Anton Slizh's

U2M10.LW.ETL Overview - Load and Transformation

Task 1

2.1. Task 01: Transformation Description

Oracle gives us the following choices for transforming data inside the database:

- Transforming Data Using SQL
- Transforming Data Using PL/SQL
- Transforming Data Using Table Functions

In the building business model, the best solutions are using the SQL and PL/SQL transformations.

The data transforming and loading processes in the preparing DWH often use the small tables and simple transformations. These actions can be successfully completed by using standard functionality of INSERT, UPDATE, MERGE statements. Also, I should note that I have used very often the 'UPSERT' functionality to INSERT new rows into the table and UPDATE existing rows. The MERGE statement is really good choice for this action.

Sometimes for realization more complex transformations the standard SQL functionality is not enough. In preparing DWH for more complex and large tables such as sales or products (with SCD2 implementation) more efficient and logically simpler is to use the PL/SQL statements. For example, a PL/SQL procedure could open multiple cursors and read data from multiple source tables, combine this data using complex business rules, and finally insert the transformed data into one or more target table. It would be difficult or impossible to express the same sequence of operations using standard SQL statements.

Task 2

3.1. Task 02: Loading to SAL Layer Data

The Main Task is to load dimension to SAL layer

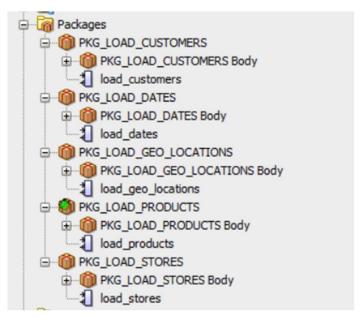
Required points:

- Create new package for Load FCT_* and DIM_* to SAL Layer
- Load Dimension
- Load SCD Dimension
- Load FCT *

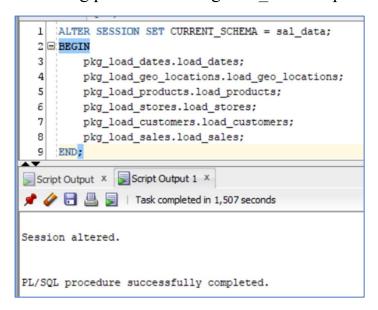
First step is to create SAL layer objects (Dimension and Fact tables) using prepared scripts.



After initializing tables we should to transform data from the DW layer into Star Layer using prepared procedures.



Executing procedures using load data script:

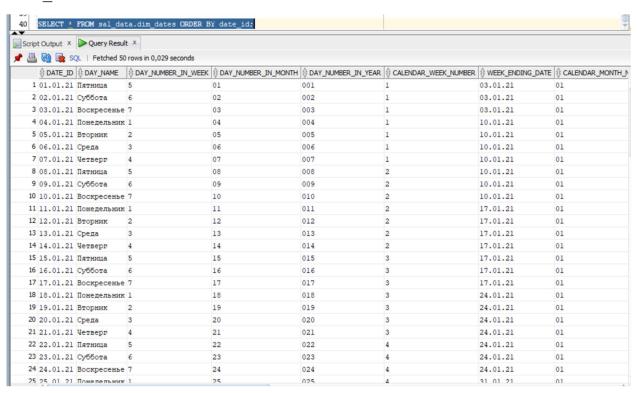


So, let's look at the data of our final Star Schema:

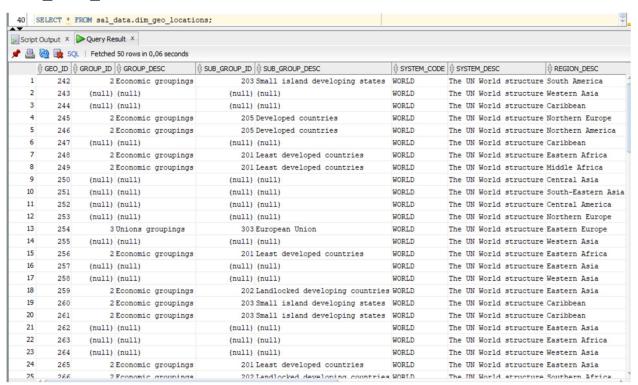
DIM CUSTOMERS

cript	Output ×	uery Result X							
	SQL	Fetched 50 rows	in 0,117 second	s					
	CUSTOMER_ID	♦ FIRST_NAME	& LAST_NAME		⊕ COUNTRY	⊕ EMAIL	⊕ BIRTHDAY	∯ INSERT_DT	⊕ UPDATE
1	104970	Ethan	Lewis	+375297914925	Belarus	EthanLewis@gmail.com	31.03.81	20.08.22	20.08.2
2	109701	Joe	Campbell	+375297917249	Belarus	JoeCampbell@gmail.com	07.10.82	20.08.22	20.08.2
3	110379	Jordan	Gonzalez	+375297912894	Russian Federation	JordanGonzalez@gmail.com	21.04.83	20.08.22	20.08.2
4	104829	Harold	Martinez	+375297912502	Ukraine	HaroldMartinez@gmail.com	22.01.98	20.08.22	20.08.2
5	104781	John	Wright	+375297915635	Belarus	JohnWright@gmail.com	19.04.01	20.08.22	20.08.2
6	105204	Michelle	Davis	+375297912296	Belarus	MichelleDavis@gmail.com	06.08.99	20.08.22	20.08.2
7	106256	Jesse	Nguyen	+375297916093	Belarus	JesseNguyen@gmail.com	08.04.96	20.08.22	20.08.2
8	110305	Ashley	Lewis	+375297914980	Belarus	AshleyLewis@gmail.com	18.10.93	20.08.22	20.08.2
9	107993	Jack	Lee	+375297913888	Belarus	JackLee@gmail.com	31.05.98	20.08.22	20.08.2
10	105228	Roy	Martin	+375297913804	Belarus	RoyMartin@gmail.com	09.08.95	20.08.22	20.08.2
11	106630	Larry	Moore	+375297913493	Belarus	LarryMoore@gmail.com	22.03.97	20.08.22	20.08.2
12	107338	Kimberly	Allen	+375297915493	Belarus	KimberlyAllen@gmail.com	06.12.94	20.08.22	20.08.2
13	107447	Roy	Green	+375297916492	Kazakhstan	RoyGreen@gmail.com	19.10.01	20.08.22	20.08.2
14	109859	Rebecca	Thompson	+375297914223	Belarus	RebeccaThompson@gmail.com	01.01.89	20.08.22	20.08.2
15	107411	Ethan	White	+375297914285	Belarus	EthanWhite@gmail.com	06.10.93	20.08.22	20.08.2
16	107344	Willie	Wright	+375297915718	Belarus	WillieWright@gmail.com	24.06.88	20.08.22	20.08.2
17	109469	Bryan	Adams	+375297916607	Belarus	BryanAdams@gmail.com	01.06.99	20.08.22	20.08.2
18	105074	Mark	Wright	+375297915647	Russian Federation	MarkWright@gmail.com	27.11.88	20.08.22	20.08.2
19	105662	Ethan	Rivera	+375297917101	Belarus	EthanRivera@gmail.com	10.05.82	20.08.22	20.08.2
20	109335	Paul	Sanchez	+375297914498	Kazakhstan	PaulSanchez@gmail.com	26.02.01	20.08.22	20.08.2
21	107390	Amanda	Martinez	+375297912554	Belarus	AmandaMartinez@gmail.com	30.08.99	20.08.22	20.08.2
22	110541	Sharon	Campbell	+375297917296	Belarus	SharonCampbell@gmail.com	09.10.95	20.08.22	20.08.2
23	105001	Lawres	Carter	+375297917499	Belarus	LawresCarter@gmail.com	10.03.95	20.08.22	20.08.2

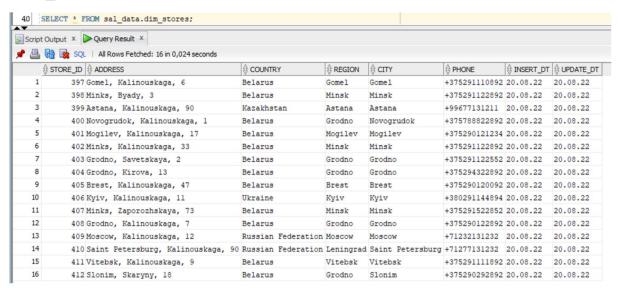
DIM DATES



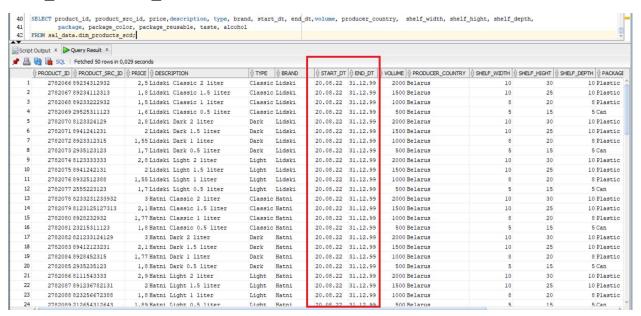
DIM GEO LOCATIONS



DIM_STORES



DIM_PRODUCTS_SCD



FCT SALES

Script Output X	Query R	esult ×							
P 🚇 🚻 🕦 squ	Fetche	d 50 rows in 0,069	seconds						
\$ SALE_ID {	DATE_ID	PRODUCT_ID	CUSTOMER_ID	\$ STORE_ID	∯ GEO_ID				∯ SUM
1 65116278 2	9.11.21	2782085	107091	410	473	2	20211129000115	20.08.22	3,
2 651162791	1.02.22	2782139	107091	399	476	3	20220211000116	20.08.22	5,
3 65116280 2	3.08.21	2782130	106489	402	412	1	20210823000201	20.08.22	2,
4 651162811	7.07.22	2782129	110641	402	412	1	20220717000301	20.08.22	1,
5 65116282 0	7.03.21	2782096	110547	404	412	2	20210307000405	20.08.22	3,5
6 65116283 2	0.12.21	2782126	110547	405	412	2	20211220000409	20.08.22	13
7 65116284 1	3.10.21	2782124	110547	406	453	2	20211013000413	20.08.22	
8 65116285 2	7.07.22	2782103	104332	402	412	3	20220727000501	20.08.22	
9 651162861	5.02.22	2782133	108251	407	412	2	20220215000603	20.08.22	3,
10 65116287 3	0.04.21	2782067	108251	406	453	1	20210430000613	20.08.22	1,
11 65116288 2	9.05.22	2782095	108251	399	476	1	20220529000616	20.08.22	2,
12 65116289 0	4.12.21	2782143	107279	410	473	2	20211204000715	20.08.22	3,
13 65116290 2	9.10.21	2782109	109922	407	412	1	20211029000803	20.08.22	2,
14 65116291 0	3.04.22	2782107	110642	412	412	3	20220403000907	20.08.22	10,
15 651162921	1.06.21	2782083	110642	409	473	1	20210611000914	20.08.22	2,
16 65116293 0	5.06.21	2782144	105675	397	412	3	20210605001012	20.08.22	4,
17 65116294 2	6.06.21	2782122	105675	406	453	3	20210626001013	20.08.22	5,
18 65116295 0	6.05.21	2782128	105675	410	473	3	20210506001015	20.08.22	4,
19 651162961	8.06.21	2782104	105675	399	476	3	20210618001016	20.08.22	6,
20 651162971	3.12.21	2782140	106066	407	412	3	20211213001103	20.08.22	4,
21 651162981	1.04.21	2782110	106066	412	412	2	20210411001107	20.08.22	1
22 65116299 3	0.04.22	2782104	110643	398	412	1	20220430001202	20.08.22	2,
23 651163001	5.08.21	2782104	110643	404	412	1	20210815001205	20.08.22	2,
24 65116301 0	5.07.22	2782129	110643	404	412	3	20220705001205	20.08.22	4,
25 651163021	7.02.21	2782118	110643	406	453	1	20210217001213	20.08.22	1,

The DIM_PRODUCTS_SCD table is SCD2 type dimension. So, for each row we have the start date and end date which represent the time period of being actual. The surrogate key for each product is *product_id* and the natural key is *product_src_id*.

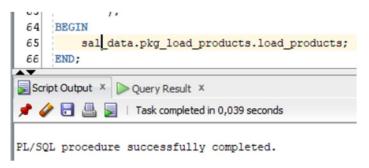
Let's try to update any product and show the SCD2 mechanism in use. For example, choose the first product in the screen – Lidski Classic 2 liter (good choice by the way).

 $Product_id-278066, Product_src_id-89234312932.$

Manually add to the DW layer new row with the same *product_src_id* attribute value, but with the updated *producer_country* attribute value and the tomorrow *insert dt*.

```
45 INSERT INTO dw_data.dw_products
     VALUES (
 46
 47
              dw_data.seq_products.NEXTVAL,
 48
              '89234312932'.
 49
              2.5,
              'Lidski Classic 2 liter',
 50
 51
              'Classic',
 52
              'Lidski',
              'Germany',
 53
              2000,
 54
 55
              10,
              30,
 56
 57
              10,
 58
              'Plastic',
              'Brown',
 59
 60
              'Reusable',
 61
              'Classic',
 62
              1.2,
              SYSDATE + 1
 63
 64
              );
Script Output X Query Result X
📌 🤌 🔒 📕 | Task completed in 0,046 seconds
1 row inserted.
```

Now let's reload products dimension data.



Look at our updated product at the dimension:



We can see that new row became actual for this product (*product_src_id*). So, this way we can store the history of changes in our product dimension.

Also let's create some Data Marts on our Star Scheme. I have used the VIEWs which select all necessary data from prepared Star Scheme.

DM_ALL_CUSTOMERS

```
CREATE OR REPLACE VIEW sal_data.all_customers
 AS
     SELECT c.customer_id,
           first_name,
           last_name,
           phone,
          country,
           birthday,
            amount,
     FROM sal_data.dim_customers c
     JOIN (SELECT customer id,
          count (amount) AS amount,
               sum(sum) AS sum
          FROM sal_data.fct_sales
          GROUP BY customer_id) s
     ON s.customer_id = c.customer_id
     ORDER BY sum DESC;
```

crip	Output X Query Result X SQL Fetched 50 row	s in 0,089 second	ls					
	CUSTOMER_ID FIRST_NAMI	& LAST_NAME	∯ PHONE	⊕ COUNTRY	⊕ EMAIL	⊕ BIRTHDAY		∯ SUM
1	109293 Donna	Smith	+375297911399	Belarus	DonnaSmith@gmail.com	19.03.97	71	345,5
2	109130 Lisa	Gonzalez	+375297912927	Belarus	LisaGonzalez@gmail.com	13.05.82	74	325,25
3	108651 Gabriel	White	+375297914307	Russian Federation	GabrielWhite@gmail.com	11.02.91	66	323,07
4	104956 Margaret	White	+375297914338	Belarus	MargaretWhite@gmail.com	28.06.91	70	321,76
5	105153 Adam	Ramirez	+375297914790	Belarus	AdamRamirez@gmail.com	30.08.96	68	315,31
6	105989 Stephanie	Lopez	+375297912814	Belarus	StephanieLopez@gmail.com	04.03.86	68	313,1
7	105209 Wayne	Martinez	+375297912521	Kazakhstan	WayneMartinez@gmail.com	18.08.89	62	313,09
8	106958 Jeremy	Allen	+375297915440	Belarus	JeremyAllen@gmail.com	20.02.86	65	312,4
9	107745 Alexan	Baker	+375297916828	Belarus	AlexanBaker@gmail.com	24.02.92	62	308,54
10	109359 Roy	Torres	+375297915980	Belarus	RoyTorres@gmail.com	05.12.85	61	308,36
11	109001 Jonat	Baker	+375297916819	Russian Federation	JonatBaker@gmail.com	28.04.92	70	308,14
12	106832 Benja	Moore	+375297913497	Kazakhstan	BenjaMoore@gmail.com	23.01.91	68	307,36
13	104935 Deborah	Lopez	+375297912813	Belarus	DeborahLopez@gmail.com	16.10.90	67	307,03
14	108818 Mark	Miller	+375297912063	Belarus	MarkMiller@gmail.com	17.11.95	63	304,52
15	105027 Frank	Wilson	+375297912989	Russian Federation	FrankWilson@gmail.com	23.03.96	64	303,94
16	108213 Emily	Young	+375297915366	Belarus	EmilyYoung@gmail.com	22.10.95	70	303,2
17	107846 Gerald	Hall	+375297916981	Belarus	GeraldHall@gmail.com	19.07.88	61	302,44
18	106823 Aaron	Wilson	+375297912996	Belarus	AaronWilson@gmail.com	25.06.90	68	302,11
19	110415 Jonat	Scott	+375297915795	Russian Federation	JonatScott@gmail.com	06.05.86	62	302,03
20	110189 Timothy	Harris	+375297914377	Belarus	TimothyHarris@gmail.com	10.10.95	65	302,03
21	105465 Joe	Hill	+375297916225	Belarus	JoeHill@gmail.com	28.08.86	65	301,76
22	109163 Karen	Allen	+375297915486	Ukraine	KarenAllen@gmail.com	21.04.89	63	300,98
23	106977 Alan	Green	+375297916487	Belarus	AlanGreen@gmail.com	13.08.96	64	300,02
24	110576 Dennis	Taylor	+375297913377	Belarus	DennisTaylor@gmail.com	26.03.84	66	297,48
25	105505 Ashley	Brown	+375297911780	Belarus	AshleyBrown@gmail.com	17.08.90	65	296,56
26	109033 Gabriel	Rodriguez	+375297912387	Relarus	GabrielRodriguez@gmail.com	06 04 83	66	296 16

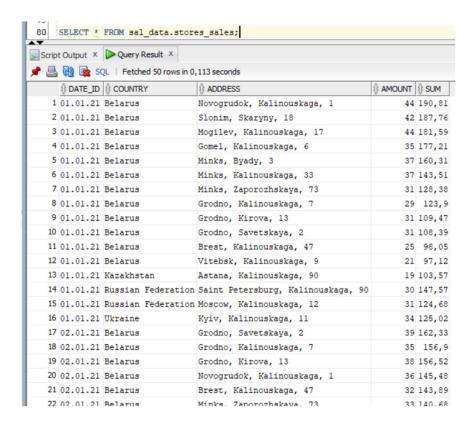
DM ALL BRANDS

```
ECREATE OR REPLACE VIEW sal data.all brands
      SELECT s.date id,
              p.brand,
               COUNT (s.amount) AS amount,
               SUM(s.sum) AS sum
     FROM sal_data.fct_sales s
      LEFT JOIN sal data.dim products scd p
      ON s.product id = p.product id
      GROUP BY s.date_id, p.brand
      ORDER BY date_id, sum DESC;
 80 | SELECT * FROM sal data.all brands;
 Script Output X Query Result X
 📌 🚇 🝓 🕵 SQL | Fetched 50 rows in 0,146 seconds
      150 624,87
    1 01.01.21 Hatni
                        152 543,55
    2 01.01.21 Lidski
                         73 472,5
    3 01.01.21 Zaporozhski
    4 01.01.21 Alivarski
                             84 360,92
    5 01.01.21 Ruski
                             62 205,5
                        174 687,45
    6 02.01.21 Lidski
                        126 510,1
    7 02.01.21 Hatni
                         78 494,3
    8 02.01.21 Zaporozhski
    9 02.01.21 Alivarski
                             74 293,46
    10 02.01.21 Ruski
                             68 213,8
   11 03.01.21 Liqsk.
12 03.01.21 Zaporozhski 89 00.,
21 Hatni 119 492,13
80 383,24
                         178 693,75
                         80 383,24
```

DM STORE SALES

15 03.01.21 Ruski

89 278,5



And at the end let's create the script which join together all load procedures to execute whole ETL process.

```
1
     BEGIN
  2
          -- Cleansing Layer
         dw_cl.pkg_load_products.load_products;
  3
  4
         dw cl.pkg load stores.load stores;
  5
         dw cl.pkg load customers.load customers;
         dw_cl.pkg_load_sales.load_sales;
  6
  7
  8
         -- DW Layer
  g
         dw_data.pkg_load_dates.load_dates;
 10
         dw_data.pkg_load_geo_locations.load_geo_locations;
 11
         dw data.pkg load products.load products;
 12
         dw_data.pkg_load_stores.load_stores;
 13
         dw data.pkg load customers.load customers;
 14
         dw_data.pkg_load_sales.load_sales;
 15
 16
         -- SAL Layer
 17
         sal data.pkg load dates.load dates;
         sal_data.pkg_load_geo_locations.load_geo_locations;
 18
 19
         sal_data.pkg_load_products.load_products;
 20
         sal_data.pkg_load_stores.load_stores;
        sal_data.pkg_load_customers.load_customers;
 21
 22
         sal data.pkg load sales.load sales;
 23 END;
Script Output X
📌 🧼 🖥 🚇 📄 | Task completed in 6,349 seconds
PL/SQL procedure successfully completed.
```