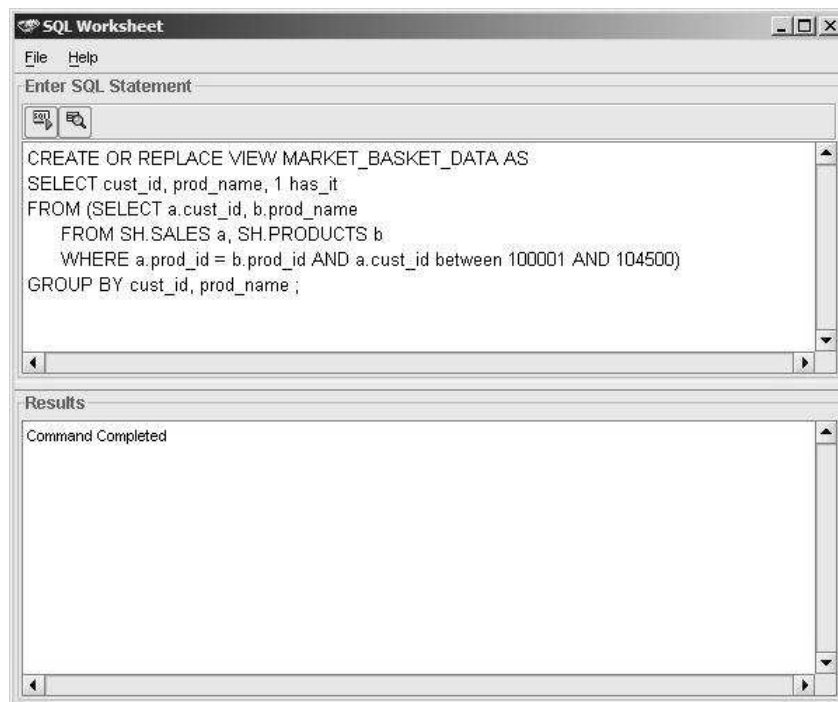


Laboratorium 3

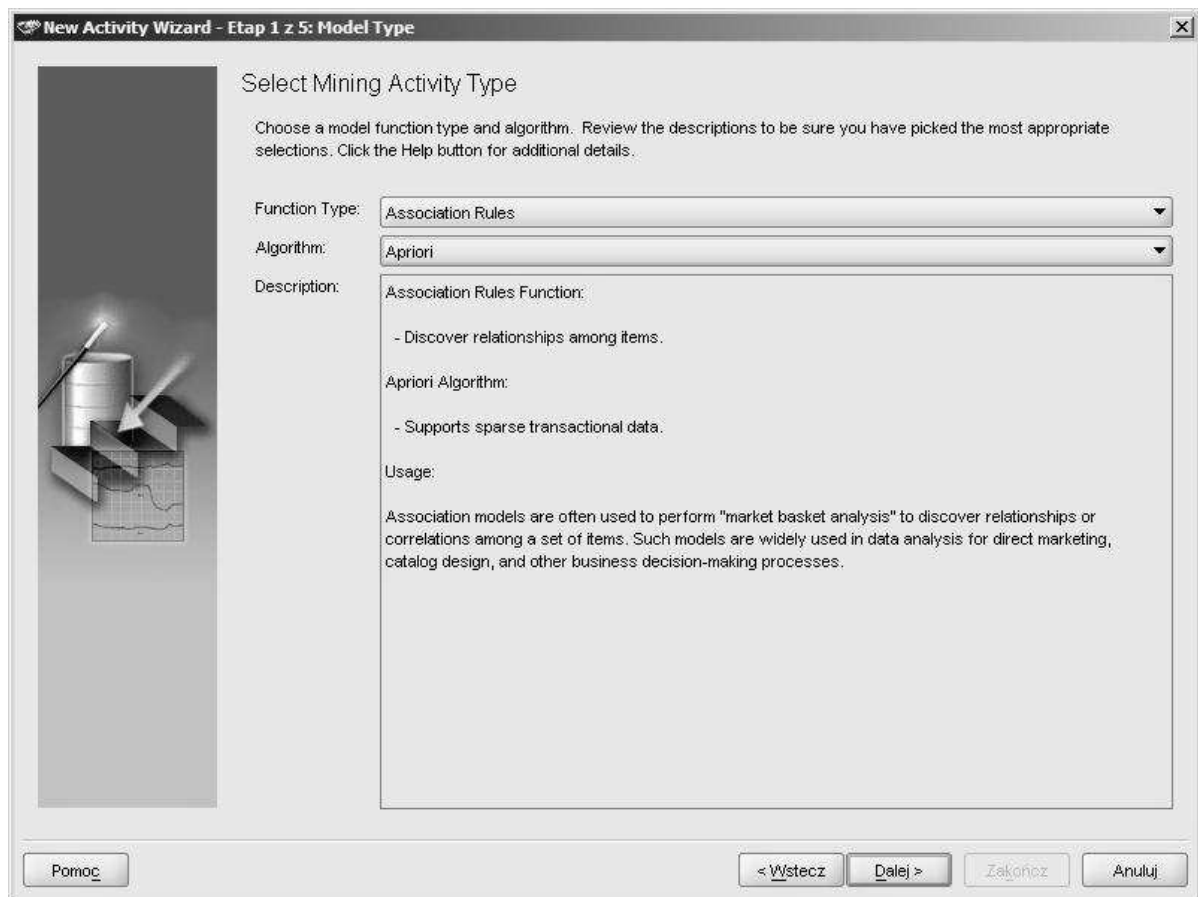
Odkrywanie reguł asocjacyjnych.

1. Uruchom narzędzie Oracle Data Miner i połącz się z serwerem bazy danych.
2. Z menu głównego wybierz Tools→SQL Worksheet. W górnym oknie wprowadź i wykonaj komendę:

```
CREATE OR REPLACE VIEW MARKET_BASKET_DATA AS
SELECT cust_id, prod_name, 1 AS has_it
FROM ( SELECT a.cust_id, b.prod_name
      FROM SH.SALES a, SH.PRODUCTS b
      WHERE a.prod_id = b.prod_id
            AND a.cust_id BETWEEN 100001 AND 104500)
GROUP BY cust_id, prod_name;
```



3. Z menu głównego wybierz Activity→Build. Na ekranie powitalnym kliknij przycisk Dalej>.
4. Z listy Function Type wybierz Association Rules. Jedyne dostępny algorytm, Apriori, zostaje automatycznie wybrany jako algorytm przetwarzający. Kliknij przycisk Dalej>.



5. Oracle Data Miner wymaga do odkrywania asocjacji pionowej organizacji bazy danych, przykład takiej organizacji przedstawiono na rysunku poniżej.

```

Run SQL Command Line
SQL*Plus: Release 10.2.0.1.0 - Production on Pt Lip 28 13:55:06 2006
Copyright (c) 1982, 2005, Oracle. All rights reserved.

SQL> connect student@miner
Proszę podać hasło:
Połączono.
SQL> select * from market_basket_data where rownum < 10 order by cust_id;

  CUST_ID PROD_NAME                                HAS_IT
-----
100100 External 8X CD-ROM                            1
100100 Mouse Pad                                    1
100200 Standard Mouse                                1
100200 Mouse Pad                                    1
100200 Extension Cable                              1
100200 CD-RW, High Speed Pack of 5                  1
100300 SIMM- 16MB PCMCIAII card                     1
100300 18" Flat Panel Graphics Monitor             1
100400 Mouse Pad                                    1

9 wierszy zostało wybranych.

SQL>

```

6. Wskaż schemat STUDENT i tabelę MARKET_BASKET_DATA jako źródło danych do eksploracji. Jako atrybut Item ID wybierz z listy atrybut PROD_NAME. Kliknij przycisk **Dalej>**.

New Activity Wizard - Etap 2 z 4: Data

Select Transactional Data

Choose Transactional Data.

Schema:

Table/View:

Item ID:

Use Name Lookup For Item

Schema

Table

Item ID

Description

Pomoc < Wstecz **Dalej >** Zakończ Anuluj

7. Jako identyfikator transakcji wskaż atrybut CUST_ID. Kliknij przycisk **Dalej>**.

New Activity Wizard - Etap 3 z 4: Data

Transaction ID Selection Step

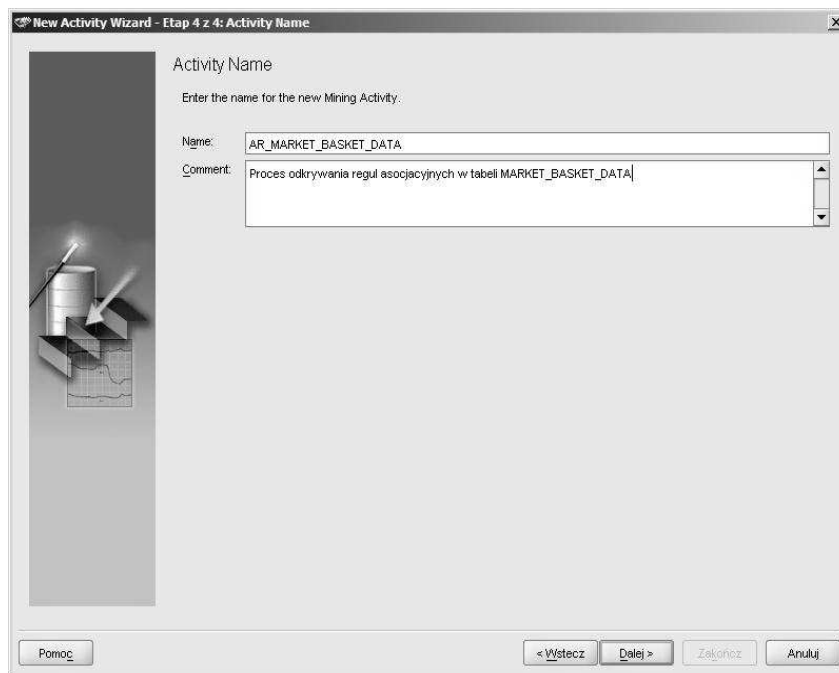
Select one or more attributes uniquely identifying a single transaction.

Transaction Identifier

Select	Attribute
<input checked="" type="checkbox"/>	CUST_ID
<input type="checkbox"/>	HAS_IT

Pomoc < Wstecz **Dalej >** Zakończ Anuluj

8. Podaj nazwę i opis dla procesu eksploracji. **Uwaga:** nie używaj polskich liter w nazwie i opisie. Kliknij przycisk **Dalej>**.



New Activity Wizard - Etap 4 z 4: Activity Name

Activity Name

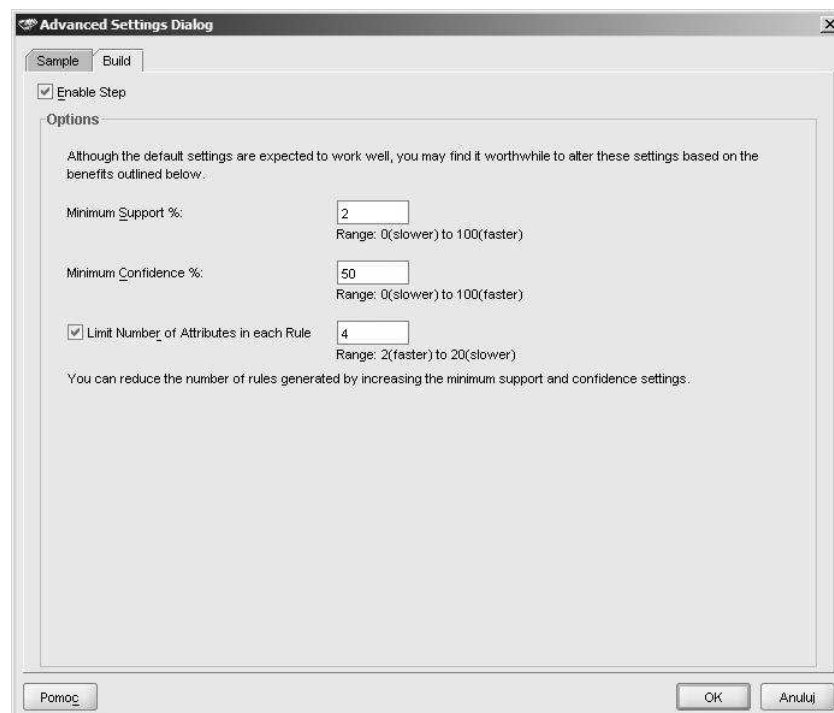
Enter the name for the new Mining Activity.

Name: AR_MARKET_BASKET_DATA

Comment: Proces odkrywania reguł asocjacyjnych w tabeli MARKET_BASKET_DATA

Pomoc < Wstecz Dalej > Zakończ Anuluj

9. Kliknij przycisk **Advanced Settings**. Upewnij się, że na zakładce **Sample** opcja **Enable Step** jest wyłączona (eksploracja zostanie przeprowadzona na całym zbiorze danych). Przejdź na zakładkę **Build**. Zmień wartość parametru minimalnego wsparcia (**Minimum Support %**) na **2**, a wartość parametru minimalnej ufności (**Minimum Confidence %**) na **50**. Ogranicz długość znajdowanych reguł asocjacyjnych do 4 (**Limit Number of Attributes in each Rule**). Kliknij przycisk **OK**. Upewnij się, że opcja **Run upon finish** jest zaznaczona. Kliknij przycisk **Zakończ** i poczekaj na zakończenie wykonywania się algorytmu Apriori.



Advanced Settings Dialog

Sample Build

Enable Step

Options

Although the default settings are expected to work well, you may find it worthwhile to alter these settings based on the benefits outlined below.

Minimum Support %: 2
Range: 0(slower) to 100(faster)

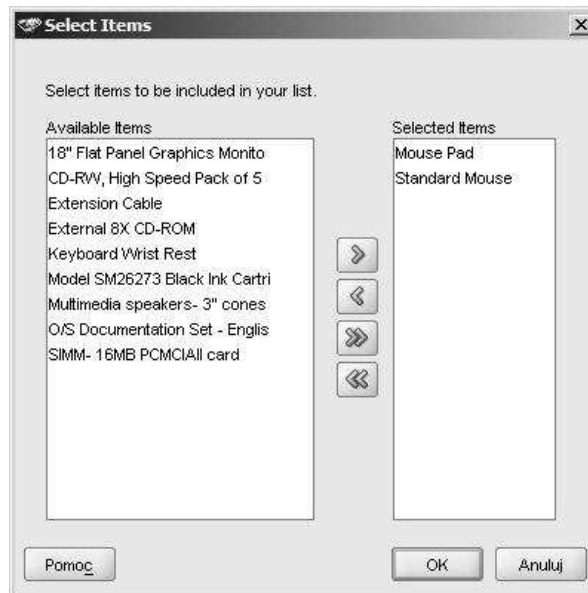
Minimum Confidence %: 50
Range: 0(slower) to 100(faster)

Limit Number of Attributes in each Rule: 4
Range: 2(faster) to 20(slower)

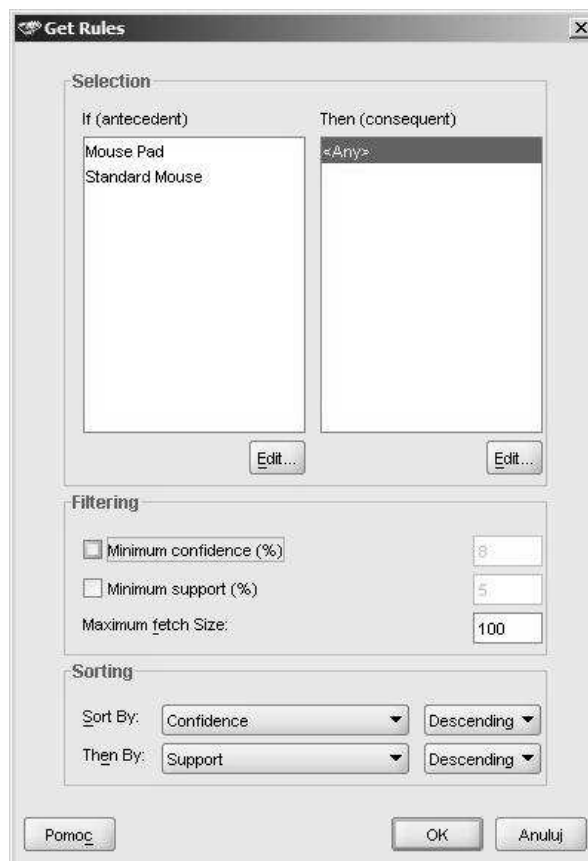
You can reduce the number of rules generated by increasing the minimum support and confidence settings.

Pomoc OK Anuluj


10. Kliknij odnośnik **Results** w bloku Build. Kliknij przycisk **Get Rules** w prawym górnym rogu okna. Kliknij przycisk **Edit** umieszczony pod lewym oknem, zaznacz wartość **Mouse Pad** i przenieś ją do prawego panelu. Podobnie wybierz element **Standard Mouse**. Kliknij przycisk **OK**.

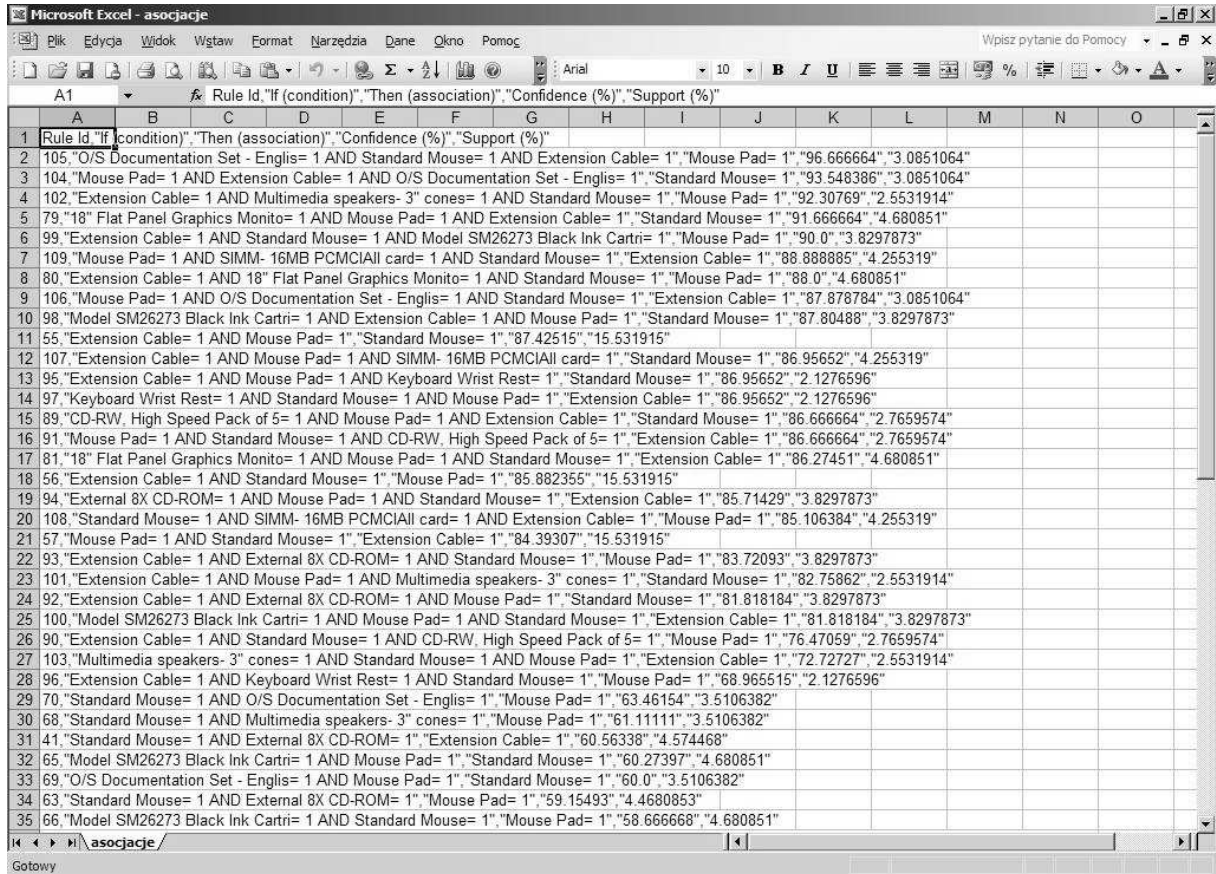


11. W prawym panelu w oknie pobierania reguł pozostaw wpisaną tam wartość **<Any>**. Upewnij się, że pola wyboru związane z filtrowaniem reguł według wsparcia i ufności są odznaczone. Kliknij przycisk **OK**.



12. Wybierz dowolną regułę z listy i zaznacz ją. W dolnym oknie pojawi się szczegółowy

raport na temat znalezionej reguły asocjacyjnej. Kliknij na  (ikonkę eksportu) umieszczoną w prawym górnym rogu tuż nad listą znalezionych reguł asocjacyjnych. Jako format pliku eksportu wybierz CSV. Kliknij przycisk **OK**. Zapisz plik w lokalnym systemie plików. Obejrzyj zawartość pliku.



13. Połącz się z bazą danych wykorzystując iSQLPlus. Wykonaj skrypt ar.sql. Po każdym kroku przeanalizuj uzyskane wyniki (komentarz jest umieszczony wewnątrz skryptu).

Ćwiczenie samodzielne

W schemacie użytkownika **sh** znajdują się tabele PRODUCTS, SALES o następujących schematach:

```
SQL> desc sales
```

Name	Null?	Type
PROD_ID	NOT NULL	NUMBER
CUST_ID	NOT NULL	NUMBER
TIME_ID	NOT NULL	DATE
CHANNEL_ID	NOT NULL	NUMBER
PROMO_ID	NOT NULL	NUMBER
QUANTITY_SOLD	NOT NULL	NUMBER(10,2)
AMOUNT_SOLD	NOT NULL	NUMBER(10,2)

```
SQL> desc products
```

Name	Null?	Type
------	-------	------

```

-----
PROD_ID                NOT NULL NUMBER(6)
PROD_NAME              NOT NULL VARCHAR2(50)
PROD_DESC              NOT NULL VARCHAR2(4000)
PROD_SUBCATEGORY      NOT NULL VARCHAR2(50)
PROD_SUBCATEGORY_ID   NOT NULL NUMBER
PROD_SUBCATEGORY_DESC NOT NULL VARCHAR2(2000)
PROD_CATEGORY         NOT NULL VARCHAR2(50)
PROD_CATEGORY_ID      NOT NULL NUMBER
PROD_CATEGORY_DESC    NOT NULL VARCHAR2(2000)
PROD_WEIGHT_CLASS     NOT NULL NUMBER(3)
PROD_UNIT_OF_MEASURE          VARCHAR2(20)
PROD_PACK_SIZE        NOT NULL VARCHAR2(30)
SUPPLIER_ID           NOT NULL NUMBER(6)
PROD_STATUS           NOT NULL VARCHAR2(20)
PROD_LIST_PRICE       NOT NULL NUMBER(8,2)
PROD_MIN_PRICE        NOT NULL NUMBER(8,2)
PROD_TOTAL            NOT NULL VARCHAR2(13)
PROD_TOTAL_ID        NOT NULL NUMBER
PROD_SRC_ID           NUMBER
PROD_EFF_FROM         DATE
PROD_EFF_TO           DATE
PROD_VALID            VARCHAR2(1)

```

Stwórz we własnym schemacie perspektywę przedstawiającą fakty sprzedaży produktów, weź pod uwagę tylko i wyłącznie nazwę towaru (PROD_NAME). Następnie, znajdź wszystkie reguły asocjacyjne tłumaczące zakup podkładek pod nadgarstki do klawiatury (tj. reguły posiadające produkt "Keyboard Wrist Rest" w następniku).

Na potrzeby eksploracji przy użyciu PL/SQL API dane wejściowe powinny się znajdować w schemacie:

```
SQL> desc sales
```

```

Name                Null?    Type
-----
TRANSACTION_ID      VARCHAR2(128)
PRODUCTS            DMSYS.DM_NESTED_NUMERICALS

```

gdzie:

```
SQL> desc dm_nested_categoricals
```

```

DM_NESTED_CATEGORICALS TABLE OF DMSYS.DM_NESTED_CATEGORICAL
Name                Null?    Type
-----
ATTRIBUTE_NAME      VARCHAR2(30)
VALUE               VARCHAR2(4000)

```

```
SQL> desc dm_nested_numericals
```

```

DM_NESTED_NUMERICALS TABLE OF DMSYS.DM_NESTED_NUMERICAL
Name                Null?    Type
-----
ATTRIBUTE_NAME      VARCHAR2(30)
VALUE               NUMBER

```

Jeśli do ćwiczenia wykorzystujesz Oracle Data Miner, dane nie muszą być transformowane do postaci tabeli zagnieżdżonej.

UWAGA:

- nie kopiuj danych do własnego schematu!!!
- zastanów się, w jaki sposób w perspektywie zbudować atrybut łączący fakty składające się na jeden zakup

- zwróć uwagę, że atrybut `TIME_ID` jest typu `DATE`
- zwróć uwagę, że atrybut `ATTRIBUTE_NAME` ma długość 30 znaków, a w bazie danych występują produkty o dłuższych nazwach.
- do stworzenia perspektywy obiektowej prezentującej produkty kupione w trakcie pojedynczej transakcji jako kolekcję typu `DM_NESTED_CATEGORICALS` lub `DM_NESTED_NUMERICALS` należy wykorzystać konstrukcję `CAST(MULTISET(SELECT ... FROM ...)) AS DM_NESTED_XXXXXXXX)`