

**EPAM Systems, RD Dep.**

# Data Load

---

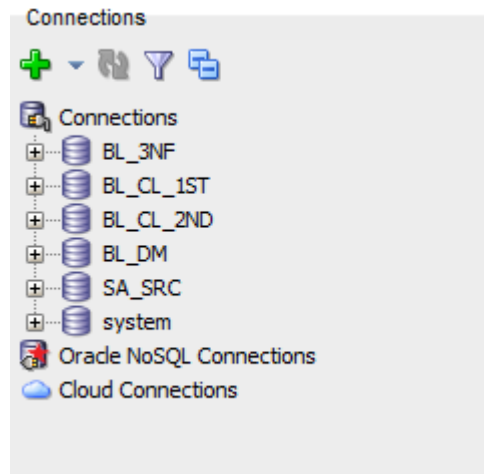
REVISION HISTORY					
Ver.	Description of Change	Author	Date	Approved	
				Name	Effective Date
1.0	Initial status	<a href="#">Valeryia Lupanova</a>	23-NOV-2017		

## Содержание

1. ОПИСАНИЕ СХЕМ .....	3
2. ОПИСАНИЕ SA_SRC .....	4
3. ОПИСАНИЕ BL_CL_1ST .....	7
4. ПРОВЕРКА ДАННЫХ В BL_CL_1ST .....	10

## 1. Описание схем

- В работе были созданы схемы, указанные ниже. Всем схемам были выданы гранты **CONNECT** и **RESOURCES**. Дополнительные гранты, понадобившиеся в ходе выполнения работы можно найти по директории **..\DWSO\SYSTEM\ADDITIONAL\_GRANTS.SQL**.



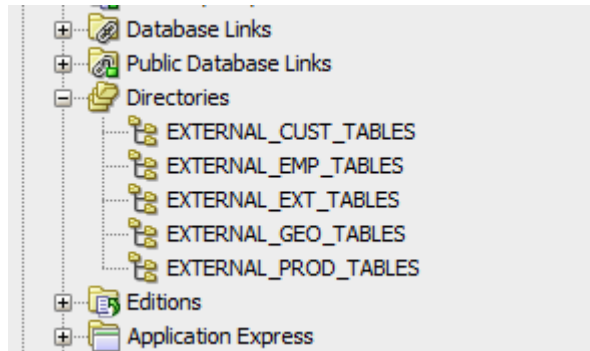
- Загрузка данных внешних источников производится в схеме SA\_SRC.
- Очистка данных проходит на слое BL\_CL\_1<sup>ST</sup>. Для первоначальной загрузки использовались WRK-таблицы. Далее очищенные данные попадают в CLS-таблицы.
- Далее очищенные данные загружаются в нормализованном виде в BL\_3NF.
- Следующий слой BL\_CL\_2<sup>ST</sup> создан для дополнительной обработки данных.
- Дименшен-модель загружается на последнем этапе в схему BL\_DM.

## 2. Описание SA\_SRC

- SA\_SRC был выдан грант на создании директорий.

```
BEGIN
  pkg_grants.user_grant(grant_name => 'CREATE ANY DIRECTORY', user_name => 'SA_SRC');
END;
```

- В схеме все таблицы имеют префикс EXT. Для каждого объекта была создана соответствующая директория для загрузки.



```
CREATE OR REPLACE DIRECTORY external_geo_tables AS
'/media/sf_Valeryia_Lupanava/Project/dwso/sa_src/sources/geography';
```

```
CREATE OR REPLACE DIRECTORY external_cust_tables AS
'/media/sf_Valeryia_Lupanava/Project/dwso/sa_src/sources/customers';
```

```
CREATE OR REPLACE DIRECTORY external_emp_tables AS
'/media/sf_Valeryia_Lupanava/Project/dwso/sa_src/sources/employees';
```

```
CREATE OR REPLACE DIRECTORY external_ext_tables AS
'/media/sf_Valeryia_Lupanava/Project/dwso/sa_src/sources';
```

```
CREATE OR REPLACE DIRECTORY external_prod_tables AS
'/media/sf_Valeryia_Lupanava/Project/dwso/sa_src/sources/products';
```

- Источники данных можно найти по директории ...\\DWSO\\SA\_SRC\\SOURCES.
- Для загрузки применялся следующий скрипт (пример одного из скриптов).

```
-- EXT_BRA_SIZES.
CREATE TABLE ext_bra_sizes
(
  bra_size_uk VARCHAR2 ( 200 CHAR ),
  bra_size_usa VARCHAR2 ( 200 CHAR ),
  bra_size_eu VARCHAR2 ( 200 CHAR ),
  bra_size_fr VARCHAR2 ( 200 CHAR ),
  bra_size_uie VARCHAR2 ( 200 CHAR )
)
ORGANIZATION EXTERNAL
(
  TYPE oracle_loader DEFAULT DIRECTORY external_ext_tables
  ACCESS PARAMETERS (fields terminated BY ',')
  LOCATION ('bra_size_grid.csv')
)
REJECT LIMIT UNLIMITED;
```

- Все скрипты организованы по директории ...\\DWSO\\SA\_SRC\\TABLES.
- Также дополнительно были выданы гранты на чтение директорий.

```

BEGIN
  pkg_grants.USER_GRANT (GRANT_NAME => 'READ ON DIRECTORY external_geo_tables', USER_NAME => 'SA_SRC');
END;

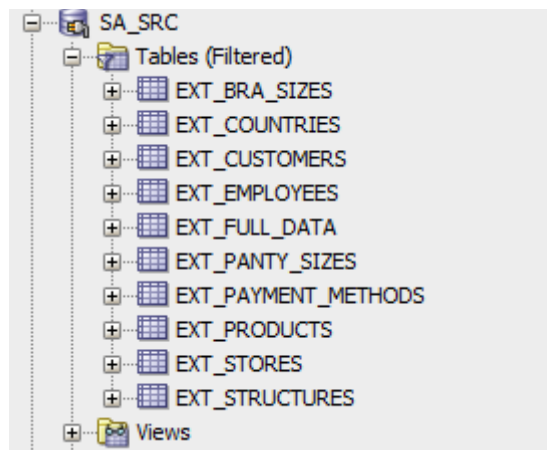
BEGIN
  pkg_grants.USER_GRANT (GRANT_NAME => 'READ ON DIRECTORY external_cust_tables', USER_NAME => 'SA_SRC');
END;

BEGIN
  pkg_grants.USER_GRANT (GRANT_NAME => 'READ ON DIRECTORY external_emp_tables', USER_NAME => 'SA_SRC');
END;

BEGIN
  pkg_grants.USER_GRANT (GRANT_NAME => 'READ ON DIRECTORY external_ext_tables', USER_NAME => 'SA_SRC');
END;

```

- После загрузки источника получилась следующая структура.



- Последним шагом схеме BL\_CL\_1<sup>ST</sup> из схемы SA\_SRC были выданы гранты на селекты из созданных в SA\_SRC таблиц.

```

BEGIN
  pkg_grants.user_grant(grant_name => 'SELECT', schema_name => 'SA_SRC', object_name => 'EXT_COUNTRIES', user_name => 'BL_CL_1ST');
END;

BEGIN
  pkg_grants.user_grant(grant_name => 'SELECT', schema_name => 'SA_SRC', object_name => 'EXT_STRUCTURES', user_name => 'BL_CL_1ST');
END;

BEGIN
  pkg_grants.user_grant(grant_name => 'SELECT', schema_name => 'SA_SRC', object_name => 'EXT_FULL_DATA', user_name => 'BL_CL_1ST');
END;

BEGIN
  pkg_grants.user_grant(grant_name => 'SELECT', schema_name => 'SA_SRC', object_name => 'EXT_BRA_SIZES', user_name => 'BL_CL_1ST');
END;

BEGIN
  pkg_grants.user_grant(grant_name => 'SELECT', schema_name => 'SA_SRC', object_name => 'EXT_PANTY_SIZES', user_name => 'BL_CL_1ST');
END;

```

- Гранты BL\_CL\_1<sup>ST</sup> из схемы SA\_SRC выдавались посредством созданного пакета для выдачи грантов.

```
CREATE OR REPLACE PACKAGE pkg_grants
AUTHID CURRENT_USER
AS
    PROCEDURE user_grant (grant_name IN VARCHAR2,
                          schema_name IN VARCHAR2,
                          object_name IN VARCHAR2,
                          user_name   IN VARCHAR2);

    PROCEDURE user_grant (grant_name IN VARCHAR2,
                          user_name   IN VARCHAR2,
                          comments    IN boolean := FALSE);
END pkg_grants;

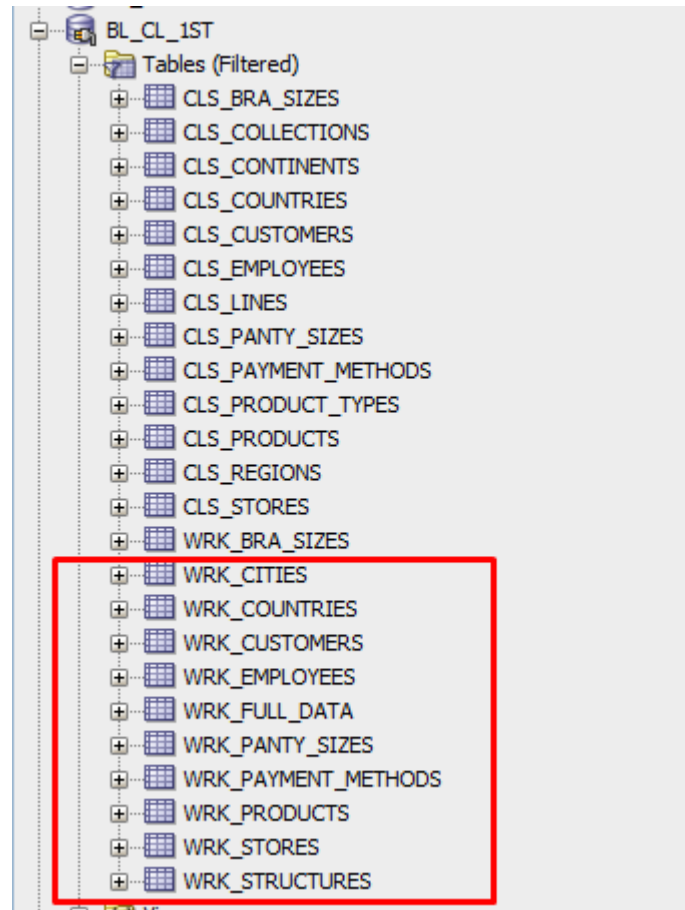
CREATE OR REPLACE PACKAGE BODY pkg_grants AS

    PROCEDURE user_grant (grant_name IN VARCHAR2,
                          schema_name IN VARCHAR2,
                          object_name IN VARCHAR2,
                          user_name   IN VARCHAR2)
    IS
    BEGIN
        EXECUTE IMMEDIATE ('GRANT ' || grant_name || ' ON ' || schema_name || '.' || object_name || ' TO ' || user_name);
    END user_grant;

    PROCEDURE user_grant (grant_name VARCHAR2,
                          user_name   VARCHAR2,
                          comments    IN boolean := FALSE)
    IS
    BEGIN
        EXECUTE IMMEDIATE ('GRANT ' || grant_name || ' ON ' || user_name || ' TO ' || user_name);
    END user_grant;
END pkg_grants;
```

### 3. Описание BL\_CL\_1ST

- В BL\_CL\_1ST первоначально были загружены данные из таблиц EXT в первичальном виде.



- Скрипты на создание объектов находятся в директории **...\DWSO\BL\_CL\_1ST\TABLES**. Для каждого объекта создана папка, в которой расположен скрипт на создание и скрипт на заполнение.
- Заполнение объектов происходило с помощью созданного пакета. Для объектов WRK один пакет загрузки, поскольку здесь необходимо залить все данные, без фильтрации. Т.е. можно передавать в параметре название объектов, не создавая дополнительные процедуры и пакеты.
- Пример заполнения одного объекта – WRK\_BRA\_SIZES.

```
BEGIN
  pkg_etl_insert_wrk.insert_table table_name_to => 'WRK_BRA_SIZES' table_name_from => 'SA_SRC.EXT_BRA_SIZES';
END;
```

- Пример пакета для заполнения WRK-таблиц.

```








CREATE OR REPLACE PACKAGE pkg_etl_insert_wrk
AUTHID CURRENT_USER
AS
    PROCEDURE insert_table (table_name_to IN VARCHAR2,
                           table_name_from IN VARCHAR2);

END pkg_etl_insert_wrk;

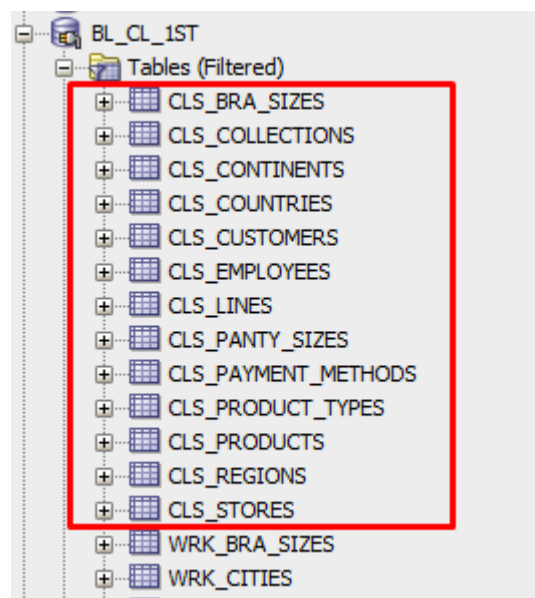
CREATE OR REPLACE PACKAGE BODY pkg_etl_insert_wrk
AS
-----
PROCEDURE insert_table (table_name_to IN VARCHAR2,
                       table_name_from IN VARCHAR2)
IS
BEGIN
    EXECUTE IMMEDIATE ('TRUNCATE TABLE ' || table_name_to);
    EXECUTE IMMEDIATE ('INSERT INTO ' || table_name_to || ' SELECT * FROM ' || table_name_from);
    COMMIT;
END insert_table;
END pkg_etl_insert_wrk;

```

- Для объектов CLS были созданы различные пакеты. Пакет включает несколько процедур только в случае, если объекты относятся к одной предметной области, как, например, GEO-данные.
















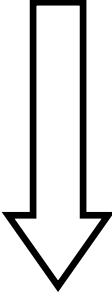


 pkg\_etl\_insert\_customers.sql  
 pkg\_etl\_insert\_employees.sql  
 pkg\_etl\_insert\_geography.sql  
 pkg\_etl\_insert\_payment\_methods.sql  
 pkg\_etl\_insert\_product\_types.sql  
 pkg\_etl\_insert\_sizes.sql  
 pkg\_etl\_insert\_stores.sql

- Структура CLS-объектов.





- Все объекты были организованы подобно WRK объектам.

Name	Date modified	Type
 cls_bra_sizes	11/22/2017 5:17 PM	File folder
 cls_collections	11/22/2017 9:37 PM	File folder
 cls_continents	11/22/2017 5:49 PM	File folder
 cls_countries	11/22/2017 5:50 PM	File folder
 cls_customers	11/22/2017 7:43 PM	File folder
 cls_employees	11/22/2017 7:52 PM	File folder
 cls_lines	11/22/2017 9:37 PM	File folder
 cls_panty_sizes	11/22/2017 5:19 PM	File folder
 cls_payment_methods	11/22/2017 8:34 PM	File folder
 cls_product_types	11/22/2017 9:37 PM	File folder
 cls_products	11/22/2017 9:36 PM	File folder
 cls_regions	11/22/2017 5:50 PM	File folder
 cls_stores	11/22/2017 7:59 PM	File folder
 wrk_bra_sizes	11/22/2017 6:21 PM	File folder
 wrk_cities	11/22/2017 7:42 PM	File folder
		
 cls_bra_sizes.sql	11/22/2017 4:45 PM	Microsoft SQL Ser... 1 KB
 cls_bra_sizes_insert.sql	11/22/2017 5:17 PM	Microsoft SQL Ser... 1 KB

#### 4. Проверка данных в BL\_CL\_1ST

- Первый запрос.

```
SELECT *
FROM cls_customers;
```

Query Result x

SQL | Fetched 50 rows in 0.254 seconds

CUSTOMER_ID	FIRST_NAME	LAST_NAME	AGE	AGE_CATEGORY_ID	EMAIL	PHONE	ADDRESS	CITY
1	WH6962478	Erina	Stein	18	1estein4@japanpost.jp	506-953-1399	40 La Follette Hill	39020
2	ZI4049515	Burnard	Zellmer	83	5bzellmerd@biglobe.ne.jp	664-251-7379	08226 Iowa Center	38733
3	UC2398683	Jacklin	Taree	60	5jtareek@techcrunch.com	518-474-2046	81 Anzinger Way	41747
4	UC2398683	Jacklin	Taree	60	5jtareek@techcrunch.com	518-474-2046	81 Anzinger Way	37138
5	SK3212168	Jaquenette	Daltry	68	5jdaltryl@umn.edu	205-368-0250	65159 Tennyson Junction	19725
6	IB2171816	Ashleigh	Bent	41	3abentia@phpbb.com	194-172-6082	3993 Fairfield Point	17770
7	HS1243709	Oneida	Huby	36	3ohuby1m@instagram.com	635-481-1497	3 Tomscot Parkway	7482
8	UY4206588	Jeremy	Ghioni	52	4jghioni73@ameblo.jp	737-554-0514	20 Jana Lane	23980
9	NE3927166	Cleavland	Caitlin	18	1ccaitlin7g@163.com	761-442-8247	136 Bellgrove Center	17694
10	RM3071685	Magdalena	Smelley	19	1msmelley8p@china.com.cn	723-757-9531	58922 Steensland Trail	16038
11	RK1776128	Perry	Stonman	63	5pstonman30@angelfire.com	372-323-5105	041 Grayhawk Street	8073
12	QZ6767142	Elke	Markovic	67	5emarkovic34@alibaba.com	940-261-2496	910 Londonderry Terrace	36411
13	XC7820217	Cristobal	Cawson	70	5ccawson39@ask.com	799-525-9915	20973 Steensland Trail	3160
14	AY7244175	Glory	Jeduch	22	1gjeduch44@nih.gov	332-671-9036	0 Tennessee Terrace	23818
15	ST8222964	Arlyne	McLaughlin	24	1amclaughlin4j@surveymonkey.com	480-414-1841	44 Graedel Alley	3359
16	MC6296113	Wileen	Goggey	41	3wgoggey57@fda.gov	392-540-1895	4981 Stuart Way	48083
17	GT7324936	Tab	Canelas	21	1tcanelasm3@dion.ne.jp	744-903-5029	1 8th Center	34260
18	TH7071714	Rafael	Blaik	55	5rblaikmb@tuttocitta.it	732-684-6910	6031 Lotheville Drive	20448
19	TP2485692	Dimitri	Zettler	33	2dzettler8v@prweb.com	754-507-5037	17722 Banding Parkway	29348

- Второй запрос.

```
SELECT *
FROM cls_employees;
```

Query Result x

SQL | Fetched 50 rows in 0.093 seconds

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	STORE_SRCID	POSITION_NAME	POSITION_GRADE_ID	WORK...	EMAIL
1	11-3923431	Jamesy	Innett	572	sales representative	novice	11jinnett1c@yandex.ru
2	04-5097138	Felicdad	Rizzo	529	financial advisor	expert	14frizzo26@storify.com
3	66-8305520	Albertina	Dodwell	980	sales representative	experienced	5adodwell2i@icio.us
4	98-5813199	Audy	Gavigan	394	sales associate	middle	7agavigan3f@facebook.
5	66-2469351	Riccardo	D'Hooge	417	store manager	middle	6rdhooge4x@imageshack
6	92-6419853	Austina	Cottis	194	financial planner	expert	3acottis57@census.gov
7	82-6220300	Allene	Fluit	995	account executive	experienced	5afluit6q@cocolog-nif
8	93-3610809	Gardner	Bristow	4	financial advisor	novice	5gbristow7x@wired.com
9	61-4469517	Jourdan	Shiels	954	sales associate	middle	6jshiels7z@webnode.co
10	74-4967445	Fran	Yakovitch	279	sales director	expert	8fyakovitch8s@oracle.
11	05-4381847	Cthrine	Killerby	316	financial planner	middle	1ckillerby91@who.int
12	25-6249483	Candie	Throssell	225	account executive	novice	15cthrossellcn@quantca
13	23-5224324	Corilla	Puckett	279	executive vice president of sales	middle	6cpuckettcu@bizjourna
14	34-8101145	Brien	Fonquernie	390	retail sales consultant	experienced	8bfonquernied2@dailyw
15	41-8546150	Sindee	Edgerton	439	financial planner	experienced	6sedgertond5@google.c
16	76-9117523	Caressa	Oleksiak	433	director of national sales	middle	3coleksiakdj@domainma
17	50-7016051	Lorianne	Bodd	708	sales coordinator	middle	1lbodddn@flavors.me
18	95-8429058	Avril	Ciccottini	541	financial advisor	experienced	10aciccottining@merria
19	96-3289646	Roana	Pickover	867	sales director	experienced	10rpickovernw@hexun.co

- Третий запрос.

```
SELECT *
FROM cls_products;
```

	PRODUCT_ID	PRODUCT_NAME	LINE_ID	PRODUCT_TYPE_ID
1	1	Fuscia	10	4
2	2	Turquoise	8	4
3	3	Crimson	10	4
4	4	Orange	4	4
5	5	Crimson	10	4
6	6	Puce	10	4
7	7	Orange	4	4
8	8	Puce	4	4
9	9	Red	7	4
10	10	Indigo	6	4
11	11	Mauv	6	4
12	12	Pink	8	4
13	13	Goldenrod	6	4
14	14	Violet	4	4
15	15	Pink	4	4
16	16	Red	4	4
17	17	Indigo	10	4
18	18	Maroon	7	4
19	19	Aquamarine	8	4
20	20	Violet	10	4

- Объекты в CLS, как можно видеть, уже в нормализованном виде, поскольку они в таком же виде будут загружаться в 3NF.