

|  |
| --- |
| EXTRACT, TRANSFORM, LOAD  Labwork 6  PL/SQL Techniques |

Оглавление

[1 Загрузка измерений из 3нф в STAR схему 3](#_Toc500148434)

[1.1 Таблицы измерений. 3](#_Toc500148435)

[1.2 Создание измерений 3](#_Toc500148436)

[1.3 Пакеты для загрузки данных 4](#_Toc500148437)

[2 Использование Advanced PL/SQL 5](#_Toc500148438)

[2.1 Bulk collect и Forall 5](#_Toc500148439)

[2.2 For loop 5](#_Toc500148440)

[3 Загрузка и перезагрузка данных. 7](#_Toc500148441)

[3.1 Процедуры загрузки. 7](#_Toc500148442)

[3.2 Тестирование перезагрузки данных измерения SCD1 7](#_Toc500148443)

[3.3 Тестирование перезагрузки данных измерения SCD2 8](#_Toc500148444)

[4 Результаты загрузки данных 10](#_Toc500148445)

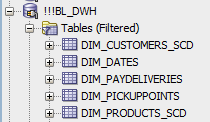
[5 Схема 3NF 12](#_Toc500148446)

# Загрузка измерений из 3нф в STAR схему

Все скрипты примеров находятся на Git в папке [BI-Lab-2017](https://github.com/mikitakandratsiuk/BI-Lab-2017)/[\_2. ETL](https://github.com/mikitakandratsiuk/BI-Lab-2017/tree/master/_2.%20ETL)/[Tasks](https://github.com/mikitakandratsiuk/BI-Lab-2017/tree/master/_2.%20ETL/Tasks)/[Maryna\_Hlazunova](https://github.com/mikitakandratsiuk/BI-Lab-2017/tree/master/_2.%20ETL/Tasks/Maryna_Hlazunova)/Task 06/. Эта директория является домашней для данной лабораторной работы и далее упоминаться не будет.

## Таблицы измерений.

В моем DWH на уровне DM реализовано 5 таблиц измерений: даты, продукты, пункты доставки, клиенты и общий дим для способов доставки и оплаты. Два из них SCD 2 типа.

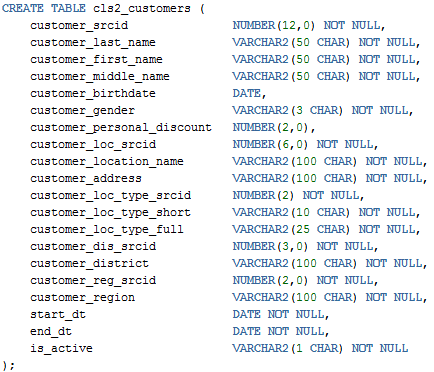


## Создание измерений

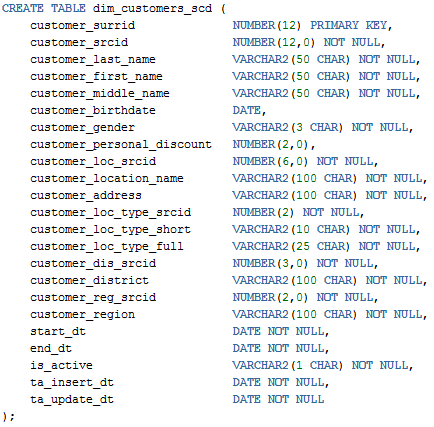
Каждое измерение, кроме даты, находятся в схеме BL\_DWH и реализуются с помощью 3 скриптов: создание таблицы, создание сиквенса и права для других схем. Также используется промежуточные таблицы cls2\_ схемы BL\_CL\_DWH, они практически полностью повторяют структуру dim\_ таблиц, не имеют ключей и используется для промежуточной заливки данных.

Пример на измерении клиентов.

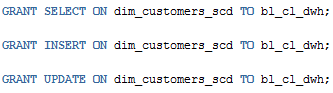
Создание таблицы cls\_:



Создание измерения dim\_:



Необходимые гранты на таблицу dim\_customers\_scd:



Создание сиквенса и грант на него для пользователя, осуществяющего загрузку данных:



## Пакеты для загрузки данных

Загрузка данных в измерения происходит с помощью процедур пакетов. Пример будет располагаться в домашней папке, остальные находятся в папке проекта:

pkg\_load\_dim\_customers\_def.sql

pkg\_load\_dim\_customers\_impl.sql

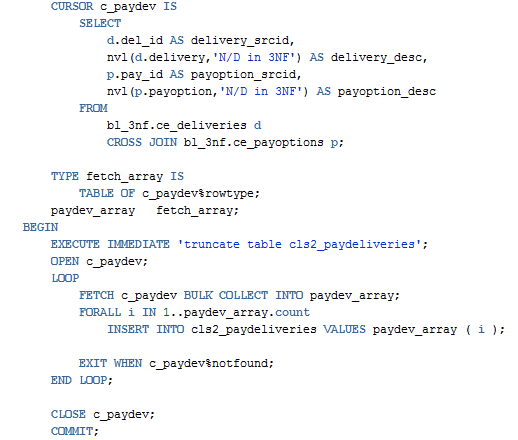
# Использование Advanced PL/SQL

## Bulk collect и Forall

Bulk collect – это select, который извлекает несколько строк с одной выборкой, улучшая скорость извлечения данных.

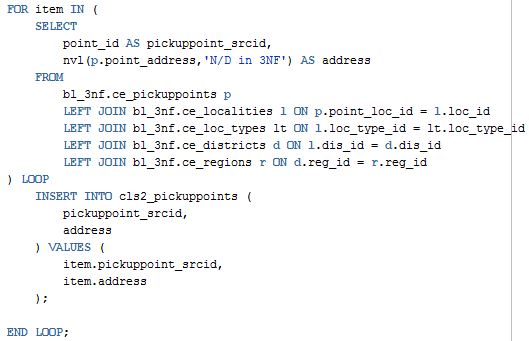
Forall используется для INSERT, UPDATE или DELETE для быстрого изменения нескольких строк данных.

Их можно объединить, тогда результат будет заполнен итерациями выражения FORALL.



## For loop

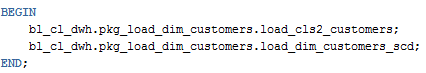
Цикл FOR LOOP в Oracle позволяет выполнить код повторно в течение фиксированного количества раз.



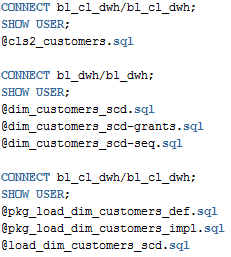
# Загрузка и перезагрузка данных.

## Процедуры загрузки.

Загрузка и перезагрузка данных осуществляется с помощью скриптов, которые вызывают процедуры пакетов. Для каждого измерения используется отдельный пакет.

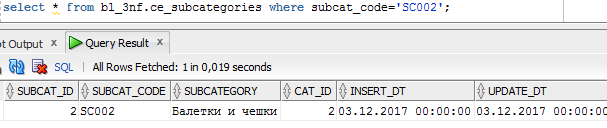


Полное пересоздание объектов и загрузка данных производится с помощью батча:

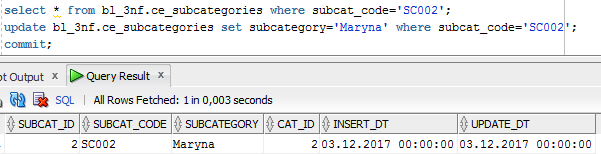


## Тестирование перезагрузки данных измерения SCD1

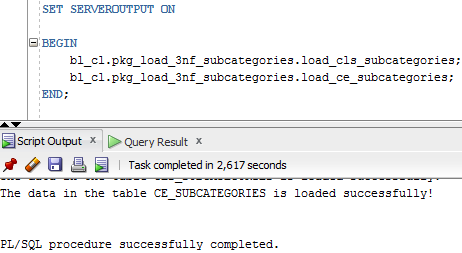
Функционирование системы приведу на одном из измерений SCD01 Подкатегории товаров. Данные на момент начала тестирования:



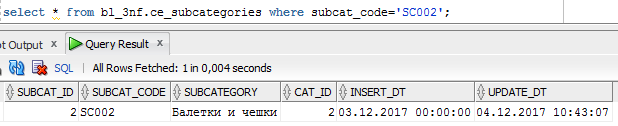
Чтобы не изменять сорс, я изменю название категории в самой таблице, как будто эти данные были изначально залиты в систему:



Далее я запускаю процедуру загрузки подкатегорий:



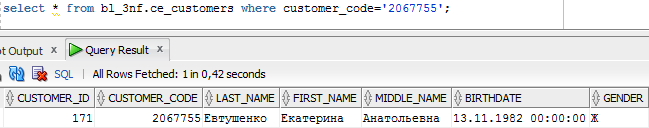
И проверяю полученный результат:

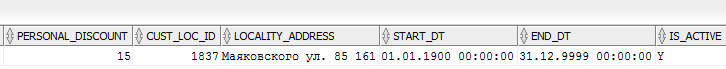


Название подкатегории изменилось на новое, а также была изменена дата обносления записи update\_dt.

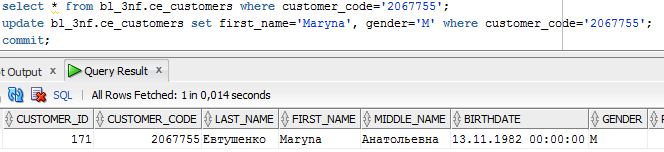
## Тестирование перезагрузки данных измерения SCD2

Функционирование системы приведу на одном из измерений SCD02 Клиенты. Данные на момент начала тестирования:



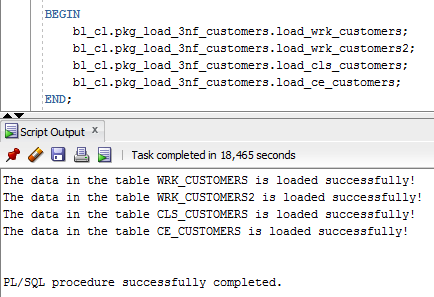


Чтобы не изменять сорс, я изменю атрибуты клиента в самой таблице, как будто эти данные были изначально залиты в систему:

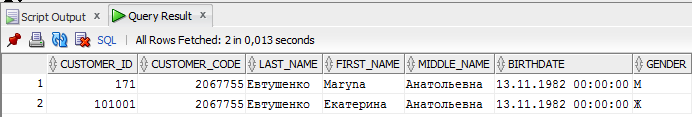


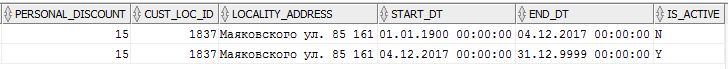


Далее я запускаю процедуру загрузки клиентов:



И проверяю полученный результат:





Данные были перезагружены, старая запись была обносвлена и стала неактивной, о чем говорят значения полей end\_dt и is\_active, а новая была вставлена