementacja SCD 2)
<b>Połączenie (Tabela wymiarów)</b>	<b>Jeden wiersz opisuje jedno połączenie.</b>		
	ID_Połączenia	Numeryczny	PK
	Nazwa połączenia	Varchar(50)	Nazwa połączenia

<b>Przystanek (Tabela wymiarów)</b>	<b>Jeden wiersz opisuje jeden przystanek należący do połączenia.</b>		
	ID_Przystanek	Numeryczny	PK
	Nazwa przystanku	Varchar(30)	Nazwa przystanku
<b>Data (Tabela wymiarów)</b>	<b>Jeden wiersz opisuje jeden dzień.</b>		
	ID_Data	Numeryczny	PK
	Data	Data	Data
	Rok	4 cyfry	Rok
	Miesiąc	Varchar(10)	Miesiąc. Dozwolone wartości: Styczeń, Luty, Marzec, Kwiecień, Maj, Czerwiec, Lipiec, Sierpień, Wrzesień, Październik, Listopad, Grudzień
	MiesiącNo	Numeryczny	Numer miesiąca
	DzieńTygodnia	Varchar(10)	Dzień tygodnia. Dozwolone wartości: Poniedziałek, Wtorek, Środa, Czwartek, Piątek, Sobota, Niedziela
	DzieńTygodniaNO	Numeryczny	Numer dnia tygodnia
	DzieńRoboczy	Varchar(15)	Dzień roboczy. Dozwolone wartości: dzień wolny, dzień roboczy
	PoraRoku	Varchar(7)	Pora roku. Dozwolone wartości: Wiosna, Lato, Jesień, Zima.
	Wakacje	Varchar(20)	Charakterystyka czasu wolnego. Dozwolone wartości: dzień powszedni, ferie zimowe, ferie letnie.
	Święto	Varchar(50)	Rodzaj święta. Dozwolone wartości: Nowy Rok, Trzech Króli, Wielki Piątek, Wielka Sobota, Wielkanoc, Poniedziałek Wielkanocny, Święto Pracy, Święto Konstytucji 3 Maja, Boże Ciało, Wniebowzięcie Najświętszej Maryi Panny, Dzień Wszystkich Świętych, Dzień Zaduszny, Dzień Niepodległości, Boże Narodzenie, Drugi dzień Bożego Narodzenia, Dzień Babci, Dzień Dziadka,

	Walentynki, Dzień Kobiet, Dzień Matki, Dzień Dziecka, Dzień Ojca, Mikołajki, Sylwester		
<b>Czas (Tabela wymiarów)</b>	<b>Jeden wiersz opisuje jedną godzinę z minutami.</b>		
	ID_Czas	Numeryczny	PK
	Godzina	Numeryczny	Godzina. Zakres wartości: 0 - 23
	Minuta	Numeryczny	Minuta. Zakres wartości: 0 - 59
	Pora dnia	Varchar(10)	Pora dnia Dozwolone wartości: rano, południe, popołudnie, wieczór, noc

## Model wymiarowy

### Definicje faktów

**Fakt 1 Wykonanie przejazdu:** Wykonanie przejazdu na dany przystanek konkretnego dnia (Przyjazd i odjazd z przystanku o konkretnej godzinie). Przejazd jest realizowany w ramach danego połączenia przez dany pociąg. W pociągu znajduje się określona liczba pasażerów. Pociąg ma ograniczoną liczbę miejsc.

Tabela faktów: Wykonanie przejazdu

Ziarnistość:

- Określone połączenie
- Określony przystanek
- Określony pociąg
- Określony kurs
- Określona data
- Określone godziny przyjazdu/odjazdu

Miary i funkcje agregujące

- Liczba pasażerów przedziałowych
- Liczba pasażerów bezprzedziałowych
- Miejsca przedziałowe w pociągu
- Miejsca bezprzedziałowe w pociągu
- Liczba pasażerów
- Miejsca w pociągu

**Średnie obłożenie** –  $\text{SUM}(\text{Liczba pasażerów}) / \text{SUM}(\text{Miejsca w pociągu}) * 100\%$

**Średnie obłożenie wagonów przedziałowych**– **SUM**(Liczba pasażerów przedziałowych) / **SUM**(Miejsca przedziałowe w pociągu) \* 100%

**Średnie obłożenie wagonów bezprzedziałowych**– **SUM**(Liczba pasażerów bezprzedziałowych) / **SUM**(Miejsca bezprzedziałowe w pociągu) \* 100%

**Liczba kursów** – **COUNT**(DISTINCT KursNO)

**Liczba wszystkich faktów przejazdu** – **COUNT**(1)

**Fakt 2 Nabycie biletu:** Nabycie biletu na dany kurs z określonym przystankiem początkowym i końcowym. W każdym wierszu opisana jest liczba sprzedanych biletów dla kursu na konkretnym połączeniu, uwzględniając przystanki, od których zaczyna i na których kończy podróż pasażer.

Ziarnistość:

- Określone połączenie
- Określony przystanek początkowy
- Określony przystanek końcowy
- Określony kurs

Miary i funkcje agregujące

- Liczba biletów

**Łączna liczba sprzedanych biletów** – **SUM**(Ilość biletów)

## Definicje wymiarów

**Wymiary dla faktu 1 Wykonanie przejazdu**

WYMIAR/ WYMIAR ATRYBUT	Tabela/Kolumna	Typ
Numer Kursu	WykonaniePrzejazdu.KursNO	Wymiar zdegenerowany
Pociąg	Pociąg	Wymiar
Numer Seryjny	Pociąg.NrSeryjny	Atrybut wymiaru
Model	Pociąg.Model	Atrybut wymiaru
KategoriaRokProdukcji	Pociąg.KategoriaRokProdukcji	Atrybut wymiaru
Model Hierarchia	<ul style="list-style-type: none"><li>• Pociąg.Model</li><li>••Pociąg.KategoriaRokProdukcji</li></ul>	Wymiar hierarchiczny
KategoriaPredkości	Pociąg.KategoriaPrędkości	Atrybut wymiaru
Klimatyzacja	Pociąg.Klimatyzacja	Atrybut wymiaru
Przystanek	Przystanek	Wymiar
Nazwa przystanku	Przystanek.NazwaPrzystanku	Atrybut wymiaru
Połączenie	Połączenie	Wymiar
Nazwa połączenia	Połączenie.NazwaPołączenia	Atrybut wymiaru

Data	Data	Wymiar
Data Rok	Data.Rok	Atrybut wymiaru
Data Miesiąc	Data.Miesiąc	Atrybut wymiaru
Data Dzień	Data.Data	Atrybut wymiaru
Data Hierarchia	•Data.Rok •• Data.Miesiąc •••Data.Data	Wymiar hierarchiczny
Dzień tygodnia	Data.DzieńTygodnia	Atrybut wymiaru
Dzień roboczy	Data.DzieńRoboczy	Atrybut wymiaru
Pora roku	Data.PoraRoku	Atrybut wymiaru
Wakacje	Data.Wakacje	Atrybut wymiaru
Święto	Data.Święto	Atrybut wymiaru
Data dzień tygodnia hierarchia	•Data.Rok •• Data.Miesiąc •••Data.DzieńTygodnia	Wymiar hierarchiczny
Data wakacje	•Data.Rok •• Data.Wakacje	Wymiar hierarchiczny
Data dzień roboczy hierarchia	•Data.Rok •• Data.Miesiąc •••Data.DzieńRoboczy	Wymiar hierarchiczny
Data pora roku hierarchia	•Data.Rok ••Data.PoraRoku	Wymiar hierarchiczny
Czas przyjazdu	Czas	Wymiar
Czas przyjazdu godzina	Czas.Godzina	Atrybut wymiaru
Czas przyjazdu minuta	Czas.Minuta	Atrybut wymiaru
Czas przyjazdu pora dnia	Czas.PoraDnia	Atrybut wymiaru
Czas przyjazdu hierarchia	•Czas.Godzina ••Czas.Minuta	Wymiar hierarchiczny
Czas przyjazdu pora dnia hierarchia	• Czas.PoraDnia ••Czas.Godzina •••Czas.Minuta	Wymiar hierarchiczny
Czas odjazdu	Czas	Wymiar
Czas odjazdu godzina	Czas.Godzina	Atrybut wymiaru
Czas odjazdu minuta	Czas.Minuta	Atrybut wymiaru
Czas odjazdu pora dnia	Czas.PoraDnia	Atrybut wymiaru
Czas odjazdu hierarchia	•Czas.Godzina ••Czas.Minuta	Wymiar hierarchiczny
Czas odjazdu pora dnia hierarchia	• Czas.PoraDnia ••Czas.Godzina •••Czas.Minuta	Wymiar hierarchiczny

### Wymiary dla faktu 2 *Nabycie biletu*

WYMIAR/ WYMIAR ATRYBUT	Tabela/Kolumna	Typ
---------------------------	----------------	-----

Numer Kursu	NabycieBiletu.KursNO	Wymiar zdegenerowany
Przystanek Początkowy	Przystanek	Wymiar
Nazwa przystanku początkowego	Przystanek.NazwaPrzystanku	Atrybut wymiaru
Przystanek Końcowy	Przystanek	Wymiar
Nazwa przystanku końcowego	Przystanek.NazwaPrzystanku	Atrybut wymiaru
Połączenie	Połączenie	Wymiar
Nazwa połączenia	Połączenie.NazwaPołączenia	Atrybut wymiaru

## Sprawdzanie wykonalności zapytań w oparciu o model wielowymiarowy

- Porównaj różnice w obłożeniu pociągów w różnych porach dnia.  
*Miara: Średnie obłożenie*  
*Wymiar: Czas (atrybuty wymiaru: Pora Dnia)*
- Porównaj obłożenie pociągów w różnych dniach tygodnia.  
*Miara: Średnie obłożenie*  
*Wymiar: Data (atrybuty wymiaru: Dzień tygodnia)*
- Podaj połączenia, które mają najczęściej obłożenie poniżej 60%.  
*Miara: Średnie obłożenie*  
*Wymiar: Połączenie (atrybuty wymiaru: Nazwa połączenia)*
- Wypisz, ile osób kończy trasę na jakim przystanku na danym kursie.  
*Miara: Łączna liczba sprzedanych biletów*  
*Wymiar: KursNO*  
*Wymiar: Przystanek Końcowy (atrybuty wymiaru: Nazwa przystanku)*
- Podaj różnice w obłożeniu pociągów w różnych miesiącach roku na danych połączeniach.  
*Miara: Średnie obłożenie*  
*Wymiar Połączenie: (atrybuty wymiaru: Nazwa połączenia)*  
*Wymiar: Data (atrybuty wymiaru: Miesiąc)*
- Podaj średnie obłożenie wagonów przedziałowych/bezprzedziałowych danego pociągu na trasach jakie wykonał w ciągu miesiąca.  
*Miara: Średnie obłożenie wagonów przedziałowych*  
*Miara: Średnie obłożenie wagonów bezprzedziałowych*



*Wymiar: Pociąg (atrybuty wymiaru: Model)*

*Wymiar: Data (atrybuty wymiaru: Miesiąc)*

7. Porównaj obłożenie wszystkich modeli pociągów wykonujących dane połączenie o podobnej godzinie.

*Miara: Średnie obłożenie*

*Wymiar: Połączenie (atrybuty wymiaru: Nazwa połączenia)*

*Wymiar: Pociąg (atrybuty wymiaru: Model)*

*Wymiar: Czas (atrybuty wymiaru: Godzina)*

8. Porównaj obłożenie pociągów jeżdżących na tej samej trasie, ale z inną maksymalną prędkością.

*Miara: Średnie obłożenie*

*Wymiar: Połączenie (atrybuty wymiaru: Nazwa połączenia)*

*Wymiar: Pociąg (atrybuty wymiaru: KategoriaPrędkości)*

9. Porównaj obłożenie pociągów jeżdżących na tej samej trasie, wyprodukowanych w innych latach.

*Miara: Średnie obłożenie*

*Wymiar: Połączenie (atrybuty wymiaru: Nazwa połączenia)*

*Wymiar: Pociąg (atrybuty wymiaru: KategoriaRokProdukcji)*

10. Porównaj średnie obłożenie pociągów kursujących w okresie letnim, wyposażonych w klimatyzację, z pociągami bez klimatyzacji.

*Miara: Średnie obłożenie*

*Wymiar: Pociąg (atrybuty wymiaru: Klimatyzacja)*

*Wymiar: Data (atrybuty wymiaru: Pora roku)*

**Sprawdzanie, czy w źródłach danych znajdują się dane potrzebne do zapełnienia hurtowni danych.**

Nazwa tabeli	Kolumna	Źródło
Wykonanie przejazdu	Jeden wiersz opisuje fakt przejazdu pociągu na dany przystanek w ramach danego kursu.	
	ID_Polaczenie	Klucz obcy z tabeli wymiarów. Uzyskany z tabeli Połączenie z bazy relacyjnej.
	ID_Przystanek	Klucz obcy z tabeli wymiarów. Uzyskany z

		tabeli Przystanek z bazy relacyjnej.
	ID_Pociag	Klucz obcy z tabeli wymiarowej. Uzyskany z tabeli Kurs bazy relacyjnej.
	Liczba pasażerów	Wyliczona jako suma liczby pasażerów przedziałowych i bezprzedziałowych.
	Liczba pasażerów przedziałowych	Miara wyliczona na podstawie tabeli Bilet z bazy relacyjnej.
	Liczba pasażerów bezprzedziałowych	Miara wyliczona na podstawie tabeli Bilet z bazy relacyjnej.
	Miejsca przedziałowe	Wyliczone jako suma dostępnych miejsc przedziałowych i bezprzedziałowych.
	Miejsca przedziałowe w pociągu	Liczba dostępnych miejsc w pociągu uzyskana z kolumny Miejsca przedziałowe z pliku Specyfikacja pociągów.
	Miejsca bezprzedziałowe w pociągu	Liczba dostępnych miejsc w pociągu uzyskana z kolumny Miejsca bezprzedziałowe z pliku Specyfikacja pociągów.
	KursNO	Numer kursu uzyskany z tabeli Kurs z kolumny KursID.
	ID_Data	Klucz obcy z tabeli wymiarowej. Uzyskany z tabeli Kurs bazy relacyjnej.
	ID_CzasPrzyjazdu	Klucz obcy z tabeli wymiarowej. Uzyskany z tabeli CzasPrzejazdu z bazy relacyjnej.
	ID_CzasOdjazdu	Klucz obcy z tabeli wymiarowej. Uzyskany z tabeli CzasPrzejazdu z bazy relacyjnej.
<b>Nabycie biletu</b>	<b>Jeden wiersz opisuje ilość nabytych biletów na dany kurs z danym przystankiem początkowym i końcowym.</b>	

	ID_Polaczenie	Klucz obcy z tabeli wymiarów. Uzyskany z tabeli Połączenie z bazy relacyjnej.
	ID_PrzystanekPocz	Klucz obcy z tabeli wymiarów. Uzyskany z tabeli Przystanek z bazy relacyjnej.
	ID_PrzystanekKon	Klucz obcy z tabeli wymiarów. Uzyskany z tabeli Przystanek z bazy relacyjnej.
	KursNO	Numer kursu uzyskany z tabeli Kurs z kolumny KursID.
	Ilość biletów	Miara wyliczona na podstawie tabeli Bilet z bazy relacyjnej.
	ID_Data	Klucz obcy z tabeli wymiarów. Data uzyskana z tabeli Kurs z bazy relacyjnej.
<b>Pociąg</b>	<b>Jeden wiersz opisuje jeden pociąg (implementacja SCD 2)</b>	
	ID_Pociąg	ID Pociągu. Klucz surogatowy - generowany przez bazę danych.
	NrSeryjny	Unikalny nr seryjny pociągu uzyskany z pliku Specyfikacja pociągów.
	Model	Model pociągu uzyskany z pliku Specyfikacja pociągów.
	KategoriaRokProdukcji	Kategorie roku produkcji pociągu Dozwolone wartości: Stare (lata 1990-1999), Nowe (lata 2000-2009), Nowsze (lata 2010-2019), Najnowocześniejsze (lata 2020+) Wyliczona na podstawie danych z pliku Specyfikacja pociągów kolumna Rok produkcji.
	KategoriaPrędkości	Kategorie maksymalnej osiągniętej prędkości

		Dozwolone wartości: Wolne (do 149km/h) Szybkie (od 150km/h do 179km/h) Najszybsze (od 180km/h) Wyliczona na podstawie danych z pliku Specyfikacja pociągów kolumna prędkość maksymalna.
	Klimatyzacja	Czy pociąg jest wyposażony w klimatyzację Dozwolone wartości: Klimatyzowany, Nieklimatyzowany Uzyskane z danych z pliku Specyfikacja pociągów kolumna klimatyzacja.
	CzyAktualny	"1" jeśli informacje aktualne, "0" jeśli nie (implementacja SCD 2)
<b>Połączenie</b>	<b>Jeden wiersz opisuje jedno połączenie.</b>	
	ID_Połączenia	ID Połączenia. Klucz surogatowy - generowany przez bazę danych.
	Nazwa połączenia	Słowna nazwa połączenia uzyskana z tabeli Połączenie z bazy relacyjnej.
<b>Przystanek</b>	<b>Jeden wiersz opisuje jeden przystanek należący do połączenia.</b>	
	ID_Przystanek	ID Przystanku. Klucz surogatowy – generowany przez bazę danych.
	Nazwa przystanku	Słowna nazwa przystanku uzyskana z tabeli Przystanek z bazy relacyjnej.
<b>Data</b>	<b>Jedna krotka opisuje jeden dzień.</b> Wszystkie dane w tej tabeli są generowane krotka po krotce na podstawie dowolnego kalendarza, przed procesem ETL.	
<b>Czas</b>	<b>Jedna krotka opisuje jedną godzinę (niezależnie od daty).</b>	

	Wszystkie dane w tej tabeli są generowane krotka po krotce na podstawie zegara, przed procesem ETL.
--	---