

|  |
| --- |
| EXTRACT, TRANSFORM, LOAD  Labwork 1  Overview of ETL and SQL Review |

Оглавление

[1 Создание объектов 3](#_Toc498887350)

[1.1 Создание пользователей 3](#_Toc498887351)

[1.2 Создание директории и внешних таблиц 3](#_Toc498887352)

[1.3 Создание таблиц слоя cleansing 4](#_Toc498887353)

[1.4 Создание таблиц слоя 3NF 4](#_Toc498887354)

[1.5 Создание публичных синонимов 5](#_Toc498887355)

[1.6 Создание последовательностей 5](#_Toc498887356)

[2 зАГРУЗКА данных 6](#_Toc498887357)

[2.1 Загрузка данных в таблицы слоя CLS 6](#_Toc498887358)

[2.2 Загрузка данных в таблицы слоя 3NF 8](#_Toc498887359)

[3 SQL\*Plus 11](#_Toc498887360)

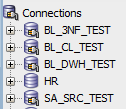
# Создание объектов

Все скрипты 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 01/. Эта директория является домашней для данной лабораторной работы и далее упоминаться не будет.

## Создание пользователей

Для работы нам будут необходимы 4 пользователя. В этой лабораторной только три, но я создам сразу всех необходимых со всеми необходимыми грантами на начальном этапе.

Скрипт на создание - \system\create\_users.sql



## Создание директории и внешних таблиц

Для пользователя SA\_SRC\_TEST нужно дать привилегию на создание директории, из которой будут читаться внешние файлы.

GRANT CREATE ANY DIRECTORY TO SA\_SRC\_TEST;

Также директорию можно создать из-под пользователя SYS, тогда для SA\_SRC\_TEST нужно выдать привилегию на чтение:

GRANT READ ON DIRECTORY ext\_tables TO SA\_SRC\_TEST;

Также подобная привилегия нужна для BL\_CL\_TEST, т.к. будет происходить импорт данных из внешних таблиц в таблицы cls\_ на стороне BL\_CL\_TEST.

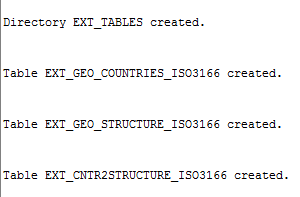
GRANT READ ON DIRECTORY ext\_tables TO BL\_CL\_TEST;

Скрипт на создание директории и выдачу привилегий на нее:

\sa\_src\create\_directory.sql

Скрипт на создание внешних таблиц и выдачу прав на их чтение для BL\_CL\_TEST:

\sa\_src\create\_ext\_tables.sql

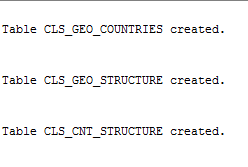


## Создание таблиц слоя cleansing

В общем случае данные из внешних таблиц перегружаются в таблицы wrk\_ c такой же структурой. В случае лабораторной работы загрузка будет сразу производится в таблицы cls\_, т.к. отсутствует задание дальнейшей очистки и валидации данных.

Скрипт на создание таблиц cls\_ и выдачу прав на их чтение для BL\_3NF\_TEST:

\bl\_cl\create\_cls\_tables.sql



## Создание таблиц слоя 3NF

Данные из таблиц cls\_ нормализуются и загружаются в таблицы ce\_ c новой структурой третьей нормальной формы. Здесь при создании таблиц важным условием является создание первичных и внешних ключей для обеспечения целостности данных.

Скрипт на создание таблиц cls\_:

\bl\_3nf\create\_3nf\_tables.sql

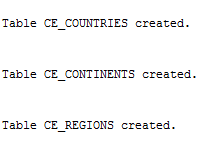
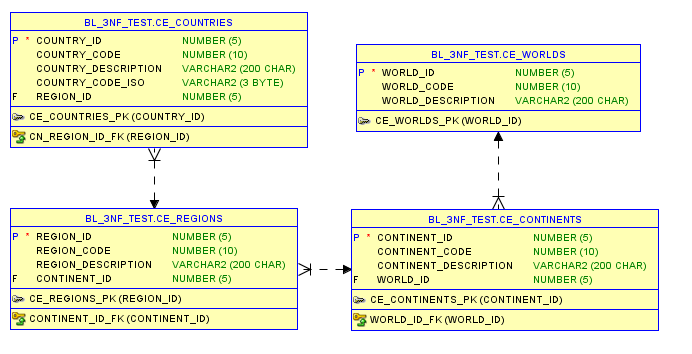


Схема БД, получившаяся в итоге:

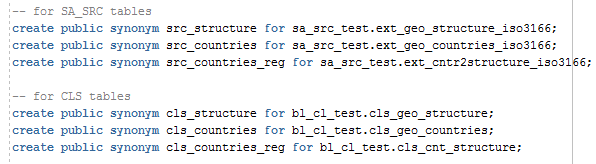


## Создание публичных синонимов

В рамках данной лабораторной работы для упрощения написания доступа к таблицам и добавления смысловой нагрузки в названия таблиц я создала публичные синонимы для таблиц слоя SA\_SRC\_TEST и BL\_CL\_TEST.

Скрипт на создание публичных синонимов:

\system\create\_public\_synonyms.sql



## Создание последовательностей

Для генерации ключей в таблицах 3NF я создавала последовательности.

Скрипты на их создание:

\bl\_3nf\create\_3nf\_sequences.sql

# зАГРУЗКА данных

Загрузка каждой таблицы производится отдельной процедурой.

## Загрузка данных в таблицы слоя CLS

Загрузка данных в таблицы cls\_ производится с помощью пакета pkg\_etl\_load\_cls. Здесь обращение к таблицам происходят через паблик синонимы.

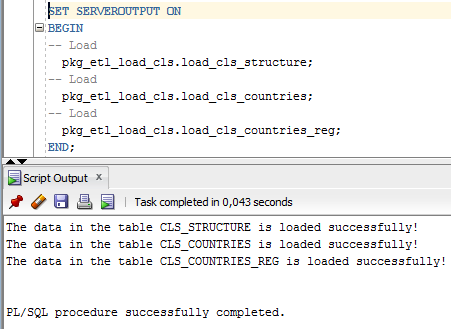
Скрипт на создание спецификации и тела пакета pkg\_etl\_load\_cls:

\bl\_cl\pkg\_etl\_load\_CLS-gef.sql

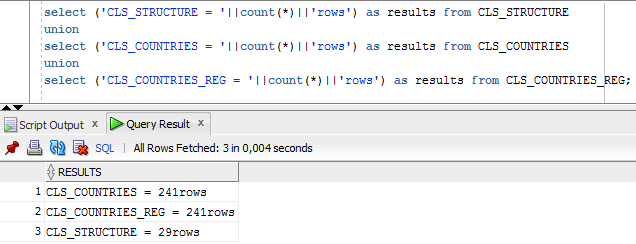
\bl\_cl\ pkg\_etl\_load\_CLS-impl.sql

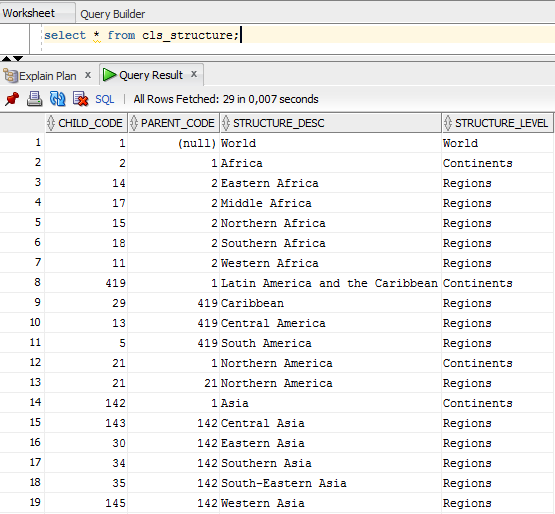
Последовательный вызов процедур для последовательной загрузки таблиц находится в скрипте:

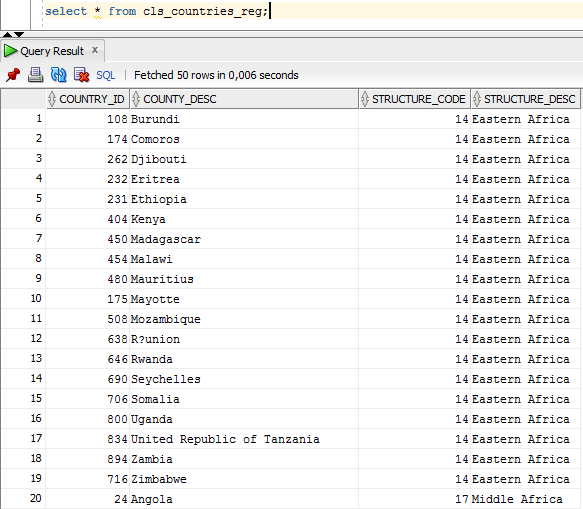
\bl\_cl\ load\_cls\_data.sql

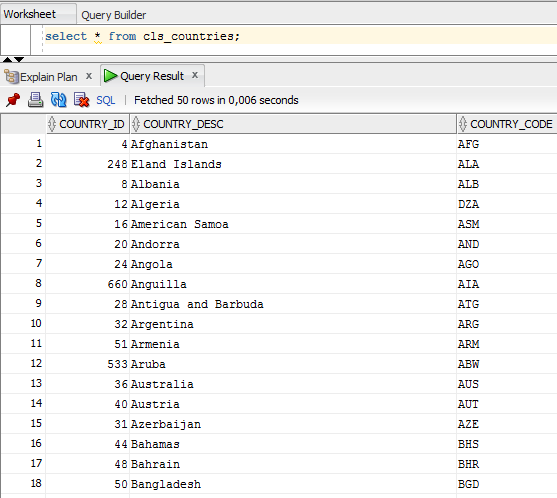


Проверяем таблицы после загрузки данных:









## Загрузка данных в таблицы слоя 3NF

Загрузка данных в таблицы cls\_ производится с помощью пакета pkg\_etl\_load\_3nf. Здесь обращение к таблицам происходят через паблик синонимы.

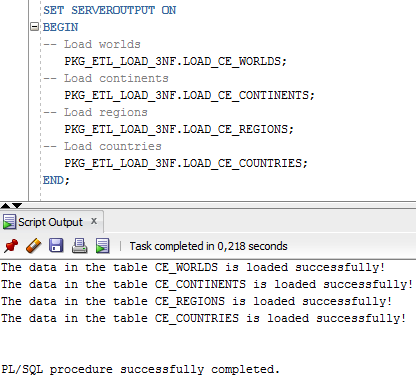
Скрипт на создание спецификации и тела пакета pkg\_etl\_load\_3nf:

\bl\_3nf\pkg\_etl\_load\_3NF-gef.sql

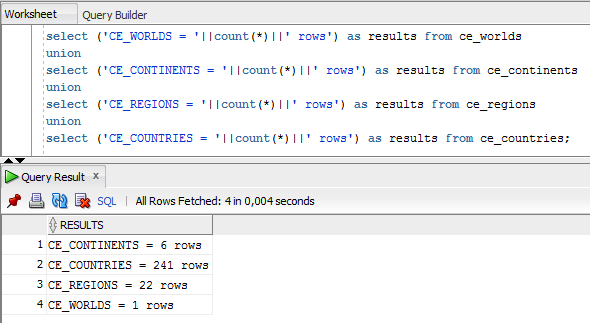
\bl\_3nf\pkg\_etl\_load\_3NF-impl.sql

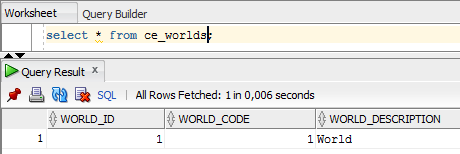
Последовательный вызов процедур для последовательной загрузки таблиц находится в скрипте:

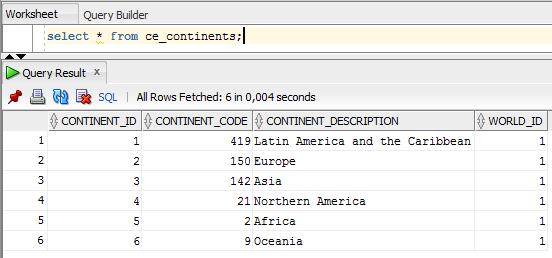
\bl\_3nf\ load\_3nf\_data.sql

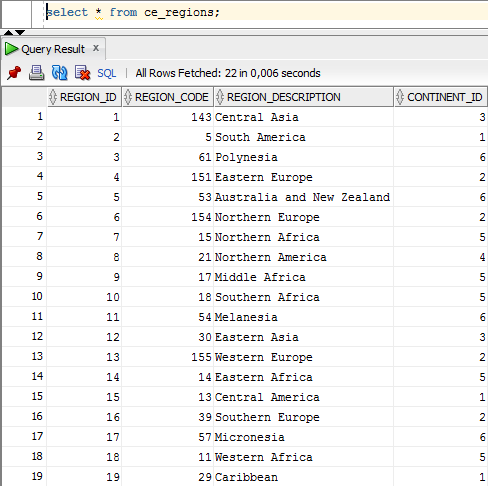


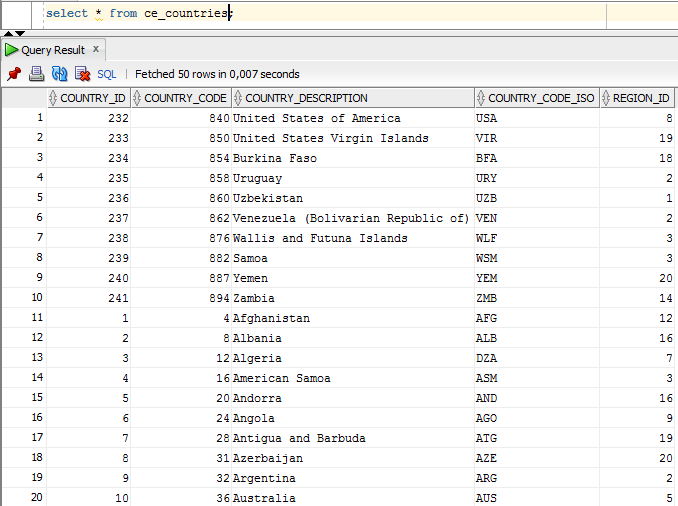
Проверяем таблицы после загрузки данных:





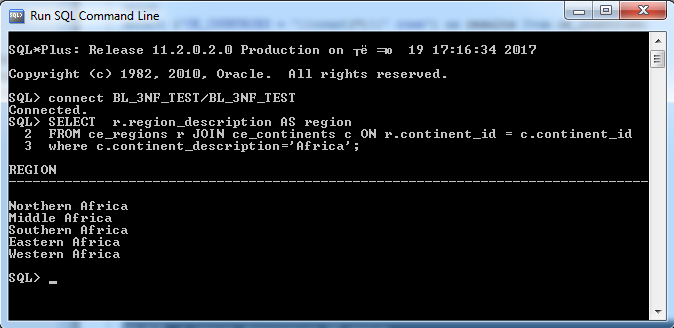




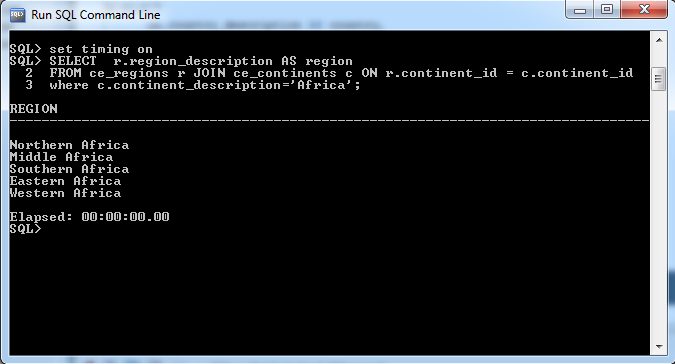


# SQL\*Plus

Подключение к схеме и выполнение запроса:

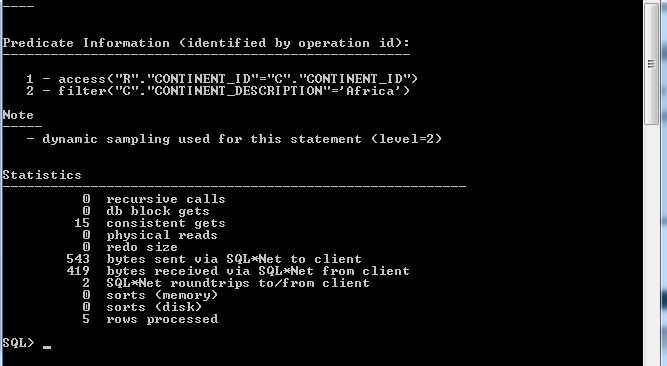


Установка SET TIMING ON позволяет показывать время выполнения запроса.

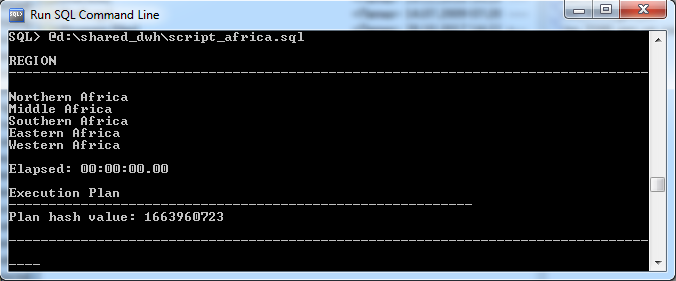


Чтобы была возможность просмотреть план выполнения запросов в SQL\*Plus необходимо воспользоваться установкой: SET AUTOTRACE ON.

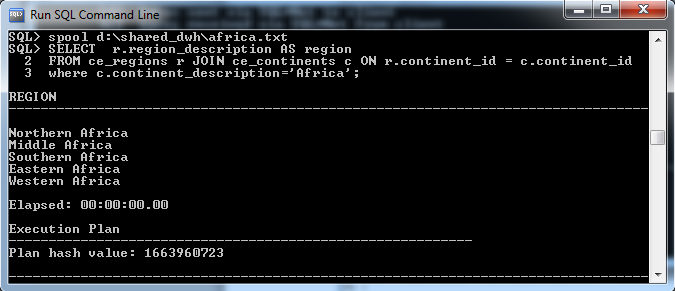




Есть возможность запустить скрипт, воспользовавшись значком @:



Выгрузку в файл можно сделать, используя команды SPOOL путь к файлу и SPOOL OFF, при этом файл создается автоматически:



Так выглядит содержимое файла после выгрузки:

